The Collapsing Method: Does the Direction of Collapse Affect Accuracy?

Sarah Greenfield, Francisco Chiclana, Robert John[1]

1. Centre for Computational Intelligence, School of Computing, De Montfort University, Leicester LE1 9BH, UK
Email: {sarahg, chiclana, rij}@dmu.ac.uk

Abstract— *The Greenfield-Chiclana Collapsing Defuzzifier comes in many variants, depending on the order of slice collapse. The accuracy of the fundamental variants of forward, backward, inward and outward, and the composite variants of forward-backward and outward right-left is compared experimentally for the discretised interval type-2 fuzzy set.*

Keywords— Centroid, Collapsing, Defuzzification, Interval Type-2 Fuzzy Set, Representative Embedded Set.

1 Introduction

A fuzzy inferencing system (FIS) is a computerised system that uses fuzzy sets and rules to support decision making. *Type-2* FISs are being developed for an increasing number of applications such as [1], [2], [3]. There are five main stages to any FIS: fuzzification, antecedent computation, implication, aggregation and defuzzification. In the case of a type-2 FIS (where at least one fuzzy set is type-2), defuzzification consists of two parts — type-reduction and defuzzification proper, as shown in figure 1. Type-reduction is the procedure by which a type-2 fuzzy set is converted to a type-1 fuzzy set, known as the type-reduced set (TRS). The TRS is then easily defuzzified to give a crisp number.

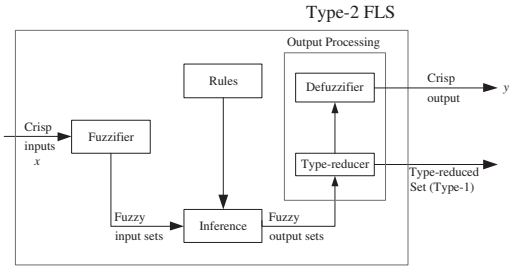


Figure 1: Type-2 FIS (from Mendel [4]).

The type-reduction stage of type-2 defuzzification is problematic owing to its computational complexity. This is the type-reduction algorithm originally described by Mendel ([4], pages 248-252):

1. All possible type-2 embedded sets ([4], definition 3-10, page 98) are enumerated.
2. For each embedded set the minimum secondary membership grade is found.
3. For each embedded set the domain value of the type-1 centroid of the type-2 embedded set is calculated.
4. For each embedded set the secondary grade is paired with the domain value to produce a set of ordered pairs (x,z). It is possible that for some values of x there will be more than one corresponding value of z.
5. For each domain value, the maximum secondary grade is selected. This creates a set of ordered pairs $(x,zMax)$ such that there is a one-to-one correspondence between x and $zMax$. This completes the type-reduction of the type-2 set to the type-1 TRS.

The resultant TRS, as with any type-1 fuzzy set, is readily defuzzified by finding its centroid.

Thus type-reduction involves the processing of *all* the embedded sets within the type-2 set. This is why we term the procedure 'exhaustive defuzzification'. Embedded sets are very numerous. For instance, when a prototype type-2 FIS performed an inference using sets which had been discretised into 51 slices across both the x and y-axes, the number of embedded sets in the aggregated set was calculated to be in the order of 2.9×10^{63}. Though individually easily processed, embedded sets in their totality give rise to a processing bottleneck simply by virtue of their high cardinality. Consequently, exhaustive defuzzification is an impractical technique.

A computationally simpler alternative to the exhaustive method is the *Greenfield-Chiclana Collapsing Defuzzifier*, introduced in [5]. This technique converts an interval type-2 fuzzy set into a type-1 fuzzy set which approximates to the *representative embedded set (RES)*, whose defuzzified value is equal to that of the original type-2 set. As a type-1 set, the RES may then be defuzzified straightforwardly.

In this paper we build on the work reported in [5]. The next section covers assumptions and definitions; section 3 presents an overview of the collapsing defuzzifier, after which section 4 introduces the theme of this article, the notion that *there are variants of the collapsing method.* Sections 5 and 6 are concerned with finding experimentally the most accurate variant.

2 Preliminaries

2.1 Assumptions

Discretisation The work presented here is concerned only with defuzzification of *discretised* type-2 fuzzy sets.

Interval Type-2 Fuzzy Set This paper is concerned with the *interval* type-2 fuzzy set.

Centroid Method of Defuzzification It is assumed that the centroid method of defuzzification ([6], page 336) is used.

2.2 Definitions

Definition 1 (Degree of Discretisation) *The* degree of discretisation *of a discretised fuzzy set is the separation of the slices.*

ISBN: 978-989-95079-6-8

Scalar Cardinality For type-1 fuzzy sets, Klir and Folger ([7], p17) define scalar cardinality as follows:

Definition 2 (Scalar Cardinality) *The* scalar cardinality *of a fuzzy set A defined on a finite universal set X is the summation of the membership grades of all the elements of X in A. Thus,*

$$| A |= \sum_{x \in X} \mu_A(x). \qquad ([7], p17)$$

To distinguish scalar cardinality from cardinality in the classical sense, we adopt the '$\| \|$' symbol for scalar cardinality.

3 Overview of the Collapsing Method

An interval type-2 set may be regarded as a blurred type-1 set. The collapsing method is a technique for deriving a type-1 fuzzy set from a type-2 fuzzy set, and may be thought of as a reversal of blurring. The type-1 set's membership function is calculated so that its defuzzified value approximates that of the type-2 fuzzy set. It is a simple matter to defuzzify the type-1 set, and to do so would be to find the defuzzified value of the original type-2 fuzzy set. The collapsing process approximates the output of the type-reducer followed by the type-1 defuzzifier, and in so doing reduces the computational complexity of type-2 defuzzification. We term this special type-1 set the *'representative embedded set approximation (RESA)'*. (Full details of the collapsing algorithm, including proof of the

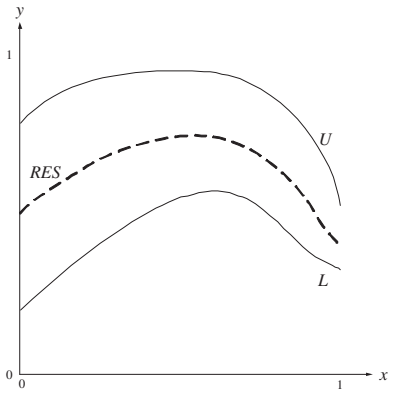


Figure 2: A Representative Embedded Set.

associated theorem, may be found at [5].) We formally state the Simple[1] Representative Embedded Set Approximation:

Theorem 1 (Simple Representative Embedded Set Approx.) *The membership function of the* embedded set *R derived by dynamically collapsing slices of a discretised type-2 interval fuzzy set $\tilde{F}$, having lower membership function L, and upper membership function U, is:*

$$\mu_R(x_i) = \mu_L(x_i) + r_i$$

with

$$r_i = \frac{\left(\|L\| + \sum_{j=1}^{i-1} r_j\right) b_i}{2\left(\|L\| + \sum_{j=1}^{i-1} r_j\right) + b_i}, \qquad (1)$$

and $b_i = \mu_U(x_i) - \mu_L(x_i)$, $r_0 = 0$.

This is an iterative formula. Collapsing proceeds vertical slice by vertical slice. The first slice is collapsed, the first y-value of the RESA calculated, the next slice is collapsed and the second y-value of the RESA calculated, and so on until all the slices have been collapsed. In this formula each b_i is the blur for vertical slice i, i.e. the difference between the upper membership function and the lower membership function for slice i. Each r_i is the amount by which the y-value of L must be increased to give the y-value of the RESA R.

4 Variants of the Collapsing Method

As we have seen, equation (1) is the formula for collapsing. This is in fact *a version* of collapsing — the most intuitive variant, whereby the slices are collapsed in the order of increasing domain value ($x = 0$ to $x = 1$). We term this *collapsing forward*. However slice collapse may be performed in any slice order giving slightly different RESAs. If the domain of the interval type-2 fuzzy set is discretised into s vertical slices, the number of permutations of these slices is $s!$ ([8], page 139). Therefore there must be $s!$ RESAs obtainable by varying the order of slice collapse. The question this piece of research addresses is, "Does the order in which the slices are collapsed affect the accuracy of the method?"

There are four fundamental variants, which we term *forward*, *backward*, *outward* and *inward*. Inward and outward may each be approached in two different ways. For the inward variant, slice collapse might start from the left or from the right. For the outward variant, the first slice is in the middle[2], but the second slice may be to the right or to the left. Added to these, there are composite variants, such as *forward-backward*, which is the mean of the defuzzified values found by collapsing forward and collapsing backward.

5 Experimental Comparison

5.1 Test Sets

We chose to test defuzzification in isolation from the rest of an FIS, on specially created test sets. Our methodology was to run different collapsing variants against each other to see which gave the most accurate results.

Symmetric Horizontal Test Set The lower membership function is the line $y = 0.2$; the upper membership function the line $y = 0.8$. The shape of this test set may be described as a horizontal stripe. The symmetry of this set tells us that its defuzzified value is 0.5.

Symmetric Triangular Test Set This is a normal test set. The lower and upper membership functions are both triangular in shape, both with vertices at $(0.4, 1)$. The symmetry of this set reveals its defuzzified value to be 0.4.

Asymmetric Gaussian Test Set This test set was deliberately designed to be asymmetrical, and hence a more realistic simulation of an FIS aggregated set. Both the lower and upper membership functions are Gaussian. As this set has no symmetry, exhaustive defuzzification (section 1) had to be employed to determine the actual defuzzified value, which, as

[1] In [5], we used the term 'simple' to describe an interval set in which each vertical slice consists of only two points, corresponding to L and U. The term is redundant in the context of this paper.

[2] We always employ an odd number of slices, giving a determinate middle slice.

ISBN: 978-989-95079-6-8

would be expected, varies slightly with the degree of discretisation.

5.2 Test Strategy

A preliminary set of tests was performed on the fundamental variants: forward, backward, inward, outward, and the composite variant forward-backward. As will be reported in the next section, the outward variant outperformed the others. This led on to further tests to discover the most accurate version of collapsing outward.

For the Gaussian test set, the only way of knowing its defuzzified value was to employ exhaustive defuzzification (section 1). This procedure only works properly for 21 slices or under. In contrast, the horizontal and triangular test sets reveal their defuzzified values by symmetry, so the number of slices used can be much higher, allowing finer discretisation.

6 Results and Conclusion

6.1 Preliminary Tests

Table 1 gives the results for the horizontal test set; table 2 gives the associated errors. Table 3 shows the triangular test set results, and table 4 the errors. The defuzzification results for the Gaussian test set are shown in table 5, with the errors in table 6. For all three test sets, the best performing variant was outward, followed by inward, then forward and backward. For the symmetrical sets (horizontal and triangular), the errors of collapsing forward were equal and opposite to those of collapsing backward. Therefore in these cases we would expect collapsing forward-backward to give exact results. This has been confirmed by experiments. For the Gaussian test set, backward performed more poorly than forward. In this case the composite of forward-backward performed worse than forward, though better than backward.

6.2 Further Tests

The outward variant may be performed in two ways, outward right and outward left (section 4). Collapsing right-left is the mean of collapsing right and collapsing left. The results and associated errors for the three versions of the outward variant as applied to the three test sets are shown in tables 7 to 9.

For the symmetrical horizontal test set, outward right and outward left gave rise to equal but opposite errors. For the composite outward right-left, these errors cancelled to zero.

The triangular test set, though symmetrical, is not placed symmetrically about $x = 0.5$. The errors of collapsing right and collapsing left were of equal sign and either equal or very close in quantity. When the errors were not equal, those of outward left were marginally smaller than those of outward right. Unsurprisingly, the performance of the composite of outward right-left fell between that of outward right and outward left.

For the Gaussian test set, the errors were all of negative sign. At all degrees of discretisation, outward left gave the best results, outward right the poorest, and outward right-left came in between.

For two of the three test sets outward left outperformed outward right. Our conjecture is that the position of the centroid is an important factor affecting which performs better out of outward right and outward left. This topic requires further research using a wider range of test sets, but for now we conclude that the optimum strategy is the composite outward right-left.

6.3 Why is Outward the Most Accurate Variant?

This explanation is based on the symmetrical horizontal test set. As each slice is collapsed, $\|L\| + \sum_{j=1}^{j=i-1} r_j$ in both the numerator and denominator of the collapsing formula (equation (1)) increases, which means that as the collapse progresses, the r_i for each collapsed slice i is a closer approximation to $\frac{1}{2}b_i$, i.e. half the 'blur' term. Thus with every successive collapsed slice, the RESA tends towards the midline of L and U, as shown in figure 3 for the forward and backward variants.

For the symmetrical horizontal test set, we take the RES to be the midline of L and U for two reasons. Firstly, by symmetry we would expect the RES to be a horizontal line. Secondly, as the number of slices is increased (either as the collapse progresses, or as the degree of discretisation is made finer), the RESA gets closer to the midline of L and U.

Therefore, as the slices are collapsed, the RESA approaches the RES. This means that the earlier slices in the RESA deviate more from the RES than the later ones. To get the best results, the collapse need to proceed symmetrically. Both the inward and outward variants meet this criterion; the inaccuracies are distributed symmetrically. However the greatest inaccuracy is associated with the first collapsed slice. To achieve maximum accuracy, the ideal place for this first slice to be positioned is centrally, as the effect on the defuzzified value obtained is then minimal. For this reason outward (figure 4) gives a more accurate defuzzified values than inward. We would expect the same reasoning to apply to all type-2 fuzzy test sets. However further investigation, using radically contrasting test sets, is planned.

References

[1] Hani Hagras. A hierarchical type-2 fuzzy logic control architecture for autonomous mobile robots. *IEEE Transactions on Fuzzy Systems*, 12:524–539, 2004.

[2] Robert I. John and Sarah Lake. Modelling nursing perceptions using type-2 fuzzy sets. In *Proc. EUROFUSE 2001*, pages 241–246, 2001.

[3] Turhan Ozen and Jonathan M. Garibaldi. Investigating adaptation in type-2 fuzzy logic systems applied to umbilical acid-base assessment. In *Proc. EUNITE 2003*, pages 289–294, Oulu, Finland, 2003.

[4] Jerry M. Mendel. *Uncertain rule-based fuzzy logic systems: Introduction and new directions*. Prentice-Hall PTR, 2001.

[5] Sarah Greenfield, Francisco Chiclana, Simon Coupland, and Robert I. John. The collapsing method of defuzzification for discretised interval type-2 fuzzy sets. 2008. Accepted for *Information Sciences Special Edition on Higher Order Fuzzy Sets*. DOI: 10.1016/j.ins.2008.07.011.

[6] George J. Klir and Bo Yuan. *Fuzzy sets and fuzzy logic*. Prentice-Hall P T R, 1995.

[7] George J. Klir and Tina A. Folger. *Fuzzy sets, uncertainty, and information*. Prentice-Hall International, 1992.

[8] ed. Carol Gibson. *The facts on file dictionary of mathematics*. Facts on File, New York and Oxford, second edition, 1988.

ISBN: 978-989-95079-6-8

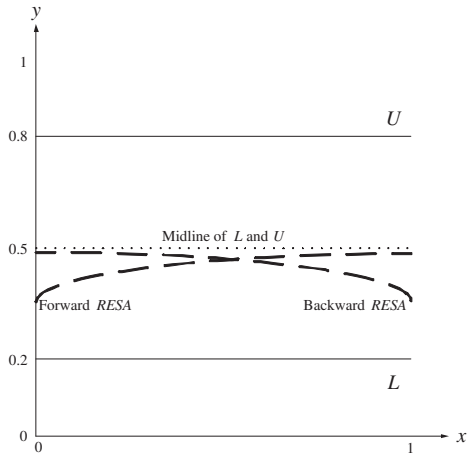


Figure 3: Forward RESA and Backward RESA.

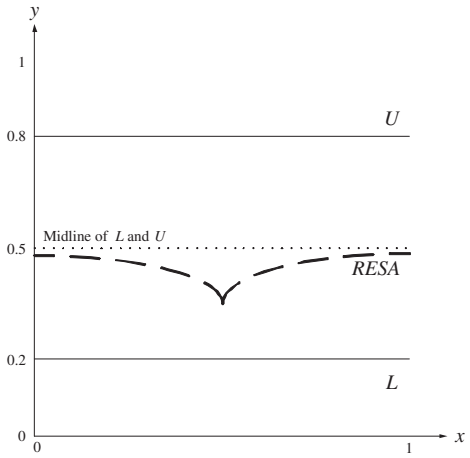


Figure 4: Outward RESA.

Table 1: Defuzzified Values Obtained by Collapsing the Symmetrical Horizontal Test Set.

Degree of Discretisation	Defuzzified Value	Collapsing Forward	Collapsing Backward	Collapsing Inward	Collapsing Outward
0.1	0.5	0.5038320922	0.4961679078	0.4993086838	0.4995891494
0.05	0.5	0.5019998917	0.4980001083	0.4998049953	0.4998891777
0.02	0.5	0.5008177226	0.4991822774	0.4999665227	0.4999815068
0.01	0.5	0.5004115350	0.4995884650	0.4999914377	0.4999953154
0.005	0.5	0.5002064040	0.4997935960	0.4999978353	0.4999988213
0.002	0.5	0.5000827100	0.4999172900	0.4999996513	0.4999998107
0.001	0.5	0.5000413793	0.4999586207	0.4999999126	0.4999999526
0.0001	0.5	0.5000041401	0.4999958599	0.4999999991	0.4999999995
0.00001	0.5	0.5000004140	0.4999995860	0.5000000000	0.5000000000

Table 2: Errors Incurred in Collapsing the Symmetrical Horizontal Test Set.

Degree of Discretisation	Defuzzified Value	Collapsing Forward	Collapsing Backward	Collapsing Inward	Collapsing Outward
0.1	0.5	0.0038320922	-0.0038320922	-0.0006913162	-0.0004108506
0.05	0.5	0.0019998917	-0.0019998917	-0.0001950047	-0.0001108223
0.02	0.5	0.0008177226	-0.0008177226	-0.0000334773	-0.0000184932
0.01	0.5	0.0004115350	-0.0004115350	-0.0000085623	-0.0000046846
0.005	0.5	0.0002064040	-0.0002064040	-0.0000021647	-0.0000011787
0.002	0.5	0.0000827100	-0.0000827100	-0.0000003487	-0.0000001893
0.001	0.5	0.0000413793	-0.0000413793	-0.0000000874	-0.0000000474
0.0001	0.5	0.0000041401	-0.0000041401	-0.0000000009	-0.0000000005
0.00001	0.5	0.0000004140	-0.0000004140	0.0000000000	0.0000000000

ISBN: 978-989-95079-6-8

Table 3: Defuzzified Values Obtained by Collapsing the Symmetrical Triangular Test Set.

Degree of Discretisation	Defuzzified Value	Collapsing Forward	Collapsing Backward	Collapsing Inward	Collapsing Outward
0.1	0.4	0.4001359091	0.3998640909	0.4001131909	0.3998916916
0.05	0.4	0.4000597189	0.3999402811	0.4000498280	0.3999505451
0.02	0.4	0.4000230806	0.3999769194	0.4000195457	0.3999808751
0.01	0.4	0.4000115326	0.3999884674	0.4000098170	0.3999904744
0.005	0.4	0.4000057773	0.3999942227	0.4000049299	0.3999952381
0.002	0.4	0.4000023153	0.3999976847	0.4000019784	0.3999980943
0.001	0.4	0.4000011585	0.3999988415	0.4000009904	0.3999990469
0.0001	0.4	0.4000001159	0.3999998841	0.4000000992	0.3999999047
0.00001	0.4	0.4000000116	0.3999999884	0.4000000099	0.3999999905

Table 4: Errors Incurred in Collapsing the Symmetrical Triangular Test Set.

Degree of Discretisation	Defuzzified Value	Collapsing Forward	Collapsing Backward	Collapsing Inward	Collapsing Outward
0.1	0.4	0.0001359091	-0.0001359091	0.0001131909	-0.0001083084
0.05	0.4	0.0000597189	-0.0000597189	0.0000498280	-0.0000494549
0.02	0.4	0.0000230806	-0.0000230806	0.0000195457	-0.0000191249
0.01	0.4	0.0000115326	-0.0000115326	0.0000098170	-0.0000095256
0.005	0.4	0.0000057773	-0.0000057773	0.0000049299	-0.0000047619
0.002	0.4	0.0000023153	-0.0000023153	0.0000019784	-0.0000019057
0.001	0.4	0.0000011585	-0.0000011585	0.0000009904	-0.0000009531
0.0001	0.4	0.0000001159	-0.0000001159	0.0000000992	-0.0000000953
0.00001	0.4	0.0000000116	-0.0000000116	0.0000000099	-0.0000000095

Table 5: Defuzzified Values Obtained by Collapsing the Gaussian Test Set.

Degree of Discretisation	Defuzzified Value (EM)	Collapsing Forward	Collapsing Backward	Collapsing Inward	Collapsing Outward	Collapsing Forward-Backward
0.5	0.2899142309	0.2947300898	0.4090097593	0.2940174555	0.2884666838	0.3518699246
0.25	0.2906756945	0.2925398791	0.3712146394	0.2923170555	0.2900969651	0.3318772592
0.125	0.3043413255	0.3052741624	0.3526142975	0.3051864643	0.3041285835	0.3289442299
0.1	0.3074987724	0.3082433183	0.3477346996	0.3081777251	0.3073450280	0.3279890090
0.0625	0.3125118626	0.3129728510	0.3393585073	0.3129362993	0.3124323840	0.3261656791
0.05	0.3142610070	0.3146278507	0.3362363800	0.3145998182	0.3142020426	0.3254321154

Table 6: Errors Incurred in Collapsing the Gaussian Test Set.

Degree of Discretisation	Defuzzified Value (EM)	Collapsing Forward	Collapsing Backward	Collapsing Inward	Collapsing Outward	Collapsing Forward-Backward
0.5	0.2899142309	0.0048158589	0.1190955284	0.0041032246	-0.0014475471	0.0619556937
0.25	0.2906756945	0.0018641846	0.0805389449	0.0016413610	-0.0005787294	0.0412015647
0.125	0.3043413255	0.0009328369	0.0482729720	0.0008451388	-0.0002127420	0.0246029044
0.1	0.3074987724	0.0007445459	0.0402359272	0.0006789527	-0.0001537444	0.0204902366
0.0625	0.3125118626	0.0004609884	0.0268466447	0.0004244367	-0.0000794786	0.0136538165
0.05	0.3142610070	0.0003668437	0.0219753730	0.0003388112	-0.0000589644	0.0111711084

ISBN: 978-989-95079-6-8

Table 7: Defuzzified Values and Errors Obtained for the Symmetrical Horizontal Test Set, Collapsed Outward.

		Collapsing Defuzzified Values			Errors		
Degree of Disc.	Defuzzified Value	Outward Right	Outward Left	Outward Right-Left	Outward Right	Outward Left	Outward Right-Left
0.1	0.5	0.4995891494	0.5004108506	0.5000000000	-0.0004108506	0.0004108506	0.0000000000
0.05	0.5	0.4998891777	0.5001108223	0.5000000000	-0.0001108223	0.0001108223	0.0000000000
0.02	0.5	0.4999815068	0.5000184932	0.5000000000	-0.0000184932	0.0000184932	0.0000000000
0.01	0.5	0.4999953154	0.5000046846	0.5000000000	-0.0000046846	0.0000046846	0.0000000000
0.005	0.5	0.4999988213	0.5000011787	0.5000000000	-0.0000011787	0.0000011787	0.0000000000
0.002	0.5	0.4999998107	0.5000001893	0.5000000000	-0.0000001893	0.0000001893	0.0000000000
0.001	0.5	0.4999999526	0.5000000474	0.5000000000	-0.0000000474	0.0000000474	0.0000000000
0.0001	0.5	0.4999999995	0.5000000005	0.5000000000	-0.0000000005	0.0000000005	0.0000000000
0.00001	0.5	0.5000000000	0.5000000000	0.5000000000	0.0000000000	0.0000000000	0.0000000000

Table 8: Defuzzified Values and Errors Obtained for the Symmetrical Triangular Test Set, Collapsed Outward.

		Collapsing Defuzzified Values			Errors		
Degree of Disc.	Defuzzified Value	Outward Right	Outward Left	Outward Right-Left	Outward Right	Outward Left	Outward Right-Left
0.1	0.4	0.3998916916	0.3998916916	0.3998916916	-0.0001083084	-0.0001083084	-0.0001083084
0.05	0.4	0.3999505451	0.3999522908	0.3999514180	-0.0000494549	-0.0000477092	-0.0000485820
0.02	0.4	0.3999808751	0.3999812363	0.3999810557	-0.0000191249	-0.0000187637	-0.0000189443
0.01	0.4	0.3999904744	0.3999905679	0.3999905212	-0.0000095256	-0.0000094321	-0.0000094788
0.005	0.4	0.3999952381	0.3999952617	0.3999966681	-0.0000047619	-0.0000047383	-0.0000047501
0.002	0.4	0.3999980943	0.3999980981	0.3999980962	-0.0000019057	-0.0000019019	-0.0000019038
0.001	0.4	0.3999990469	0.3999990479	0.3999990474	-0.0000009531	-0.0000009521	-0.0000009526
0.0001	0.4	0.3999999047	0.3999999047	0.3999999047	-0.0000000953	-0.0000000953	-0.0000000953
0.00001	0.4	0.3999999905	0.3999999905	0.3999999905	-0.0000000095	-0.0000000095	-0.0000000095

Table 9: Defuzzified Values and Errors Obtained for the Gaussian Test Set, Collapsed Outward.

		Collapsing Defuzzified Values			Errors		
Degree of Disc.	Defuzzified Value	Outward Right	Outward Left	Outward Right-Left	Outward Right	Outward Left	Outward Right-Left
0.5	0.2899142309	0.2884666838	0.2890645675	0.2887656257	-0.0014475471	-0.0008496634	-0.0011486052
0.25	0.2906756945	0.2900969651	0.2902918203	0.2901943927	-0.0005787294	-0.0003838742	-0.0004813018
0.125	0.3043413255	0.3041285835	0.3041906758	0.3041285835	-0.0002127420	-0.0001506497	-0.0001816959
0.1	0.3074987724	0.3073450280	0.3073862653	0.3073656467	-0.0001537444	-0.0001125071	-0.0001331257
0.0625	0.3125118626	0.3124323840	0.3124493111	0.3124408476	-0.0000794786	-0.0000625515	-0.0000710150
0.05	0.3142610070	0.3142020426	0.3142130418	0.3142075422	-0.0000589644	-0.0000479652	-0.0000534648

ISBN: 978-989-95079-6-8

Data Integration, Approximate Categorisation and Fuzzy Associations

Trevor Martin[1,2] Yun Shen[1]

1 Artificial Intelligence Group, Department of Engineering Mathematics,
University of Bristol, University Walk, Bristol BS8 1TR, UK
2 BT Intelligent Systems Lab, BT Innovate, Orion Building
Adastral Park, Ipswich IP5 3RE, UK

Email: firstName.lastName@bristol.ac.uk

Abstract—*The use of hierarchical taxonomies to organise information (or sets of objects) is essential to the semantic web and is also fundamental to many aspects of web 2.0. In most cases, the seemingly crisp granulation of a taxonomy disguises the fact that categories are based on loosely defined concepts which are better modelled by allowing graded membership. Fuzzy categories may also arise when integrating information from multiple sources which do not conform to precisely the same taxonomy definitions. Knowledge of relations between categories can be summarised by association rules. In this paper, we outline a new method to calculate fuzzy confidences for association rules between fuzzy categories from different hierarchies. We illustrate with examples drawn from a system that integrates information from web-based sources.*

Keywords— fuzzy association rules, fuzzy data mining, fuzzy hierarchies, soft semantic web, mass assignment.

1 Introduction

The semantic web [1] is a co-ordinated attempt to develop "*common formats for integration and combination of data drawn from diverse sources*"* whilst Web2.0 is an emergent trend involving changes in web use, including user participation, content sharing, collaboration, etc. Both rely heavily on hierarchical taxonomies as a fundamental mechanism for information organisation. In the case of the semantic web, ontologies provide a formal definition of terms, relations, constraints, etc. as well as their correspondence to objects in the real world. For example, an ontology for wine retailers would define terms such as grape type, vintage, region, descriptions of flavour, etc. Web documents using these terms can be processed by machines, knowing that the "meaning" is defined by the ontology. The use of "tags" from the ontology enables automated reasoning as well as assisting human understanding.

In the case of Web 2.0, tagging and categorisation also play a central role. Photo and video sharing sites, wikis, blogs, and large parts of e-commerce rely on hierarchical categorisation of content. This may be implicit rather than explicit - a tag such as *sport* can encompass *athletics*, *swimming*, *football*, etc - but the hierarchical nature is unarguable. Rather than relying on formal, agreed ontologies, the meaning of a tag is essentially defined by the user who adds the tag, and a degree of shared understanding is necessary for the system to work.

The ability to categorise and summarise data is a key feature of human intelligence, enabling us to group multiple entities together into an (approximately) uniform whole and represent / reason about the group as a single concept. Many familiar ways to access information such as books, libraries, computer file structures, the web and other networks provide evidence that hierarchical categorisation is an efficient way to organise and access information.

There is rarely a single unique hierarchy - instead, a person chooses the most appropriate way to split up data according to their expertise / background knowledge and the problem at hand. For example, Fig 1 shows two possible divisions of customers. Here, as in most real-world hierarchies, groups of entities (or conceptual categories) are loosely defined, able to admit elements according to some scale of membership rather than according to an absolute yes/no test - indicating that fuzzy set theory [2] is appropriate.

An association exists when the extensions of two concepts overlap significantly, as indicated in Fig 1 by the directed link. Association rules (in their crisp form) are a well-established technique for knowledge discovery in databases,

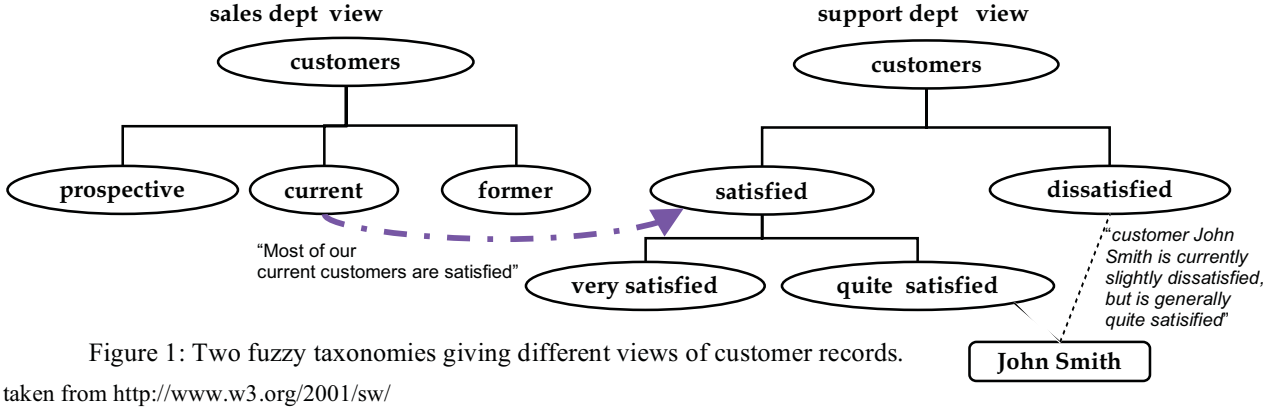


Figure 1: Two fuzzy taxonomies giving different views of customer records.

* taken from http://www.w3.org/2001/sw/

ISBN: 978-989-95079-6-8

enabling "interesting" relations to be discovered. There have been a number of proposals to develop fuzzy association rules, that is to discover the degree of association between fuzzy categories. Some of our recent work has used mass assignment theory [3-5] to find a point valued association strength between fuzzy categories [6], later extended to an interval-valued version [7]. In common with other work on fuzzy association rules, most of our previous work assumes there is a crisp value for the rule confidence.

The main contribution of this paper is a novel mass assignment-based method for calculating a *fuzzy* confidence in associations between fuzzy categories. It relies on a new method of converting fuzzy relations to mass assignments and a membership function for fuzzy confidence related to the movement of mass needed to produce that confidence value. The minimum and maximum values for the confidence can be found quickly, and memberships calculated based on the corresponding mass assignments. A full version, including proofs, will appear subsequently [8].

2 Background

The Smart Queries and Adaptive Data (SQuAD) project is concerned with adding structure (tags) to data and refining approximate knowledge from this data. It aims to assist in the extraction of useful information from diverse sources of semi-structured text data and consists of four main components:

- approximate fuzzy grammars [9] to tag small text fragments.
- SOFT (Structured Object Fusion Toolkit) [10] to help determine that entities from different sources are the same (a process also known as instance-matching, entity resolution)
- iPHI (intelligent Personal Hierarchies of Information) [11, 12] which aims to combine and integrate multiple sources of information and to configure access to the information based on an individual's personal categories.
- TRACK (Time-varying Relations in Approximately Categorised Knowledge). [7] The previous three processes leave us with a collection of objects organised into fuzzy categories, where the taxonomic structure reflects a view of the underlying data. Insight can be gained by considering relations between fuzzy categories, particularly from different category hierarchies. Of particular interest is the change over time in the degree of association between fuzzy categories - for example, in Fig 1, an executive in charge of the company would be interested to know whether the set of current customers (in the sales department's categorisation) is mostly the same as the fuzzy set of satisfied customers (in the support department's categorisation) and how the relation has changed over recent time, over the medium term and over the long term - particularly if there have been significant changes such as company practice, number/nature of competitors, etc during any of these periods.

2.1 Fuzzy Sets - the Conjunctive Interpretation

Many authors (e.g. [13]) have proposed fuzzy sets to model uncertain values in databases and knowledge based applications. The standard (disjunctive) interpretation of a fuzzy set in this context is as a possibility distribution - a single valued attribute which is not known exactly.

The conjunctive interpretation of a fuzzy set occurs when the attribute can have multiple values. These values are not repeated so it is a set, not a multi-set. For example, a person may be able to speak several languages; we could model this as a fuzzy set of languages, where membership would depend on the degree of fluency. This is formally a relation rather than a function on the underlying sets. We distinguish between the conjunctive interpretation - modelled by a monadic fuzzy relation – and the disjunctive interpretation – modelled by a possibility distribution by using the notation

$$F(a) = \{x/\mu(x) \mid x \in U\}$$

to denote a single valued attribute F of some object a (i.e. a possibility distribution over a universe U) and

$$R(a) = [x/\chi(x) \mid x \in U]$$

to denote a multi-valued attribute (relation). Fuzzy categories represent the latter case, since we have multiple values that satisfy the predicate to a greater or lesser degree.

2.2 Extending Association Rules to Fuzzy Categories

In creating association rules within transaction databases (e.g. [14], see also [15] for a clear overview), the standard approach is to consider a table in which columns correspond to items and each row is a transaction. A cell contains 1 if the item was bought, and 0 otherwise. The aim of association rule mining is to find links between disjoint subsets of items – for example, do customers generally buy biscuits and cheese when buying beer and wine? These disjoint subsets represent categories, as described earlier.

Let I denote the set of items, so that any transaction can be represented as $tr \subseteq I$, and consider X, the set of all transactions (strictly speaking, X is a multi-set but can be made into a set by adding a unique identifier to each transaction). We also specify two categories (itemsets) s and t, which are non-empty, non-overlapping subsets of I,

$$t \subset I \qquad s \subset I \quad \text{where} \quad s \cap t = \varnothing$$

and the sets of transactions containing s and t

$$S = \{x | x \in X \wedge s \subseteq x\}$$

$$T = \{x | x \in X \wedge t \subseteq x\}$$

An association rule is of the form $s \Rightarrow t$ and is interpreted as stating that when the items in s appear in a transaction, it is likely that the items in t will also appear i.e. it is not an implication in the formal logical sense. A slight abuse of notation allows us to use $S \Rightarrow T$ or $s \Rightarrow t$ as the rule.

Most authors use two measures to assess the significance of association rules. The support of a rule $s \Rightarrow t$ is the number (or relative number) of transactions in which both s and t appear, and the confidence of the rule is an estimate (based on the samples) of the conditional probability of t being contained in a transaction given that it contains s

$$Support(s,t) = |S \cap T| \qquad (1)$$

$$Conf(s,t) = \frac{|S \in T|}{|S|} \qquad (2)$$

Typically a threshold is chosen for the support, so that only frequently occurring sets of items s and t are considered; a second threshold filters out rules of low confidence.

ISBN: 978-989-95079-6-8

name	sales	salary
a	100	1000
b	80	400
c	50	800
d	20	700

good sales	*high salary*	*confidence*
sales≥80	*high≥400*	1
sales≥50	*high≥500*	0.667
sales>50	*high>500*	0.5
sales≥50	*high>800*	0.333

Figure 2 : A simple database of names (*a, b, c, d*), sales and salaries (top) and (bottom) confidences for the rule *good sales => high salary* arising from different crisp definitions of the terms *good sales* and *high salary*

For example, consider a database of sales employees, salaries and sales figures. A mining task might be to find out whether the good sales figures are achieved by the highly paid employees. Given the database table in Fig 2, we can obtain rule confidences ranging from 1/3 up to 1 by different crisp definitions of "*good sales*" and "*high salary*".. Although this is a contrived example, such sensitivity to the cut-off points adopted for crisp definitions is a good indication that a fuzzy approach is more in line with human understanding of the categories.

Various approaches to fuzzifying association rules have been proposed e.g. [15-17]. The standard extension to the fuzzy case is to treat the sets *S, T* as fuzzy and find the intersection and cardinality using a t-norm and sigma-count respectively.

$$Conf(S,T) = \frac{\in \mu_{S \cap T}(x)}{\in \mu_S(x)} \qquad (3)$$

In the example of Fig 2, a fuzzy approach would categorise employees according to simple membership functions for *good sales* (S) and *high salary* (T), which could lead to

$S = [a/1, b/0.8, c/0.5, d/0.2]$

and

$T = [a/1, b/0.4, c/0.8, d/0.7]$

and confidence 0.72 for the association $S=>T$ using (3).

As pointed out by [15], using min and the sigma count for cardinality can be unsatisfactory because it does not distinguish between several tuples with low memberships and few tuples with high memberships - for example,

$S = [x_1/1]$

$T = [x_2/1]$

leads to *Conf(S, T) = 0* but

$S_1 = [x_1/1, x_2/0.01, x_3/0.01, \ldots, x_{1000}/0.01]$

$T_1 = [x_1/0.01, x_2/1, x_3/0.01, \ldots, x_{1000}/0.01]$

leads to

$$Conf(S_1,T_1) = \frac{1000 \in 0.01}{1 + 999 \in 0.01} \in 0.91$$

which is extremely high for two almost disjoint sets (this example originally appeared in [18]). Using a fuzzy cardinality (i.e. a fuzzy set over the possible cardinality values) is also potentially problematic since the result is a possibility distribution over rational numbers, and the extension principle [19] gives a wider bound than it should, due to neglect of interactions between the numerator and denominator in Eq 2. For example, given

$S = [x1/1, x2/0.8]$

$T = [x1/1, x2/0.4].$

the fuzzy cardinalities are

$|S \cap T| = \{1/1, 2/0.4\}$,

$|S| = \{1/1, 2/0.8\}$

leading (by the extension principle) to a confidence of

$\{0.5/0.8, 1/1, 2/0.4\}$,

clearly incorrect as the confidence cannot be greater than 1.

2.3 Fuzzy Relations as Mass Assignments

As with previous work [6, 7] we start from the fact that a relation represents a conjunctive set of ordered *n*-tuples i.e. a conjunction of *n* ground clauses. For example, if *U* is the set of dice values then we could define a (crisp) predicate *differBy4or5* on $U \times U$ as the set of pairs

[(1,6), (1,5), (2,6), (5,1), (6,1), (6, 2)]

This is a conjunctive set, in that each pair satisfies the predicate. In a similar way, a fuzzy relation represents a set of *n*-tuples that satisfy a predicate to some degree. Thus *differByLargeAmount* could be represented by

[(1,6)/1, (1,5)/0.6, (2,6)/0.6, (5,1)/0.6, (6,1)/1, (6,2)/0.6]

The interpretation is not that a single pair satisfies this predicate, but that one set of pairs satisfies it (out of several possible sets of pairs).

A mass assignment [3-5] on a universe *U* is a distribution over the power set of *U*. Here, the mass assignment is on possible *relations* :

$R_1 = [(1,6),(6,1)]$

$R_2 = [(1,6),(1,5),(2,6),(5,1),(6,1),(6,2)]$

$m = \{\{R_1\} : 0.4, \{R_1, R_2\} : 0.6\}$

This is equivalent to treating the fuzzy relation as a fuzzy set of crisp relations:

$differByLargeAmount = \{R1/1, R2/0.6\}$

Similarly, a monadic fuzzy predicate *largeValue* defines a set of 1-tuples such as [6/1, 5/0.8, 4/0.3] which is written as a fuzzy set of crisp monadic relations:

$largeValue = \{[6]/1, [6,5]/0.8, [6,5,4]/0.3\}$

and has the mass assignment

$m_{largeValue} = \{\{[6]\} : 0.2, \{[6],[6,5]\} : 0.5, \{[6],[6,5],[6,5,4]\} : 0.3\}$

Our subsequent studies [7] show that this approach can sometimes overestimate the difference between full and nearly-full membership, which can lead to unreasonably large intervals calculated for the confidence of association rules. For example, under this interpretation, the monadic fuzzy relation $S = [a/1\ b/0.98]$ has the mass assignment

$m_S = \{\{[a]\} : 0.02, \{[a],[a,b]\} : 0.98\}$

ISBN: 978-989-95079-6-8

The normal mass assignment interpretation allows us to redistribute the mass on {*[a], [a,b]*} to either of the relations *[a]* or *[a,b]* which leads to the family of distributions:

$S = [a] : 1\text{-}x \quad , \quad [a,b] : x \quad$ where $0 \le x \le 0.98$

This flexibility in re-assigning mass means that for a source

$S = [a/1\ b/0.98]$ and target relation $T = [a/1\ b/0.98\ c/0.02]$

we get an interval [0.51, 1] which is surprisingly wide considering the two relations are so similar. NB this behaviour arises mostly in contrived cases, smaller intervals are calculated in the vast majority of "real" association rules that have arisen in our experimental studies such as [7].

2.4 Alternative Interpretation of Relations as Mass Assignments

The approach discussed above, which we will refer to as an open world approach, treats partial membership of a tuple x in a relation R (i.e. $0 < \cap_R(x) < 1$) as an upper bound for the mass that can be assigned to any set of tuples including x. This leads to a wide range of mass distributions that can be derived from the fuzzy relation R.

In the open world approach, for any tuple x where $\cap_R(x)<1$, the total mass that can be assigned to relations containing x is given by

$$0 \le \sum_{\substack{t=[x1,\ldots,xn] \\ x \in t}} m(t) \le \chi_R(x)$$

In the *largeValue* example above, consider the element $x=5$, which has $\cap_R(x) = 0.8$; the relations containing 5 are [5, 6] and [4, 5, 6] and we have

$$0 \le m_{largeValue}([5,6]) + m_{largeValue}([4,5,6]) \le 0.8$$

This gives a considerable degree of flexibility in assigning mass. The alternative interpretation used here - the closed world approach - regards partial membership of a tuple x in a relation R (i.e. $0 < \cap_R(x) < 1$) as *strictly equal* to the total mass assigned to the sets of tuples which include x, i.e.

$$\sum_{\substack{t=[x1,\ldots,xn] \\ x \in t}} m(t) = \cap_R(x)$$

This means there is no flexibility in the range of mass distributions that can be derived from the fuzzy relation R. However, there is flexibility in the mass assignments when R is combined with an assignment corresponding to another relation, for example in calculating association confidences as described later.

Under this interpretation, the monadic fuzzy relation *largeValue* discussed above has

$$m_{largeValue} = \{[6]:0.2,\ [6,5]:0.5,\ [6,5,4]:0.3\}$$

and clearly

$$m_{largeValue}([5,6]) + m_{largeValue}([4,5,6]) = 0.8$$

2.5 Closed world Mass-based Association Rules

For a source category

$$S = \left[x_1/\chi_S(x_1),\ x_2/\chi_S(x_2),\ldots,\ x_{|S|}/\chi_S(x_{|S|})\right]$$

and a target category

$$T = \left[x_1/\chi_T(x_1),\ x_2/\chi_T(x_2),\ldots,\ x_{|T|}/\chi_T(x_{|T|})\right]$$

we can define the corresponding mass assignments as follows. Let the set of distinct memberships in S be

$$\Lambda_S = \left\{\chi_S^{(1)}, \chi_S^{(2)}, \ldots, \chi_S^{(n_S)}\right\}$$

where

$$\chi_S^{(1)} > \chi_S^{(2)} > \ldots > \chi_S^{(n_S)}$$

and $n_S \le |S|$

Let

$$S_i = \left\{\left[x \,\middle|\, \chi_S(x) \ge \chi_S^{(i)}\right]\right\}$$

Then the mass assignment corresponding to S is

$$\left\{S_i : m_S(S_i)\right\}, \quad 1 \le i \le n_S$$

where

$$m_S(S_k) = \chi_S^{(k)} - \chi_S^{(k+1)} \qquad (7)$$

and we define

$$\chi_S^{(i)} = 0 \quad if \quad i > n_S$$

For example, the fuzzy category

$S = [a/1,\ b/0.8,\ c/0.5,\ d/0.2]$

has the corresponding mass assignment

$$M_S = \{[a]:0.2,\quad [a,b]:0.3,\quad [a,b,c]:0.3,\quad [a,b,c,d]:0.2\}$$

We can now calculate the confidence in the association between the categories S and T using mass assignment theory. In general, this is not a unique value as we are free to move mass (consistently) between the cells corresponding to S_i and T_j for each i, j.

For two mass assignments

$$M_S = \left\{S_i : m_S(S_i)\right\}, \quad 1 \le i \le n_S$$

$$M_T = \left\{T_j : m_T(T_j)\right\}, \quad 1 \le j \le n_T$$

the composite mass assignment is

$$M = M_S \oplus M_T = \left\{X : m(X)\right\}$$

where m is specified by the composite mass allocation function, subject to

$$\sum_{j=1}^{n_T} m_{ij} = m_S(S_i) \qquad \sum_{i=1}^{n_S} m_{ij} = m_T(T_j)$$

This can be visualised using a mass tableau (see [3]) as shown in Fig 3. Each row (column) represents a relation of the source (target) mass assignment. We label the rows $S_1, S_2, \ldots S_{nS}$ and columns $T_1, T_2, \ldots T_{nT}$, and assign mass m_{ij} to cell *(i, j)* subject to row and column constraints. The confidence in the association rule is given by

$$conf(M) = \frac{\sum_{i,j}\left(m_{ij} \times |S_i \cap T_j|\right)}{\sum_{i=1}^{n_S}\left(\sum_{j}^{n_T} m_{ij} \times |S_i|\right)} = \frac{n}{d}$$

$$where \quad n = \sum_{i,j}\left(m_{ij} \times |S_i \cap T_j|\right)$$

$$d = \sum_{i=1}^{n_S}\left(\sum_{j}^{n_T} m_{ij} \times |S_i|\right) = \sum_{i=1}^{n_S} |S_i| \times m_S^{(i)} \qquad (7)$$

ISBN: 978-989-95079-6-8

Clearly $n \geq 0$, $d > 0$ and d is a constant for a given source relation S, irrespective of T and M.
For example consider the fuzzy categories

$S = [a/1, b/0.8, c/0.5, d/0.2]$ and

$T = [a/1, b/0.4, c/0.8, d/0.7]$

One notable assignment is the least prejudiced distribution, obtained by taking the product of source and target masses for each cell as shown in Fig 3. This corresponds to the minimum entropy combination of the two mass assignments.

		0.2 [a]	0.1 [ac]	0.3 [acd]	0.4 [abcd]
0.2	[a]	[a] 0.04	[a] 0.02	[a] 0.06	[a] 0.08
0.3	[ab]	[a] 0.06	[a] 0.03	[a] 0.09	[ab] 0.12
0.3	[abc]	[a] 0.06	[ac] 0.03	[ac] 0.09	[abc] 0.12
0.2	[abcd]	[a] 0.04	[ac] 0.02	[acd] 0.06	[abcd] 0.08

Figure 3: mass tableau, showing intersections $S_i \in T_j$ and the least prejudiced mass distribution. The corresponding point valued rule confidence is 1.86 / 2.5 = 0.744

It is possible to show [8] that maximum rule confidence is obtained by moving mass towards the diagonal (top left to bottom right) and that minimum confidence is obtained by moving mass towards the bottom left - top right diagonal, as illustrated in figs 4 and 5.

		0.2 [a]	0.1 [ac]	0.3 [acd]	0.4 [abcd]
0.2	[a]	[a] 0.2	[a] 0	[a] 0	[a] 0
0.3	[ab]	[a] 0	[a] 0.1	[a] 0.2	[ab] 0
0.3	[abc]	[a] 0	[ac] 0	[ac] 0.1	[abc] 0.2
0.2	[abcd]	[a] 0	[ac] 0	[acd] 0	[abcd] 0.2

Figure 4 : mass tableau, showing intersections $S_i \in T_j$ and the mass distribution leading to minimum rule confidence 2.1 / 2.5 = 0.84

		0.2 [a]	0.1 [ac]	0.3 [acd]	0.4 [abcd]
0.2	[a]	[a] 0	[a] 0	[a] 0	[a] 0.2
0.3	[ab]	[a] 0	[a] 0.1	[a] 0.1	[ab] 0.2
0.3	[abc]	[a] 0	[ac] 0.1	[ac] 0.2	[abc] 0
0.2	[abcd]	[a] 0.2	[ac] 0	[acd] 0	[abcd] 0

Figure 5 : mass tableau, showing intersections $S_i \in T_j$ and the mass distribution leading to minimum rule confidence 1.5 / 2.5 = 0.6

2.6 Membership Function for Fuzzy Confidence

We define the membership function in terms of the mass which must be moved (relative to the least prejudiced distribution, where confidence has membership 1). Any other assignment of mass requires one or more elementary mass transfers relative to the LPD, and we are particularly interested in mass assignments corresponding to minimum and maximum confidence, M^{MIN} and M^{MAX}. We define a fuzzy interval C representing the confidence such that

$$\mu_C(conf(M)) = 1 - \frac{pos(M^{LPD} - M)}{N}$$

where $$N = \max\left(pos(M^{\max} - M^{LPD}), pos(M^{\min} - M^{LPD})\right)$$

Because the membership varies linearly with the amount of mass moved, it is triangular and can be calculated quickly by considering the end points. We note that it is possible for the membership function to be discontinuous at one end (i.e. to drop abruptly to zero). Fig 6 shows the membership function for the fuzzy confidence in the *good sales - high salary* example, with the calculations shown in figs 3, 4, and 5.

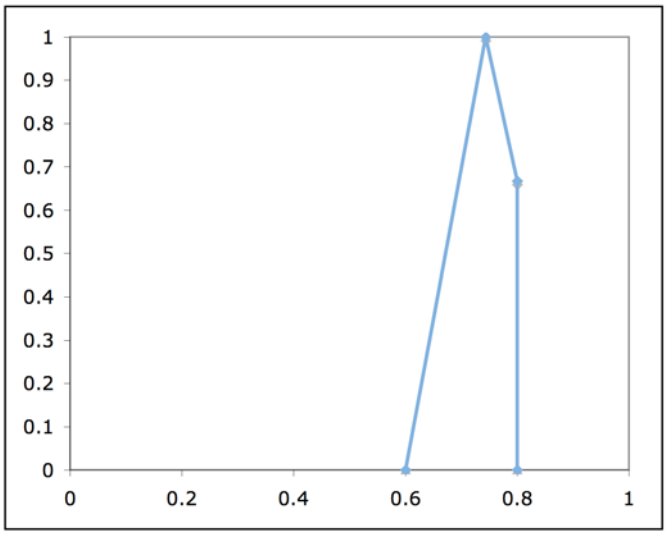


Figure 6 : Membership function for fuzzy confidence in the good sales - high salary example. Note the discontinuity at 0.8

3 Demonstrator Application

We have applied the algorithm to calculation of associations in an integrated taxonomic database of terrorist incidents, as described in [7]. Our primary source is the Worldwide Incidents Tracking System (WITS) [20] augmented by additional information from the MIPT Terrorism Knowledge Base (TKB) database[†]. WITS consists of incidents in which *"subnational or clandestine groups or individuals deliberately or recklessly attacked civilians or noncombatants (including military personnel and assets outside war zones and war-like settings)"*. We are not concerned with the correctness or otherwise of this definition, and treat the data simply as a testbed for SQuAD .

Once data has been integrated and categorised into the iPHI hierarchies, the methods outlined in this paper and [20]

[†] TKB ceased operation on 31 March 2008 and is now part of Global Terrorism Database (GTD) at www.start.umd.edu/data/gtd/

ISBN: 978-989-95079-6-8

have been applied to find significant associations between categories in different hierarchies (e.g. *geographic region*, *weapon type, perpetrator, casualty level* etc). Since the data is not static, we cannot assume that significant associations remain significant – indeed, valuable insight arises from changes in association levels relative to other associations, and trends in the strength of an association. 15,900 incidents between January 2005 and January 2006 were analysed.

Various associations can be extracted by consideration of fuzzy categories in different taxonomies. Although the vast majority of results from [7] led to reasonable intervals, there were a few cases in which intervals were quite large. As can be seen from Fig 7, much smaller intervals were obtained under the closed world method. The interval for closed world rules indicates the extremes of the fuzzy association confidence, with a symmetric triangular membership function (possibly truncated on one side). The interval for open world calculations has membership 1 throughout.

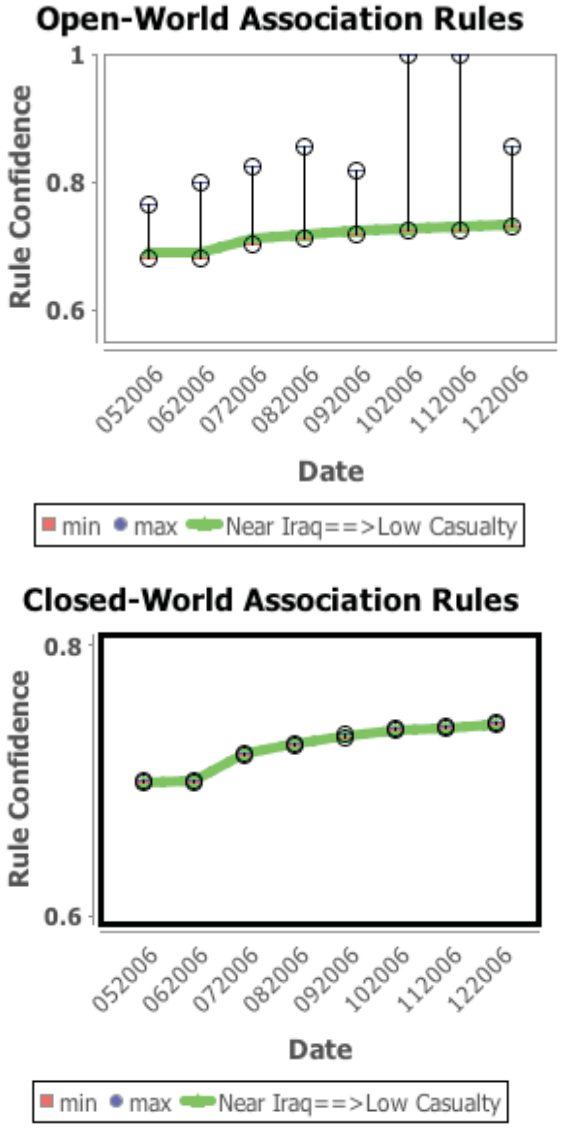


Figure 7 : Confidence intervals and point values for associations between the fuzzy categories *in/near Iraq* and *Medium casualty levels*. The fuzzy confidence (closed world) gives a much tighter interval than the open world calculation

4 Summary and Future Work

We have described a new method for calculating the confidence in an association rule between two (or more) fuzzy categories, based on mass assignment theory and yielding a fuzzy interval within which the confidence must lie. Combined with the point value confidence described in previous work, this enables us to order and plot association strengths with an indication of the degree of uncertainty.

Acknowledgment

This work was partly funded by BT and the Defence Technology Centre for Data and Information Fusion.

References

[1] Berners-Lee, T., J. Hendler, and O. Lassila, "The Semantic Web," in *Scientific American*, vol. 284, 2001, pp. 28-37.

[2] Zadeh, L. A., "Fuzzy Sets," *Inf. and Control*, 8, pp. 338-353, 1965.

[3] Baldwin, J. F., "Management of Fuzzy and Probabilistic Uncertainties for KB Systems," in *Encyclopedia of AI*, S. A. Shapiro, Ed., 2nd ed: John Wiley, 1992, pp. 528-537.

[4] Baldwin, J. F., "Mass Assignments and Fuzzy Sets for Fuzzy Databases," in *Adv. in the Shafer Dempster Theory of Evidence*, M. Fedrizzi, J. Kacprzyk, and R. R. Yager, Eds.: Wiley, 1994.

[5] Baldwin, J. F., T. P. Martin, and B. W. Pilsworth, *FRIL - Fuzzy and Evidential Reasoning in AI*. Wiley, 1995.

[6] Martin, T. P., B. Azvine, and Y. Shen, "A Mass Assignment Approach to Granular Association Rules for Multiple Taxonomies," presented at Uncertain Reasoning in the Semantic Web / Semantic Web 2007, 2007.

[7] Martin, T. P. and Y. Shen, "TRACK - Time-varying Relations in Approximately Categorised Knowledge," *Int. J. Comp Intelligence Research*, (to appear), 2009.

[8] Martin, T. P. and Y. Shen, "Fuzzy Association Rules to Summarise Multiple Taxonomies," in *Fuzziness and Scalability*, A. Laurent and Lesot. M-J, Eds.: 2009.

[9] Martin, T. P., Y. Shen, and B. Azvine, "Incremental Evolution of Fuzzy Grammar Fragments to Enhance Instance Matching and Text Mining," *IEEE Transactions on Fuzzy Systems*, 2008.

[10] Martin, T. P. and B. Azvine, "Soft Integration of Information with Semantic Gaps," in *Fuzzy Logic and the Semantic Web*, E. Sanchez, Ed.: Elsevier, 2005.

[11] Martin, T. P. and B. Azvine, "Acquisition of Soft Taxonomies for Intelligent Personal Hierarchies and the Soft Semantic Web," *BT Tech Jl*, 21, pp. 113-122, 2003.

[12] Martin, T. P., B. Azvine, and Y. Shen, "Intelligent Hierarchy Mapping: A Soft Computing Approach," *Information Tech and Intelligent Computing*, 2008.

[13] Bosc, P. and B. Bouchon-Meunier, "Databases and Fuzziness - Introduction," *Int J Intel Systems*, 9, 419, 1994.

[14] Agrawal, R. and R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases," VLDB, 1994.

[15] Dubois, D., E. Hullermeier, and H. Prade, "A systematic approach to the assessment of fuzzy association rules," *Data Mining and Knowl Discovery*, 13, 167-192, 2006.

[16] Bosc, P. and O. Pivert, "On Some Fuzzy Extensions of Association Rules," proc IFSA , 2001.

[17] Kacprzyk, J. and S.Zadrozny, "Linguistic Summarization of Data Sets Using Association Rules," proc Fuzz-IEEE, 2003.

[18] Martin-Bautista, M.J., M.A.Vila, H.L. Larsen, D. Sanchez, "Measuring Effectiveness in Fuzzy Information Retrieval," Flexible Query Answering Systems (FQAS), 2000.

[19] Zadeh, L. A., "The Concept of a Linguistic Variable and its Application to Approximate Reasoning (Part 1)," *Information Sciences*, vol. 8, pp. 199-249, 1975.

[20] WITS, "WITS - Worldwide Incidents Tracking System," National Counterterrorism Center Office of the Director of National Intelligence 2007.

ISBN: 978-989-95079-6-8

Around fuzziness. Some philosophical thoughts

Alejandro Sobrino M. Pereira-Fariña

Faculty of Philosophy, University of Santiago de Compostela
Praza de Mazarelos, s.n., 15782 Santiago de Compostela, Spain
Email: alejandro.sobrino@usc.es, martin.pereira@rai.usc.es

Abstract— *Fuzzy logic has a considerable success in the area of applications. However, this active growth in the implementation in applied electronic devices has not always been accompanied by a discussion about the theoretical grounds underlying those uses. This paper deals with some theoretical aspects that seem convenient to discuss for the purpose of clarifying the status of fuzzy logic, to get into its true nature and to appoint its achievements and limits. In order to carry out this goal, the paper is organized around three questions: 1) Fuzzy logic, is it a logic or a techno-logic?; 2) How fuzzy is fuzzy computing?; 3) Are there fuzzy objects?*

Keywords— Fuzziness, fuzzy computation, fuzzy objects, philosophy of soft computing, vagueness.

1 Introduction

Borderline cases are initially addressed by Ch. S. Peirce in 1902 [1]. In 1923, B. Russell positioned vagueness as an interesting subject in the philosophy arena [2]. Later, Hempel and Copi dealt with vagueness in separate articles published in the journal *Philosophy of Science* [3, 4]. Up till them, vague language was not considered a central topic for the analytic tradition, a school of thought characterized by an emphasis on clarity and argument, achieved by formal logic and analysis of language. M. Black introduced in 1937 [5] the 'consistency-profiles' as a way to represent the meaning of a vague predicate, defined by the standard use that a community make of it. Consistency-profiles are represented by curves, showing the degree of consistency in the application of a word to an object or concept. The notion involves three characteristics present in the modern forms to represent vagueness: the user of a language, a situation in which the user is trying to apply a symbol to an object and a way to represent the consistency of this application through a logic formalism.

Among the modern initiatives to deal with the formal representation of vagueness, understood as imprecision, the most successful has been Zadeh's fuzzy logic [6]. Fuzzy logic enables to symbolize vague propositions and to manage approximate inferences in a suitable way. As a result, it has had a tremendous growth, supported mainly by the success of their applications. However this active growth in the area of applications has not always been accompanied by a reflection on the theoretical grounds involved in the use of the underlying tool. As an applied logic it was not surprising that, in the beginning, Zadeh claimed that *"clearly the problems, the aims and the concerns of fuzzy logic are substantially different from those which animate the traditional logical systems. Thus, axiomatization, decidability, completeness, consistency, proof procedures and other issues which occupy the center of the stage in such systems are, at best, of peripheral importance in fuzzy logic"* [6, p.151]. And perhaps the same could be said of the metatheoretical discussions. In this paper we will deal with some theoretical problems that we believe that they are important for the purpose of clarifying the status of fuzzy logic, to get into its true nature and to draw its achievements and limits. The fundamental issues considered here take the form of three questions that, in turn, will name the next three sections of this work.

2 Fuzzy logic, in its true nature, is it a logic or rather than a *techno*-logic?

The reflection about the logical statute of fuzzy logic is involved with two traditional topics in Philosophy:

i. The classification of sciences.

ii. The fundamentation of sciences.

The classification of sciences is frequently revised in function of new emergent branches of knowledge, forcing to modify standard taxonomies. In the field of deductive sciences, fuzzy logic came as a new discipline that questioned some of the classical principles, prompting a debate about the nature of the logic and provoking a discussion about the status of the deviant logics, among which fuzzy logic was included [7].

Since its birth, fuzzy logic was valued in different ways by engineers (applied scientists) and by mathematicians (theoretical thinkers). Due to the success of the applications, fuzzy logic was welcomed in Engineering Schools. Nevertheless, it also caused rejection among mathematicians because, during a long time, it was a matter under construction, provisional and tentative.

Engineers use fuzzy logic according, at least, three criteria:

- Fuzzy logic allows us to represent complex scenarios that involve uncertainty using few rules. In the same vein, it permits to manage approximate inferences reaching good and quite stable solutions to the pretended conclusions.
- Although fuzzy logic is a (precise) logic about imprecision, it manages outputs that provides enough information to make good decisions. As Popper says: *"Although clarity is valuable in itself, exactness or precision is not: there can be no point in trying to be more precise than our problem demands"* [8, p.28]. So, in presence of imprecision, alternatives to the classical bivalent logic should be suggested.
- Systems with incorporated fuzzy rules are quite robust. Fuzzy logic admits a tolerance fringe between positive and negative outputs, although it is difficult to supply *a priori* guaranties about the stability of a fuzzy controller.

ISBN: 978-989-95079-6-8

These points agree with the Zadeh's *dictum*, known as 'Principle of Incompatibility': *"Stated informally, the essence of this principle is that as the complexity of a system increases, our ability to make a precise and yet significant statements about its behavior diminishes until a threshold is reached beyond which precision and significance (or relevance) become almost mutually exclusive characteristics"* [9, p. 28].

In summary, from the applied sciences point of view, as engineering, fuzzy logic is a useful discipline. Its success are the achievements of its applications and any further evidence is not requested about the logical properties of the formal tool. For engineers, fuzzy logic is a techno-logic, i.e., a logic to be implemented in electronic machines.

But beyond applications, fuzzy logic is a kind of logic and, therefore, a branch of mathematics. Mathematics is a theoretical science, not an applied one and its currency are precision, rigor and proof. So, frequently, fuzzy logic was not welcomed in mathematical circles while it was used as an *ad hoc* tool for solving quite particular problems in a large extent and specially in early applications. This partial and temporary knowledge was not well received in the field of mathematics, accustomed to dealing with eternal truths. Attending to its provisional character, mathematicians think that fuzzy logic is not a logic, but a kind of formalism for managing specific problems that involves imprecision; fuzzy logic, they say, does not manage any rationality, even not that it say to grasp: the imprecise or vague reasoning.

Attending to classical logic, the Verificability Principle allows us to substitute a proposition by its truth value. But in the field of fuzzy logic, this unanimous correspondence is problematic. A vague predicate can be represented, in a reasonable way, by similar linguistic labels or fuzzy numbers. As a matter of degree, fuzzy meaning is grasped in an imprecise way; hence, some margins of tolerance should be admitted in order to represent it. Accordingly, in fuzzy logic it is not possible to forget the world and focus only in the proofs. We need to reconcile both of them. Any attempt to give a fuzzy logic axiomatic approach hits with the absence of the stability of truth, with the lack of a perfect truth. Therefore, for mathematicians, fuzzy logic is not a true logic.

From the point of view of the fundamentation of sciences, logic is a main part of the scientific method. The hypothetico-deductive system is considered, frequently, as the scientific method *par excellence*. From Aristotle, Science is strongly related to deductive reasoning. In [10], it is defined a deductive science as a set S of sentences that satisfies four conditions:

1. *Reality postulate*: Sentences of S are about a particular subject matter, about real entities.

2. *Truth postulate*: Each sentence of S is true.

3. *Deductive postulate*: If something can be deduced from a subset of S, then it is also a member of S –closure property under logical consequence.

4. *Evidence postulate*:

   (a) *Definability*: The sentences of S are built from a finite set of terms, which have a meaning immediately obvious.

   (b) *Axiomatizability*: The truth of every sentence of S can be inferred logically from a certain finite subset S_1 of S whose truth is also obvious.

Postulates 1 and 2 require comments. It is known that both mathematics and physics are scientific theories. But while mathematics is an ideal science, physics is an experimental one. So, the objects of mathematics are not real objects, but ideal entities. Therefore, their statements can be absolutely true; timeless true. This forces to not cast doubts about the postulate 3. The simplification of the reality and the idealism related to the mathematical thinking lead to not questioning the postulate 4. But empirical statements are about real entities, and material objects change and vary over time. So, postulate 2 –the *truth postulate*– is not immediately obvious. Sometimes empirical sentences are true, sometimes are partially true (true in a degree, true with a probability,...) but they are always temporarily true; that is, true in absence of falsation. Postulate 2 is known as the Principle of Bivalence and it was traditionally rejected in other applications which are different from the logic of mathematical entities.

In twentieth century, with the birth of artificial intelligence, a new way to see deductive sciences came. The logic is not only applied to specific languages, such as the mathematical language –crisp in nature–, or to regimented natural language –specified artificially–, but also to statements that verbalize normal situations related to ordinary life. So, the *Reality Postulate* was extended to any real entity, including these with ill-defined borders and the *Principle of Bivalence* was replaced by the *Principle of Valence* [11], which states that every proposition has a truth value, not fixing how many or which. The relaxation of principles 1 and 2 casts doubt on the fourth: in real life it is difficult to think about meanings or truths immediately obvious. More attention is demanded by the third principle. Fuzzy logic is a deductive logic that grasps imprecise consequences. As a logic, it aims to conclude deductively, but as a theory of imprecise consequences, calls for margins of tolerance or imprecise intervals of plausible conclusions. Whether fuzzy logic verifies, or not, the deductive postulate depends on how the meanings of vague sentences are represented, as well as the approach to combine them into complex propositions. For example, Trillas et al. showed that the unique way to obtain consequences in the Tarski's sense from the Zadeh's compositional rule of inference is using the bold t-norm if the rule is represented by the Reichenbach's conditional definition [12].

From the beginning until now, fuzzy logic has evolved dramatically. Today it is still a formal tool with successful applications, but also a logic that seeks its theoretical consolidation. In this goal highlights –among others– the contributions made by the Spanish school of fuzzy logic, particularly those made by Esteva & Godo and Trillas. Esteva & Godo deal with fuzzy logic as a symbolic logic developed in the spirit of classical logic (syntax, semantics, axiomatization, truth-preserving deduction, completeness, etc. both in propositional and predicate logic) [13]. Hajek and Novak et al. contributed too to credit fuzzy logic in this vein [14, 15]. Hence, their work was directed to confer credit to fuzzy logic as a kind of logic. Trillas has been from a long time devoted to evolve a fuzzy metatheory of fuzzy logic; so, he has contributed to develop fuzzy logic as a fuzzy theory [12]. All of them, helped to

ISBN: 978-989-95079-6-8

credit fuzzy logic as an area of interest in the contemporary scientific panorama.

Despite these developments, fuzzy logic needs to assume new challenges. In the field of applications is a truism the success of the fuzzy controllers in many household devices. But to avoid falling in auto-complacency, it would be desirable to develop applications where vagueness or imprecision are the main component of a complex system; e.g., implementing fuzzy logic into devices in order to summarize perceptions [16]. In the theoretical perspective, fuzzy logic needs to debate about the ways to represent the meaning of a vague proposition, how to justify one choice or another among the various alternatives. Its credibility will be critically linked to this task. Another job will be to mix fuzzy logic with others logics while imprecision in not the only characteristic of human language or human decision-making. Context, time, intentionality, beliefs and its corresponding logics are jointly relevant to model human reasoning and, therefore, to build electronic devices that simulate human behavior.

The new fuzzy logic proposed by Zadeh under bright titles as 'Computing with Words and Perceptions', 'Precisiated Natural Languages', 'Protoforms' and so on [17], should make progress in this regard. Until now we have ashy headlines but low content.

3 How fuzzy is fuzzy computing?

'Turing machine' (TM) is perhaps the main reference in computer science. So, it is said that any function is computable if it is Turing computable. If a problem is computable, it admits an algorithmic definition; i.e., its solution can be reached by a finite sequence of instructions that ends in a finite time.

A TM requires that:

- Each cell of the tape contains only a symbol, although there can be an infinite number of cells.
- States are finite and discrete: the machine is in one state or another, but never in two states at the same time or in no state.

According to the discrete nature of the alphabet and the states, the computation performed by a TM is linear and precise. But although the representation and the computation of a problem are bivalent in nature, the output may be trivalent: in effect, a TM computation can conclude in three different states:

- Final state or success: The problem is decidable and it has a solution.
- Non-final state or failure: The problem is decidable but it does not have solution.
- Loop state or auto-referential state: The machine computes indefinitely and never stop. The problem is indecidable and it does not have any solution. Hence, a TM may have not solution because it stops on a not final state or because it does not stop.

There are so many TM as problems can be putted in an algorithmic form. But a nice achievement of the theory of computation is that any TM admits to be simulated by a special one: the binary TM. A binary TM works as follows: any symbol or state is represented by binary digit or a set of binary digits. So, an instruction as $10 \longrightarrow 10001l$ denote that the machine is in the internal state *1*, reads *0* in the tape, change to the state *1000*, writes *1* and it moves the tape one step to the left. A binary TM can simulate any other specific TM; so, it is an universal computing machine.

At this point, we wish to emphasize the capability of the binary TM to simulate the behavior of any other TM. But the TM not only support be simulated. It is also an excellent simulator; in particular, mimic goods solutions to problems which was not designed to. For example, several TM working simultaneously can imitate a parallel computation, although the way in which TM process is sequential and not parallel.

Fuzzy TM goes back to the late 1960s, when Zadeh spoke about the notion of fuzzy algorithm [18]. Santos gave the formal description of a fuzzy algorithm by fuzzy variants of Turing machines, Markov algorithms and finite automata [19]. He established the fundamentals of the fuzzy language too. Fuzzy TM are linked to classical TM through the support and crisp part of a fuzzy set. Fuzzy algorithms were drawn to perform imprecise computations.

What is meant by imprecise computations? At least, two answers are possible:

- In a soft sense, imprecise computations refer to problems whose working process can be expressed in an algorithm form through imprecise instructions; that is, rules that are or are not completely satisfied, but only fulfilled at some extent. From this perspective, a typical imprecise computation is a cooking recipe: add a pinch of salt, cook quite slowly, serve very cold,...
- In a hard sense, imprecise computations should refer to a genuinely form to compute vague instructions. If so, there would be a new model of computing, but up till today, the TM is the only machine for computing. Afterwards, we come back to this topic.

Imprecise computations in a soft sense can be simulated by a TM. A fuzzy instruction, as "add a pinch of salt', can be represented by a generalized membership function. Fuzzy logic provides a fuzzy arithmetic for managing fuzzy commands. So, a TM with a fuzzy arithmetic module can operate fuzzy computations and output a final state of partial success; that is, a solution qualified by a degree. The appearance of this fuzzy output may be continuous and not discrete. Discrete outputs passing serially, one after the other, in a small temporary sequence, may emerge as a continuous appearance –cartoons are a god metaphor: snapshot sequences of static pictures seem to have movement by themselves.

In a soft sense, a fuzzy computation can be performed by a classical TM, at least, in two ways:

- Weighting the symbols of the tape alphabet.
- Qualifying the states with degrees of preference.

So, from a theoretical point of view, fuzzy computation is classical computation augmented with a module that operates with degrees. "Fuzzy" refers to data high level representation not to a form of computing. Therefore, there is not a new or genuine way of computation [20]. However, from an applied

ISBN: 978-989-95079-6-8

perspective, the important thing is not to valuate the nature of the process, but its outcome. Today there is an enormous amount of computational processes that simulate approximate solutions to interesting problems. Still having the TM as a substrate, all of them have a very realist appearance, as it is showed by the software that control the smoothness of a train brake or the camera that stabilizes the frame for taking a good picture show. From this point of view, fuzzy computation is a fact regardless if the true nature of the computational process is fuzzy or not.

As we have said before, if it is possible to get good simulations of fuzzy instructions using classical methods, would still be desirable a true fuzzy computer? The answer of this question leads to discuss the differences and analogies between the real and the artificial, between the authentic and the simulated, directing, ultimately, the old philosophical question of what is 'being'. But this inquiry is out the limits of this paper.

Computing in the traditional sense involves inputs with numbers and symbols that are crisp. However, human beings use words to solve daily problems that attend to fuzzy rules; so to speak, humans use words to make computations. In the last years, some new idea has being emerging in this field. Fuzzy logic is emphasized to play a key role in granular computing, the methodological base for computing with words and perceptions. So, while in classical theory a sensor records a temperature by a number (30 C^o), in fuzzy theory a device should to convert the numerical input in the name of a perception: its hot. According to Ying [21], the start point of models for computing with words is to deal with fuzzy finite-state and fuzzy pushdown automata through extending their inputs to be strings of fuzzy subsets of input an alphabet. However, a study about the correlation between the strings or words and the symbols as inputs has not taken place.

In the context of translating human perceptions to computational perceptions and in the frame of a practical application, Triviño et al. [16] developed a methodology to generate a linguistic summary from data provided by sensors. A chain of transducers perform this task. A 'transducer' is a kind of automata capable of translating perceptions between two levels of granularity. In the project of a Linguistic Fuzzy Transducer the following actions are involved:

- Representing the granular word of the input by its associated vector of linguistic state variables.
- Representing the granular word of the output by its associated vector of linguistic state variables.
- Defining a set of rules that describes the evolution of the system in order to reach the outputs as a function of the values of the inputs.
- Generating a generic representation of the system by the methodology of Zadeh's protoform.

Besides fuzzy computation, there are other initiatives to deal with imprecision. The most remarkable is quantum computing. In quantum physics, vagueness or imprecision take the form of indetermination. But quantum indetermination and linguistic vagueness or imprecision are quite different concepts in nature. So, today quantum computing does not seem an adequate model for imprecise computing. Despite the different nature of vagueness and indetermination, the *continuum* and not deterministic nature of quantum computing encouraged to find alternative models in the general area of imprecision. Although today TM is the computational tool *par excellence*, other emerging tools, impelled by the problem of imprecision, are coming. If they can be an alternative to the classical way of computation is still an enigma, as they are in a starting point.

4 Are there fuzzy objects?

As previously noted, in 1923 Bertrand Russell published a seminal paper about vagueness. In that article, he provided a starting point to introduce the vagueness as a topic of discussion in the field of the analytic philosophy, that attempt to clarify, by analysis, the meaning of statements and concepts. After Russell, other philosophers like Hempel, Black, Dummet or Fine wrote about this subject [22].

Russell discussed the role of vagueness in the language and wrote: *"Vagueness and precision alike are characteristics which can only belong to a representation, of which language is an example. The have to do with the relation between a representation and that which it represents. Apart from representation, whether cognitive or mechanical, there can be no such thing as vagueness or precision; things are what they are, and there is an end of it"*[2, p. 2]. So, vagueness is confined to the realm of the language and excluded as a property of the objects.

Saying that there are only vague sentences and not vague objects seems to be a truism. Objects are what they are. Only the way to access them by the language is imperfect or imprecise in nature. In the vein of positivism, some philosophers thought the opposite: so, the first Wittgenstein said in the *Tractatus* that sentences, if they are to mean anything, must mirror reality in the same way that a picture does. But even in natural sciences, words picture objects in different ways: in that task, some words are more vague that others.

So, we can say unequivocally that the Everest is higher than Mont Blanc. But is it is intrinsically imprecise where the Mont Blanc starts or which area encompasses. A cloud is too a typical case of object with fuzzy boundaries. Quantum mechanics brought a new field of examples of vague objects. The 'spin' is a fundamental property of atomic nuclei, hadrons, and elementary particles. 'Having a spin direction x' is a property clearly defined and it is possible to know exactly when a particle is or not in a given spin direction. Making an experiment we can identify a number of electrons that have this property and the result obtained will be the same if the experiment is repeated. Nevertheless, we cannot notice if the particles that had the property in the first experiment are the same as the second. The particles fall or not under the extension of the predicate 'Having a spin direction x', but it is impossible to fix the identity of the resulting set.

We can refer to vague objects from three points of view:

- Objects are well defined but there is not enough information or appropriate knowledge for grasping them precisely.
- Objects are well defined but it is not possible to know their true being.
- Objects are ill defined because we cannot observe them

ISBN: 978-989-95079-6-8

directly and the instruments to watch them distort irremediably their nature.

The first two points are related to the representation of reality through language. The third refers to the essential nature of the objects. We approach them in an inverted order.

The mental experiment of Schrödinger's cat provides a good example of essential vagueness: suppose a cat into a box. Suppose you cannot see in it and the cat cannot see out. Into the box there are a rock, a Geiger counter and a vial of poison. Now, the rock is slightly radiactive, with an exactly 50% chance of emiting a subatomic particle during an hour. If the rock emits a particle, the Geiger counter activates a device that breaks the vial of poison, killing the cat. Quantum mechanics suggests that after an hour the cat is simultaneously alive and dead, so it is an indeterminate or vague object.

From a philosophical point of view, an essentialist perspective is approached by G. Evans, who address the impossibility of vague objects in the frame of the Kripke's theory of rigid designators. A term is a rigid designator if it always designates the same object in any world where the object exists. Proper nouns are rigid designators. Like this, we can imagine a world where the name 'cat' do not refer to a domestic, hairy, medium size, enemy of the mice,..., animal, but we cannot imagine that the name 'Lotfi Aliasker Zadeh' do not refer to Lotfi Aliasker Zadeh. If a word denotes always to the same object in every possible world, we can say that there are a strong correspondence between the word and the object: when we are talking about something, we can say that the words replace objects and the words have an unambiguous and precise meaning. Evans uses this preconception for showing the impossibility of vague objects.

In a short paper published in *Mind* in 1978, he introduced the question about the existence of vague objects [23]. The underlying question posed under the title: *Can there be vague objects?* This question adress the possibility of a vague identity relationship when both sides of the equal sign are occupied by terms that are rigid designators. Evans answers 'no' arguing the following proof (see table 1): Denote ∇ 'indeterminate',

Table 1: Evans' Theorem.

(1)	$\nabla(a = b)$	$a = b$ is indeterminate	(assumption)
(2)	$\lambda x[\nabla(a = x)]b$	b is indeterminely equal to a	(from 1)
(3)	$\neg\nabla(a = a)$	$a = a$ is determinate	axiom
(4)	$\neg\lambda x[\nabla(x = b)]a$	a is not indeterminely equal to b	(from 3)
(5)	$a \neq b$	a is not equal to b	(from 2 and 4)

Evans' argument shows that there are not vague entities if the terms are rigid designators: briefly, if it is indeterminate that *a* is equal to *b* and it is determinate that *a* and *a* are equals, whatever *a* be, *a* and *b* differ at least in one property; so, are different. But the Evans' argument entails a mistake: as D. Lewis point out in [24], Evans' proof start attaching vagueness to the identity relation involving rigid designators. But the own proof demonstrates that the premise is false; therefore, the starting point of the argument was inadequate. Evans' argument shows the inconvenience to judge the imprecision in a small subset of the language –that of rigid designators–, and the mistake to extend this conclusion to the overall language.

In general, language is imprecise in nature and its formal approach, a matter of degree. Furthermore, some words are essentially vague. In the sequel, we analyse the second way in which an object can be vague, according to the bottom clasification. There are objects of which we know all that it is required for knowing their meaning, for attributing them a valuation, but, in some cases, we cannot do that in a unique way. A 'Bald man' or a 'heap of sand' are some examples.

Sorites or *Falakros* is a well-known paradox since Greek philosophers that comes from the use of predicates as 'bald' or 'heap' in the argumentation. To analyse a *Sorites* argument with a classical look leads to paradoxical and unacceptable conclusions. Note that baldness can be mainly referred in terms of hairs and their distribution, that is, in terms of numerable instances. So, a man with 0 hairs is absolutely bald and a man with 10^{10} hairs, not. But counting does not ever clarify the status of a vague predicate. Adding to every man that is bald one hair does not modify its attribute and preserve bald. But the key point is how many hairs are neccesary to change a man from bald to not bald. If you choose a number, I can select another very close to the previous and to claim the truth with the same legitimacy. There is not a bald object that determines, in a unanimous way, a unique use of the word 'bald', but only different uses depending on the context. Different uses mean diverse representations, leading to variations –between some margins– in the conclusions reached by argumentations. Trillas and Urtubey proposed a reasonably cut point that divide when a person is going to be more bald that not bald [25]. But following a kind of metatheoretical game and applying a type-*sorites* reasoning to it, the solution comes irremeably affected by high-order vagueness: the objection that emerge is why not suggest a very similar, but distinct, cut point, to the one supplied. So, it seems quite difficult to solve problems of vagueness with precise measures.

Finally, there are objects well defined but there is not enough information or appropriate knowledge for grasping them precisely; that is, objects that are imprecisely verbalized. Linguistic imprecision should be analysed with fuzzy armamentarium. Fuzzy logic provides tools for managing problems that involve imprecise statements or approximate reasoning. In the field of fuzzy logic, there is not a debate about what is a vague object but how to represent words with vague meaning. From some philosophical traditions (phenomenology, e.g.), an object is how we describe it using words in context. Objects are represented by words, and sentences are represented by logic. If objects have a crisp shape, perhaps precise sentences can designate them and classical membership functions should be employed to formalize them. If the objects have loose frontiers, the way to speak about them are vague in nature and generalized membership functions should be used.

For fuzzy logic there are vague objects as there are fuzzy representations of them. There is no other consideration about its true nature. Fuzzy logic is not a physical theory, but a formal one. So, it does not pretend to reveal the essential nature of the objects, but knowing how people verbalize and represent them in a manageable way in order to solve problems. For fuzzy logic, fuzzy objects are mental or physical objects acceded by sentences uttered in a culture or context that point out the consensus required for an adequate formalization.

Fuzzy logic repairs on linguistic vagueness, not vague objects, anything they that be. Therefore, words or predicates

ISBN: 978-989-95079-6-8

that are instead of the objects are the objective of fuzzy logic. So, its function is to be faithful to how that ordinary people describe reality trough language, generating models to catch the meaning of the predicates according their use in context, generalising solutions whenever possible, making hypothesis on its explanatory character and contrasting the reached conclusions with those awaited. This is a dialectical process, typical of experimental sciences. As a formal tool, fuzzy logic must pursue eternal truths –even imprecise truths–, but like an experimental one, facts must show that truth is obtained only in a degree.

In addition to pursuing affordable challenges, as providing models for a large number of vague sentences in real scenarios, fuzzy logic must address others questions that are still elusive. One of them is how to justify and to manage a suitable specification of the context, basic for representing, in an adequate way, the meaning of vague predicates. There are some formal theories of context, but seems too complex to be manageable or too simple to account the complexity of the meaning. Other challenge is how to summarize a complex perception, where attention is frequently directed to an issue, with a unique membership function. Those are applied inquiries that, however, also require a theoretical debate.

5 Final Remarks

Philosophical discussion about fuzzy logic and its applications is convenient. The popularity of fuzzy logic is based, in a large extent, on the success of its applications. Applications show the working face of a theory, but its consolidation and credit come through the development of a meta-theory, that shed light over its foundations and limits. More work should be done in this direction.

Although vagueness is treated from fuzzy logic mostly by the side of applications, theoretical studies are useful if shed intriguing doubts or suggest illuminating challenges to applied developments. Theory without applications is lame, but applications without theory are blind.

So, to justify the election of a membership function consistent with the true context of a used word or to argue about alternatives for combining atomic sentences into a composite one are tasks that will help to approach more credible solutions to problems every time more generals; so, to transfer credit to fuzzy logic as logic.

The discussion about alternatives to classical computation and the characteristic of a genuine fuzzy TM serve to distinguish the real and the imitated behavior. Also it is interesting to emphasize the simulation power of the classical TM to get fuzzy outputs managing approximate algorithms with crisp implementations.

Finally, to consider how we represent objects permits us to focus on tools as fuzzy logic that manage the meaning of the words that verbalize not only objects, but objects in context.

Acknowledgments

The authors would like to acknowledge the valuable comments provided by the anonymous referees. Partially supported by HUM2007-66607-C04-02 grant from the Spanish Ministry of Education and Science, PEIC09-0196-3018 grant from the Autonomous Government of Castilla-La Mancha and by the Spanish Ministry of Science and Innovation (grant TIN2008-00040 and the FPU Fellow Program).

References

[1] J. M. Baldwin (ed.), editor. *Dictionary of Philosophy and Psychology*. MacMillan, 1902.

[2] B. Russell. Vagueness. *The Australasian Journal of Psychology and Philosophy*, 1:84–92, 1923.

[3] G. Hempel. Vagueness and logic. *Philosophy of Science*, 6, 1939.

[4] I. M. Copi. Borderline cases, vagueness and ambiguity. *Philosophy of Science*, 6, 1939.

[5] M. Black. Vagueness: An exercise in logical analysis. *Philosophy of Science*, 4:427–455, 1937.

[6] R.E. Bellman and L.A. Zadeh. Local and fuzzy logics. In J.M. Dunn and G. Epstein, editors, *Modern Uses of Multiple-Valued Logic*, pages 103–165. D. Reidel, 1977.

[7] S. Haack. *Deviant Logic, Fuzzy Logic: Beyond the Formalism*. The University of Chicago Press, Chicago, 1996.

[8] K. Popper. *Conjectures and Refutations: The Growth of Scientific Knowledge*. Routledge Classics, London, 1963.

[9] L. A. Zadeh. Outline of a new approach to the analysis of complex systems and decision process. *IEEE Transactions Systems, Man, and Cybernetics*, SMC-3:28–44, 1973.

[10] Aristotle. *Prior and posterior analytics*. Clarendom Press, Oxford, W. Ross edition, 1949.

[11] D. H. Sandford. Competing semantics of vagueness: many values vs. super truth. *Synthese*, 33(2-4):195–210, 1937.

[12] E. Trillas et al. *Introducción a la lógica borrosa*. Ariel, Barcelona, 1995.

[13] F. Esteva et al. On the standard and rational completeness of some axiomatic extensions of the monoidal t-norm logic. *Studia Logica*, 71:199–226, 2002.

[14] P. Hajek. *Metamathematics of fuzzy logic*. Kluwer, 1998.

[15] V. Novak, I. Perfilieva, I. Mockor, editor. *Mathematical Principles of fuzzy logic*. Kluwer, 2000.

[16] G. Triviño and A. van der Heide. Fuzzy languages and their relation to human and machine intelligence. In *Proc. of the Conf. IPMU*, 2008.

[17] L. A. Zadeh. A new frontier in computation - computation with information described in natural language. In *3rd International IEEE Conference Intelligent Systems*, pages 5–6, 2006.

[18] L. A. Zadeh. Fuzzy algorithms. *Information and Control*, 12(2):94–102, 1968.

[19] E. Santos. Fuzzy algorithms. *Information and Control*, 17:326–339, 1970.

[20] C. Moraga. Towards a fuzzy computability. *Mathware and Soft Computing*, 6:163–172, 1999.

[21] M. S. Ying. A formal model of computing with words. *IEEE Trans. Fuzzy Systems*, 10:640–652, 2002.

[22] R. Keefe and P. Smith, editors. *Vagueness: a reader*. The MIT Press, Cambridge, Massachusetts, 1999.

[23] G. Evans. Can there be *Vague* objects? *Analysis*, 38:208, 1978.

[24] D. Lewis. Vague identity: Evans misunderstood. *Analysis*, 48:128–130, 1988.

[25] E. Trillas and L. A. Urtubey. Towards the dissolution of the sorites paradox. *Applied Soft Computing (in press)*.

ISBN: 978-989-95079-6-8

Fuzziness as a Model of User Preference in Semantic Web Search

Veronika Vaneková[1] Peter Vojtáš[2]

1.Pavol Jozef Šafárik University in Košice, Slovakia
2.Charles University in Prague, Czech Republic
Email: veronika.vanekova@upjs.sk, peter.vojtas@mff.cuni.cz

Abstract— *Web 2.0 is based on personalized access to imprecise and incomplete information from heterogeneous sources. In this paper we present a web-based system enabling preference-based search. We use modified Fagin's model of fuzzy preferences based on aggregation of attribute preferences. Aggregation is generated by user ranking of objects. Our contributions are twofold. First we present a formal model and implementation suitable for efficient preferential search. Second we provide several fuzzy measures to evaluate experimental results of our system.*

Keywords— web search, fuzzy logic, fuzzy systems, user preference

1 Introduction

From its birth in 1965, fuzzy logic has spread to many areas. It has influenced logic, databases, neural networks, expert systems, natural language processing and other fields of both theoretical and practical research. Fuzzy logic allows us to model imprecise, vague or uncertain information and to specify vague requirements. Any theory or application can possibly be fuzzified. However, we should first consider the purpose, especially in case of applications. The important question is how much fuzziness we really deal with and where it comes from.

In this paper we propose one such application of fuzzy logic: we use fuzzy sets to represent a notion of user preference in the process of web search. Web search tasks are recently changing from finding a document to solving specific problem for specific user. A typical problem is to buy some item online, make a flight reservation, find a transportation line, apply for a job. Most of these problems actually require searching for objects from some *domain* (e.g. flights, books, notebooks, job offers). The search depends on various attributes of objects (e.g. price, destination, date of departure, airlines and class in the domain of flights). Fulltext search often fails to support such tasks.

Therefore we search data from one arbitrary, but fixed domain. Data is extracted from heterogeneous sources on the web and stored in a standard database or ontology. Our extraction tool does not create fuzzy data. Although the formal model can be easily extended to handle fuzzy data, we focus on the situations where the main source of fuzziness is not in the data itself, but in the imprecise way of human thinking and specifying requirements. Most web users do not know precisely what they want to find, but they are able to distinguish between better and worse objects. Hence we allow users to rate objects and we use inductive learning tool [1] to adjust user preferences. But users also have an opportunity to specify their preferences precisely by means of a special graphical interface. The best objects can be retrieved with top-k threshold algorithm (see [2, 3]).

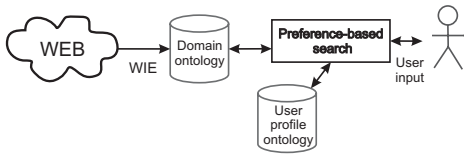


Figure 1: Overview of User-Dependent Search System.

This paper investigates how fuzzy logic can be used to make web search easier and more suitable for imprecise requirements. We compare our approach with other systems and formal models and point out the features necessary for Web 2.0 applications [4] – modeling vague preferences, dealing with huge amounts of data and ordering search results by degree of preference. We also analyze user data from our system with various measures of fuzzy similarity, equality, correlation and we compare the results.

2 Formal Models

Throughout the rest of this paper we will use the domain of job offers which is used also in the implementation. But the formal models described below are suitable for other domains as well. We tested the whole system or its parts in domain of cars, flats and investments.

Web search process is divided into two phases. The first phase is called *WIE* – web information extraction. Information is extracted from HTML pages, each of which contains data for a single object (job offer). Such pages are often generated by server side scripts, so that their structure and HTML tags are similar. The extraction tool [5] compares the page structure with *diff* algorithm and analyzes HTML tags with regular expressions. Differing parts of page sources contain relevant information, most often various attributes of domain objects. We use heuristics (like regular expression patterns for HTML structure) to avoid irrelevant information. Extracted crisp data is stored in domain ontology of job offers. Data repository is the only point that connects extraction phase with searching phase (see Fig. 1). Therefore it is easy to use other extraction tools or search in some existing standard database as well. It is also possible to have fuzzy domain data. Some extraction tools consider reliability of web sources, quality of source data or other conditions to determine fuzzy degree of extracted objects. Our formal model could be extended to consider this type of data as well. Full description of the extraction phase is beyond the scope of this paper, for more information see [6, 7, 5].

ISBN: 978-989-95079-6-8

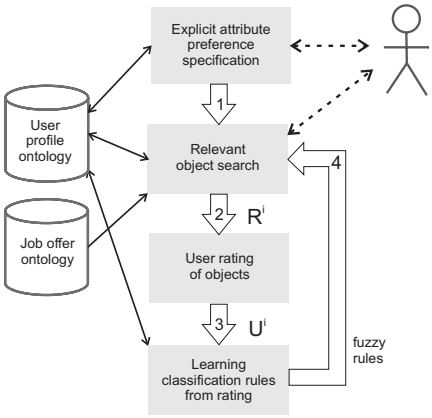


Figure 2: Flow Diagram of User-Dependent Search.

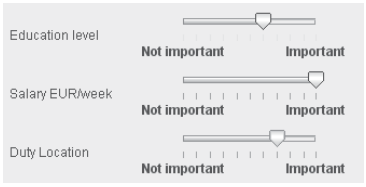


Figure 3: Interface for Attribute Importance.

The second phase, preference-based search, is detailed on Fig. 2. The task of job offer search can be repeated. The system reads user preferences (specified directly) or user feedback (previous search results rated in five-degree scale), updates stored preferences and finds new results. Every such *search cycle* provides more information about the user and helps the system to find more exact preferences.

In the rest of this section we will describe all phases of user-dependent search from Fig. 2 in more detail, together with underlying formal models. We especially emphasize the sources of fuzziness and their influence on the formal model.

2.1 Explicit Preference Specification

Current web portals often provide either a text field for fulltext search or some form to specify required attribute values. Job offer portals usually allow applicants to specify at least place, job position and contract type (i.e. permanent or temporary). But it is impossible to express things like "I prefer York to Birmingham" or "my decision depends on salary more than on location". Our system allows for vague user preferences and provides a suitable graphical interface.

First, the user is presented a list of all *attributes* from the domain of job offers that could influence his decision (like salary, benefits, required education level, required experience, place, job position, contract type, start date). The interface displays a slider marked with "Not important" and "Important" for each attribute (Fig. 3). User can drag the slider to specify how much he cares about the corresponding attribute. The settings from Fig. 3 will produce weighted average

$$@^U = \frac{6 \cdot \mu^U_{education_level} + 10 \cdot \mu^U_{salary} + 7 \cdot \mu^U_{location}}{23}$$

Weighted averages and other types of aggregation will be described below.

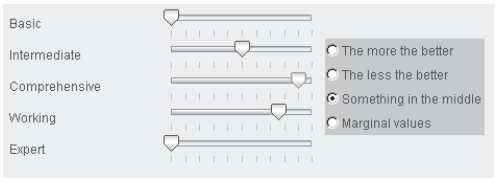


Figure 4: Interface for Attribute Preference.

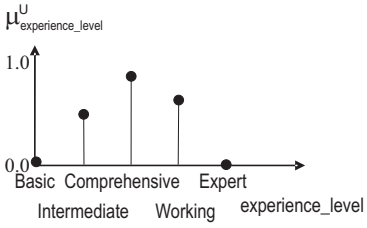


Figure 5: Preference to experience level, acquired from user input.

The next step is to specify preferred attribute values (only for attributes that are marked as "Important" to some non-zero degree). Now the graphical interface contains one slider for every attribute value. Fig. 4 shows the input interface for required experience level. The user prefers those job offers that require comprehensive experience (possibly because he has such experience in the profession).

Formally, user preference related to one particular attribute is a typical example of fuzzy subset. We consider a domain Δ of objects (job offers) and a finite set of domain attributes $\{A_1, \ldots, A_n\}$. Every attribute A_i is a function $A_i : \Delta \longrightarrow \Delta_{A_i}$, where Δ_{A_i} is a set of possible attribute values. Then *attribute preference* for user U and attribute A_i is defined as $\mu^U_{A_i} = \Delta_{A_i} \longrightarrow [0,1]$. Thus $\mu^U_{A_i}$ defines a fuzzy subset of Δ_{A_i}, which is usually a finite set with some natural element ordering.

Fuzzy set $\mu^U_{experience_level}$ generated from user input on Fig. 4 is shown on Fig. 5. Job offers that require basic or expert experience are unacceptable, while job offers with comprehensive experience are the most preferred. Every attribute preference defines non-strict linear ordering of all objects based on their membership values. The ordering generated by the fuzzy set $\mu^U_{experience_level}$ would not preserve original ordering of attribute values. Note that some distinct attribute values (e.g. basic and expert experience) have equal membership values. Such objects are *indiscernible* with respect to attribute $experience_level$.

The most suitable membership functions to represent real user preferences are monotonic and trapezoidal. Some attributes are preferred similarly by all users (e.g. high salary is better than low), however, other attribute preferences may differ significantly (e.g. preference for required language level). These *preference types* can be seen on the right side of Fig. 4.

If we consider more than one attribute, we need to determine the final preference degree for every object, based on attribute preferences. User requirements are often conflicting so that the orderings are different, sometimes even reverse. Therefore we use *aggregation function* to acquire overall ordering. Aggregation function is n-ary monotone function $@^U : [0,1]^n \longrightarrow [0,1]$.

ISBN: 978-989-95079-6-8

Let us consider an object $x \in \Delta$ with attribute values $A_1(x), \ldots, A_n(x)$ and attribute preferences $\mu^U_{A_1}(A_1(x)) = p_1, \ldots, \mu^U_{A_n}(A_n(x)) = p_n$. The global preference degree of x will be $@^U(p_1, p_2, \ldots, p_n)$. A typical aggregation function is weighted average which reflects the idea that some attributes can be more important than the others. Our system acquires the weights from the interface shown on Fig. 3. We will describe other types of aggregation functions in Section 2.3.

2.2 Relevant Object Search

After storing user preferences the system proceeds to the next phase - relevant object search. It is a modification of Fagin's threshold algorithm [3] called *top-k* [2], based on the idea that users are interested only in a few best results from potentially huge dataset and that different users can have different attribute preferences.

Top-k uses a preprocessing phase to generate lists of objects ordered by attribute values. A typical structure for indexing is a B+ tree which allows us to follow records (attribute values) from highest to lowest or vice versa. User preference can define a different ordering of objects, but it can be simulated as well. If the user specified a trapezoidal membership function ("Something in the middle" preference), we start traversing values in both directions from the middle and we merge two non-increasing lists into one non-increasing list. Thus we gain a list of objects ordered from most preferred to least preferred. (The original threshold algorithm considered only one fixed ordering for each attribute.)

If our domain contains attributes $A_1, \ldots, A_n$, we prepare n ordered lists $L_1, \ldots, L_n$. Lists can be stored in heterogeneous repositories, or even acquired from web services. Top-k algorithm uses sorted access to the lists, reads attribute values and calculates preference values. One object can have different position in each ordered list, so some attribute values will be unknown in the meantime. After reading all attribute values of some object x, top-k algorithm calculates aggregation function to determine overall preference value of the object. A threshold value is determined to see if any object with some unknown values has a chance to be more preferred than x. If not, object x is written on the output.

This algorithm is very efficient for complex preferential search. The preprocessing phase causes that we can find k best objects without traversing the whole dataset. The output is an ordered list of k best objects (according to given preferences) together with overall preference values. Despite the fact that source data is crisp, we obtain fuzzy results similar to "object $x1$ is preferred by user U to degree $0,8$". This fuzziness originates from vague user preference representation and it is reflected in the results.

2.3 Learning Classification Rules from Ratings

Although the preference model is fuzzy, it is intended to represent real user preferences as precisely as possible. If they do not correspond, the user obtains different results, or results in wrong order. The graphical interface gives user a chance to rate the results in 5-degree scale (bad, poor, average, good, excellent) as can be seen on Fig. 6. Degrees can also be represented as percents, stars, smilies, etc. Then we use ordinal classification [8] to find a new aggregation function. As we showed in [9], it is sufficient to learn an aggregation function,



Figure 6: Interface to view and rate search results.

provided that types of attribute preferences did not change (e.g. from "The more the better" to "The less the better"). The new aggregation function has a form of *fuzzy rules*.

Fuzzy rules are well-known from fuzzy controllers [10], fuzzy inference systems [11] and many other application areas. A fuzzy rule consists of antecedent and consequent. *Antecedent* is typically a (fuzzy) conjunction or disjunction of fuzzy condition clauses. *Consequent* can be some action (in case of fuzzy controllers) or a fuzzy predicate (in case of fuzzy inference systems).

controller rule: IF temperature IS cold THEN heater IS high

inference system rule: IF wind IS strong AND barometer IS falling THEN forecast IS bad

Different types of rules look syntactically similar, but they differ in the semantics of consequent. In the first case the consequent "heater IS high" means that the heater device will be *set* to high level. In the latter case the consequent "forecast IS bad" means that the variable forecast will be assigned a fuzzy value bad, but no physical action will take place. Condition clauses contain crisp variables (temperature, wind, barometer) and fuzzy sets (cold, strong, falling) in both cases. Connectives AND, OR are fuzzy t-norms and t-conorms, IF-THEN implications are also fuzzified. There is usually a larger set of rules which are processed, their outputs are combined (aggregated) in some way into a single fuzzy set. It is then defuzzified to determine crisp output value. The most common defuzzification method is *centroid calculation*, which returns the geometric center of area under the fuzzy membership function.

Our usage is more similar to fuzzy inference systems because we use fuzzy rules to find a preference value, not to control some electronic device. It is somewhat simplified to reflect our preference model. The following example shows a set of classification rules. We chose fuzzy rules as aggregation method because they are human-understandable and very close to natural way of thinking.

IF $\mu^U_{salary}(salary(x)) \geq 0.7$ AND $\mu^U_{place}(place(x)) \geq 0.6$ AND $\mu^U_{position}(position(x)) \geq 0.5$ THEN $good^U(x) \geq 0.8$

IF $\mu^U_{experience_level}(experience_level(x)) \geq 0.3$ AND $\mu^U_{salary}(salary(x)) \geq 0.6$ THEN $good^U(x) \geq 0.6$

It is easy to see some differences compared to fuzzy rules mentioned above. Our fuzzy rules are induced from a small set of rated objects, so they are more specific. Condition clause $\mu^U_{salary}(salary(x)) \geq 0.7$ can be written more intuitively as $salary(x)$ IS good ≥ 0.7. We use arbitrary fuzzy sets $\mu^U_{A_i}$ instead of linguistic variables like "cold".

ISBN: 978-989-95079-6-8

Our condition clauses use α-cut defuzzification and thus our conjunctions and implications are crisp. If object x satisfies the antecedent, then the aggregation function $good^U$ will be greater or equal to the value specified in the consequent. If some object satisfies more antecedents, we choose the greatest output value. If it does not satisfy any rule antecedent, the result will be set to 0. A set of rules makes up a monotone aggregation function. An object y with all attribute preference values greater or equal to object x will be "good" to greater or equal degree than x.

3 Implementation

The system for user-dependent web search is implemented and available on http://x64.ics.upjs.sk:8080/nazou/. It is a web application created in Java with Wicket[1] framework. Java servlets run under Tomcat application server. The server also supports Sesame[2] database to store both our ontologies, i.e. job offer ontology and user preference ontology (see Fig. 2). User preference ontology schema is fixed, while domain ontology is arbitrary (see [12, 9] for details).

The application is divided into relatively independent tools: top-k search, inductive learning, and user preference acquisition called UPreA. UPreA provides user preferences for other tools, processes user inputs and manages all changes in user profile. It communicates with Sesame database through Java API to store and update user information.

4 System and Model Evaluation

Our goal is to experiment with the implemented system and to find out if our model is a suitable representation of real user preferences. As our model is fuzzy, we are interested in fuzzy degree of correspondence between our results and some "ideal" results. Analysis of experiments is further complicated by the fact that user preferences are evolving in time and user inputs can be inconsistent.

We anticipate vague user preferences and therefore we use inductive learning to adjust the aggregation function. Our preferences define an ordering of all objects from the domain, while user ratings define another ordering of some small subset of objects. Fig. 2 shows orderings that are generated in search cycles: R^i is the ordering of search results and U^i is the ordering specified by user ratings. We explore the correspondence of R^i and U^i. If these orderings are similar, the user must have assigned higher ratings to objects from the front of the result list and therefore our preference model reflected real user preferences well. Throughout the rest of this section we assume that R^i and U^i are permutations of objects $\{o_1, \ldots, o_n\}$. Also note that results R^0 are retrieved with user-defined aggregation function (weighted average), while other results R^i, $i \geq 1$ are retrieved with aggregation (fuzzy rules) learned from previous ratings.

The main problem is how to determine the correspondence of two fuzzy subsets. We present several measures below and compare the results. We analyse how R^i corresponds with U^i in six successive search cycles. In fact, some users stopped searching after the first or the second cycle, so the number of analyzed ordered lists is slightly decreasing. Data for this analysis (search results R^i and user ratings U^i) comes from real users who tested the implemented system.

Table 1: Correlations of result ordering and user evaluation.

Cycle	0.	1.	2.	3.	4.	5.
τ-correlation	0.73	0.77	0.79	0,77	0,81	0,81
Result lists	174	134	102	53	26	15

Table 2: Weighted similarity of result ordering and user evaluation.

Cycle	0.	1.	2.	3.	4.	5.
similarity	0.53	0.41	0.39	0,56	0,61	0,64

4.1 Correlation of Orderings

The first possibility is Kendall τ-correlation coefficient [13]. It is defined as $\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)} - 1$, where n_c is the number of concordant pairs and n_d is the number of discordant pairs in both ratings. A pair of elements is discordant if their order is reversed. We modified this formula to $\tau = \frac{2P}{\frac{1}{2}n(n-1)}$, so that results belong to interval $[0, 1]$ to conform with our fuzzy model. However, in this case only results above 0.5 mean a positive correlation.

Kendall τ-correlation originally deals with strictly ordered lists. However, in our case many objects from R^i or U^i can share the same fuzzy value. We distinguish ties according to the (strict) order in which they are presented to user. This order in fact depends on indexing methods used in top-k search.

4.2 Weighted order similarity

If the order of some objects change, τ-correlation decreases evenly, regardless of the position where the discordance arose. We can also assume that the position of discordance is important – discordance on the 1^{st} and the 2^{nd} position is more significant than discordance on the 8^{th} and the 9^{th} position. This idea is reflected in *weighted order similarity*. We choose a decreasing vector of weights such that differences between adjacent values are also decreasing, e.g. $w = \langle 20, 15, 11, 8, 6, 5, 4, 3, 2, 1\rangle$ for $n = 10$. Weighted order coefficient is defined as $\mathcal{W} = 1 - \sum_{j=1}^{n} \left| w_{p_j} - w_{q_j} \right|$ where p_j is a sequence number of R^i_j and q_j is a sequence number of U^i_j. Note that this coefficient does not consider all pairs of elements as τ-correlation, only the sequence number of the same element in both orderings.

4.3 Fuzzy Equality

Another suitable measure is *fuzzy equality* [14], which is stronger condition than similarity. In this case we consider fuzzy aggregation values $@^U$, not only orders. We define fuzzy equality of fuzzy sets μ_A, μ_B as: $\equiv^\bullet (\mu_A, \mu_B) = \inf_{x\in\Delta}\{\vee^\bullet(\rightarrow^\bullet (\mu_A, \mu_B), \rightarrow^\bullet (\mu_B, \mu_A))\}$. It means that μ_A, μ_B are equal if $\mu_A(x)$ implies $\mu_B(x)$ and vice versa, for all x from the domain. We can choose different types of fuzzy implication and conjunction, namely Łukasiewicz, Gödel and product. The universal quantifier is always represented as infimum for all elements from the domain.

Finally, we compare results of these fuzzy measures on Fig. 7. τ-correlation and weighted similarity seem to be the most

[1] http://wicket.apache.org/
[2] http://www.openrdf.org/

ISBN: 978-989-95079-6-8

Table 3: Fuzzy equality of result ordering and user evaluation.

Cycle	0.	1.	2.	3.	4.	5.
$\equiv^{\bullet}_{L}$	0.19	0.41	0.32	0,35	0,42	0,42
$\equiv^{\bullet}_{G}$	0.23	0.22	0.14	0,20	0,22	0,24
$\equiv^{\bullet}_{p}$	0.17	0.28	0.16	0,23	0,31	0,27

Table 4: τ_R-correlation of result lists R^i, R^{i+1}.

i	0	1	2	3	4
$\lvert R^i \cap R^{i+1} \rvert$	4,50	6,21	5,95	4,88	5,52
$\tau_{\cap}$	0,65	0,80	0,76	0,63	0,76
τ_R	0,53	0,67	0,64	0,56	0,64

suitable measures. Fuzzy equalities impose too strict conditions on fuzzy values. We do not expect result lists and user ratings to be totally equal because we learn new preferences from their difference. Correlation increases most significantly between 1^{st} and 2^{nd} cycle where we start to use fuzzy rules. So experiments proved that fuzzy rules are suitable representation of user preference.

Another interesting point is that τ-correlation and weighted similarity have the same trends up to fifth cycle, while all fuzzy equalities have the same trends from second to fifth cycle. Łukasiewicz equality has the greatest results and Gödel equality has the smallest results because the corresponding fuzzy implications also have this feature.

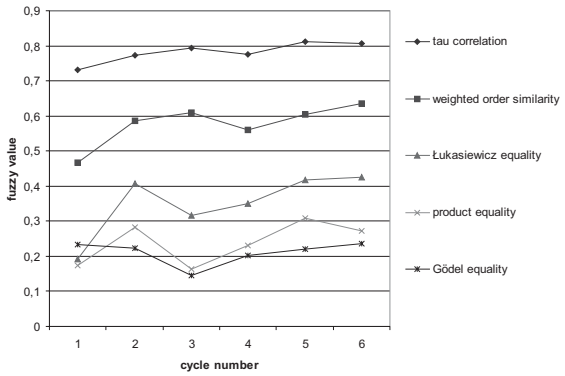


Figure 7: Comparison of different fuzzy measures.

4.4 Correlation of Subsequent Result Lists

We are also interested in the correlation of k-tuples R^i, R^{i+1}. If R^i and R^{i+1} have empty intersection, this correlation will be 0. In case of non-empty intersection, we can compute usual τ-correlation coefficient of the common elements $\tau_{\cap}$ and multiply the result with $\frac{\lvert R^i \cap R^{i+1} \rvert}{n}$. Thus we obtain τ_R. Average results for the first 5 cycles are shown in Table 4.

5 Related Work

In this section we present a short survey of similar preference models and searching methods. Many searching methods allow user to select desired attribute values, but they mostly produce a conjunctive query from user requirements. Our model is more general because aggregation functions are generalization of both (fuzzy) conjunctions and disjunctions.

A similar problem is addressed by *fuzzy multicriteria decision making* [15, 16]. It supports decision based on various attributes called *criteria*. Attribute preferences are usually represented as *fuzzy preference relations* – binary relations defined for every two-tuple of objects. Preference relation defines a degree to which one object is better than another object, based on the corresponding attribute. Other models of attribute preference mentioned in [16] are *utility functions* (identical with our fuzzy functions) and crisp *preference orderings* of objects. Aggregation is defined as *OWA operator* [17]. It is ordered weighted average associated with a normalized weight vector $w = (w_1, \ldots, w_n)^T$. Note that weight w_i is not connected with preference value of i^{th} attribute, but with i^{th} greatest preference value. This operator produces one "collective" preference relation from attribute preferences.

The *main difference* is that multicriteria decision is mainly designed for a small number of objects (alternatives). Each preference relation requires quadratic space w.r.t. number of objects. Decision making process consists of computing aggregated preference relation and transforming this relation to a global ranking of objects. Efficiency issues are left to implementations. Our model is closely connected with efficient top-k searching method which allows finding best objects without computing aggregation for all of them and sorting the results. Advantages of fuzzy multicriteria decision include support of sophisticated aggregation methods like Choquet integral and various modifications of OWA operator.

Another method to find interesting objects according to more attributes is based on *Skyline* [18]. Every attribute forms one dimension and every object is represented as a point in multi-dimensional space. Skyline is defined as a set of points which are not *dominated* by any other point (Pareto optimal), which means that no other point is equally good or better in all dimensions and strictly better at least in one dimension. A special Skyline clause can be integrated to relational database systems. Efficient methods of finding Skyline are proposed in [18, 19], including partitioning, indexing with B-trees and searching with top N algorithm (similar to top-k).

Compared to our system, Skyline produces a set of *unordered*, potentially interesting objects. Despite the similar searching methods, results can be different from our top-k results. First object o returned by top-k always belongs to Skyline set, but not necessarily the following objects, which can be in fact dominated by o, but still very close to it. Skyline also does not support attribute preferences other than simple "The more the better" or its reverse for ordinal attributes, thus it assumes the same attribute preferences for all users.

Another similar approach is *rank-aware querying* [20]. It uses general aggregation (scoring) functions to compute object ranking and returns a list of results ordered by ranking. Aggregations and new rank-join operators are integrated into relational database system. Rank-join operator performance is optimized by transforming a query into an execution plan and finding the most efficient plan. As in the case of Skyline operator, rank-aware querying does not support various attribute preferences or fuzzy rules. Simple preferences like "The less the better" can be simulated with scoring functions.

All-purpose fuzzy database system *FSQL Server* is proposed in [21]. It supports attribute preferences (called *linguistic labels*) for both ordered and unordered attribute do-

ISBN: 978-989-95079-6-8

mains. It is possible to define special fuzzy connectives to aggregate the results, but weighted average and fuzzy rules are not supported. In addition to fuzzy features needed in our system, FSQL supports fuzzy comparison, fuzzy constants, special values and α-cuts (fulfillment thresholds). FSQL is intended to support all common fuzzy features, while our system is designed for a single purpose of preferential web search.

From theoretical point of view, *rough set theory* [22] also has some common ideas with our model. It uses conditional attributes and decision attributes (similar to our user rating). Objects with the same attribute values belong to the same equivalence class and they are called *indiscernible* by those attributes. However, a set that contains one element from an equivalence class and does not contain another object from the same equivalence class, cannot be defined exactly (using only available attributes). Such set is called *rough set* and it can be defined with help of lower and upper approximation.

The main goal of rough set theory is to synthesize *decision rules* to determine the value of decision attribute from conditional attributes. The main difference is that a rough set *does not generate any ordering*, we can only distinguish objects inside the set from "boundary" objects.

6 Conclusions

Some approaches mentioned in the previous section are focused on theoretical support of decision process, some deal more with efficiency and optimization. We address both issues to achieve a clear-cut formal model and efficient implementation. Our main purpose is to enable user-specific web search in one arbitrary domain. All parts described in section 2 are important for personalized search: web information extraction, interface for explicit preference specification, top-k search, user rating and inductive learning. To the best of our knowledge, other fuzzy preference systems do not address all these phases to connect web and user.

So our approach has the following contributions:

- *Formal model* – fuzzy sets and aggregation as a model of user preference, suitable for preferential search.
- *Expressive power* – arbitrary fuzzy sets and fuzzy rules are strong tools to represent preferences.
- *Implementation* – web-based system implements specified functionality.
- *Model evaluation* – various fuzzy measures are used to analyze the results.

Acknowledgment

Partially supported by Czech projects 1ET100300517, MSM-0021620838 and Slovak project VEGA 1/0131/09.

References

[1] Tomáš Horváth. A model of user preference learning for content-based recommender systems. *Accepted for publication to COMPUTING AND INFORMATICS*.

[2] Peter Gurský and Peter Vojtáš. On top-k search with no random access using small memory. *Proceeding of ADBIS*, pages 97–111, 2008.

[3] Ronald Fagin, Amnon Lotem, and Moni Naor. Optimal aggregation algorithms for middleware. *PODS '01: Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 102–113, 2001.

[4] John Musser and Tim O'Reilly. *Web 2.0 Principles and Best Practices*. O'Reilly Radar Reports, 2006.

[5] D. Maruščák, R. Novotný, and P. Vojtáš. Unsupervised structured web data and attribute value extraction. *Proceedings of Znalosti*, 2009.

[6] Robert Baumgartner, Sergio Flesca, and Georg Gottlob. Visual web information extraction with lixto. pages 119–128, 2001.

[7] Eugene Agichtein. Web information extraction and user modeling: Towards closing the gap. *IEEE Data Eng. Bull.*, 29(4):37–44, 2006.

[8] Tomáš Horváth and Peter Vojtáš. Ordinal classification with monotonicity constraints. *ICDM 2006*, pages 217–225, 2006.

[9] Peter Gurský, Tomáš Horváth, Jozef Jirásek, Stanislav Krajči, Róbert Novotný, Veronika Vaneková, and Peter Vojtáš. Knowledge processing for web search - an integrated model. *Studies in Computational Intelligence (vol. 78)*, pages 95–104, 2007.

[10] Michio Sugeno. *Industrial Applications of Fuzzy Control*. Elsevier Science Inc., New York, NY, USA, 1985.

[11] E. H. Mamdani and S. Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *Int. J. Hum.-Comput. Stud.*, 51(2):135–147, 1999.

[12] Peter Gurský, Veronika Vaneková, and Jana Pribolová. Fuzzy user preference model for top-k search. *Fuzzy Systems, 2008 (IEEE World Congress on Computational Intelligence)*, pages 1606–1612, 2008.

[13] Maurice Kendall and Jean D. Gibbons. *Rank Correlation Methods*. A Charles Griffin Title, 5 edition, 1990.

[14] Bernard De Baets and Radko Mesiar. Metrics and t-equalities. *Journal of Mathematical Analysis and Applications*, 267(2):531 – 547, 2002.

[15] Ronald R. Yager. Extending multicriteria decision making by mixing t-norms and owa operators: Research articles. *Int. J. Intell. Syst.*, 20(4):453–474, 2005.

[16] F. Chiclana, F. Herrera, and E. Herrera-viedma. Integrating three representation models in fuzzy multipurpose decision making based on fuzzy preference relations. *Fuzzy Sets and Systems*, 97:33–48, 1998.

[17] Ronald R. Yager and Janusz Kacprzyk, editors. *The ordered weighted averaging operators: theory and applications*. Kluwer Academic Publishers, 1997.

[18] Stephan Börzsönyi, Donald Kossmann, and Konrad Stocker. The skyline operator. *Proceedings of the 17th International Conference on Data Engineering*, pages 421–430, 2001.

[19] Jan Chomicki, Parke Godfrey, Jarek Gryz, and Dongming Liang. *Intelligent Information Processing and Web Mining*, volume 31/2005 of *Advances in Soft Computing*, chapter Skyline with Presorting: Theory and Optimizations, pages 595–604. Springer-Verlag, 2005.

[20] Ihab F. Ilyas, Rahul Shah, Walid G. Aref, Jeffrey Scott Vitter, and Ahmed K. Elmagarmid. Rank-aware query optimization. *SIGMOD '04: Proceedings of the 2004 ACM SIGMOD international conference on Management of data*, pages 203–214, 2004.

[21] Jose Galindo. *Fuzzy Databases: Modeling, Design, and Implementation*. IGI Publishing, 2006.

[22] Lech Polkowski and Andrzej Skowron. Rough sets: A tutorial. *Rough Fuzzy Hybridization: A New Trend in Decision-Making*, 2000.

ISBN: 978-989-95079-6-8

IFSA-EUSFLAT 2009

Multi-Goal Aggregation of Reduced Preference Relations Based on Fuzzy Interactions between Decision Goals

Rudolf Felix
F/L/S Fuzzy Logik Systeme GmbH, Joseph-von-Fraunhofer Straße 20, 44 227 Dortmund, Germany
Tel. +49 231 9 700 921, Fax. +49 231 9700 929, e-mail: felix@fuzzy.de

Abstract– *Many aggregation operators like fuzzy integrals and similar require a preference relation as input. If the number of decision alternatives increases it becomes almost impossible to provide the exponentially increasing number of preference statements which are needed because a preference relation is defined upon the power set of the set of decision alternatives. If, as in many real world applications, the number of decision goals increases, it is also almost impossible to supervise the consistency of preference statements with respect to the interactions between the decision goals. In this paper it is discussed in which way a decision making approach based on interactions between goals is applied in order to use preference information that is not defined upon the power set of the set decision alternatives but simply on the decision set itself. Instead of defining the preferences upon the power set of the decision alternatives with respect to all the goals, for every decision goal the preference of the decision alternatives for this goal is defined upon the set of decision alternatives by a linear preference ranking of the decision alternatives. Consequences for both decision modeling based on classical preference relations and based on weighted sums are discussed.*

Keywords– Aggregation complexity, decision making, interactions between goals, reduced preference relation, weighted sum.

1 Introduction

Many decision making approaches are limited with respect to the management of complexity [4], [5], [8], [9], [10]. The consequence of this limitation is that these models are rather not applicable for real world problems. Many aggregation approaches require a preference relation as input. For more complex decision problems such preference relations are difficult to obtain because a preference relation is defined upon the power set of the set of the decision alternatives. If the number of decision alternatives increases it becomes almost impossible to provide the exponentially increasing number of preference statements needed. If additionally, as in many real world applications, the number of decision goals increases, it is almost impossible to supervise the consistency of preference statements with respect to the interactions between the decision goals. In this paper it is discussed in which way a decision making approach based on interactions between goals is adapted in order to aggregate in complex decision situations requiring only reduced linear preference information. The required input preference information is not defined upon the power set of the decision alternatives. Instead of this, for every single decision goal a linear preference ranking of the decision alternatives with respect to that goal is required. Of course, the single goal preference rankings may rank the decision alternatives in different way for different goals as the goals usually are partly conflicting. The presented aggregation approach ascertains for each pair of goals their conflicts (and correlations) by computing the so called interactions between the goals from the initial input single goal rankings. The only additional information needed when the interactions between the goals have been calculated is a linear importance information of each goal which is expressed in terms of goal priorities. The goal priorities are numbers of [0,1] and are comparable with a fuzzy measure that ranks not the decision alternatives but the decision goals itself with respect to their importance (priority).

It turned out that the decision making based on interactions between goals is less limited because the complexity of both the input information required and the aggregation process is not higher than polynomial. Since the model has successfully been applied to many real world problems [3] it clearly helped to manage many complex aggregation processes in which the number was varying between 50 and 100. The number of decision alternatives was in some applications even more than several thousands.

In the subsequent sections first for better readability of the paper we repeat a brief description of the decision making approach based on interactions between decision goals. Then we show how the approach is applied in order to work with single goal preference rankings and a statement with respect to classical preference based decision modeling is made. Subsequently it is discussed under which conditions decision models based on weighted sums help and how they are related to decision situations with interacting decision goals. Finally, the consequences of the results are discussed.

2 Decision Making based on Interactions between Goals

In the following it is shown how an explicit modeling of interaction between decision goals that are defined as fuzzy sets of decision alternatives helps to manage complexity of the decision making and aggregation. This modeling of the decision making and aggregation process significantly differs from the related approaches and the way they manage complex decision situations. First the notion of positive and negative impact sets is introduced. Then different types of interaction between goals are defined. After this it is shown how interactions between goals are used in order to aggregate pairs of goals to the so called local decision sets. Then it is described how the local decision sets are used for the aggregation of a final decision set. The complexity of the different steps is discussed.

2.1 Positive and Negative Impact Sets

Before we define interactions between goals as fuzzy relations, we introduce the notion of the positive impact set and the negative impact set of a goal. A more detailed discussion can be found in [1],[2] and [3].

Def. 1a) Let A be a non-empty and finite set of decision alternatives, G a non-empty and finite set of goals, $A \cap G$

ISBN: 978-989-95079-6-8

$=\emptyset, a \in A, g \in G, \delta \in (0,1]$. For each goal g we define the two fuzzy sets S_g and D_g each from A into [0, 1] by:

1. Positive impact function of the goal g: $S_g(a) := \delta$, if a affects g positively with degree δ, $S_g(a) := 0$ else.
2. Negative impact function of the goal g: $D_g(a) := \delta$, if a affects g negatively with degree δ, $D_g(a) := 0$ else.

Def. 1b) Let S_g and D_g be defined as in Def. 1a). S_g is called the positive impact set of g and D_g the negative impact set of g.

The set S_g contains alternatives with a positive impact on the goal g and δ is the degree of the positive impact. The set D_g contains alternatives with a negative impact on the goal g and δ is the degree of the negative impact.

2.2 Interactions between Goals

Let $P(A)$ be the set of all fuzzy subsets of A. Let X, Y ∈ $P(A)$, x and y the membership functions of X and Y respectively. Assume now that we have a binary fuzzy inclusion $I: P(A) \times P(A) \to [0,1]$ and a fuzzy non-inclusion $N: P(A) \times P(A) \to [0,1]$, such that $N(X,Y) := 1 - I(X,Y)$. In such a case the degree of inclusions and non-inclusions between the impact sets of two goals indicate the degree of the existence of interaction between these two goals. The higher the degree of inclusion between the positive impact sets of two goals, the more cooperative the interaction between them. The higher the degree of inclusion between the positive impact set of one goal and the negative impact set of the second, the more competitive the interaction. The non-inclusions are evaluated in a similar way. The higher the degree of non-inclusion between the positive impact sets of two goals, the less cooperative the interaction between them. The higher the degree of non-inclusion between the positive impact set of one goal and the negative impact set of the second, the less competitive the relationship. The pair (S_g, D_g) represents the whole known impact of alternatives on the goal g. Then S_g is the fuzzy set of alternatives which satisfy the goal g. D_g is the fuzzy set of alternatives which are rather not recommendable from the point of view of satisfying the goal g.

Based on the inclusion and non-inclusion between the impact sets of the goals as described above, 8 basic fuzzy types of interaction between goals are defined. The different types of interaction describe the spectrum from a high confluence between goals (analogy) to a strict competition (trade-off) [1].

Def. 2) Let S_{g_1}, D_{g_1}, S_{g_2} and D_{g_2} be fuzzy sets given by the corresponding membership functions as defined in Def. 1). For simplicity we write S_1 instead of S_{g_1} etc.. Let $g_1, g_2 \in G$ where G is a set of goals. *T* is a t-norm.

The fuzzy types of interaction between two goals are defined as relations which are fuzzy subsets of $G \times G$ as follows*:*

1. g_1 *is independent of* g_2*:* <=> $T(N(S_1,S_2), N(S_1,D_2), N(S_2,D_1), N(D_1,D_2))$
2. g_1 *assists* g_2*:* <=> $T(I(S_1 S_2), N(S_1,D_2))$
3. g_1 *cooperates with* g_2*:* <=> $T(I(S_1,S_2), N(S_1,D_2), N(S_2,D_1))$
4. g_1 *is analogous to* g_2*:* <=> $T(I(S_1,S_2), N(S_1,D_2), N(S_2,D_1), I(D_1,D_2))$
5. g_1 *hinders* g_2*:* <=> $T(N(S_1,S_2), I(S_1,D_2))$
6. g_1 *competes with* g_2*:* <=> $T(N(S_1,S_2), I(S_1,D_2), I(S_2,D_1))$
7. g_1 *is in trade-off to* g_2*:* <=> $T(N(S_1,S_2), I(S_1,D_2), I(S_2,D_1), N(D_1,D_2))$
8. g_1 *is unspecified dependent from* g_2*:* <=> $T(I(S_1,S_2), I(S_1,D_2), I(S_2,D_1), I(D_1,D_2))$

The interactions between goals are crucial for an adequate orientation during the decision making process because they reflect the way the goals depend on each other and describe the pros and cons of the decision alternatives with respect to the goals. For example, for cooperative goals a conjunctive aggregation is appropriate. If the goals are rather competitive, then an aggregation based on an exclusive disjunction is appropriate.

Note that the complexity of the calculation of every type of interaction between two goals is O(card(A) * card(A)) = O((card(A))^2) [4].

2.3 Two Goals Aggregation based on the Type of their Interaction

The assumption, that cooperative types of interaction between goals imply conjunctive aggregation and conflicting types of interaction between goals rather lead to exclusive disjunctive aggregation, is easy to accept from the intuitive point of view. It is also easy to accept that in case of independent or unspecified dependent goals a disjunctive aggregation is appropriate. For a more detailed formal discussion see for instance [1],[2]. Knowing the type of interaction between two goals means to recognize for which goals rather a conjunctive aggregation is appropriate and for which goals rather a disjunctive or even exclusively disjunctive aggregation is appropriate. This knowledge then in connection with information about goal priorities is used in order to apply interaction dependent aggregation policies which describe the way of aggregation for each type of interaction. The aggregation policies define which kind of aggregation operation is the appropriate one for each pair of goals. The aggregation of two goals gi and gj leads to the so called local decision set $L_{i,j.}$. For each pair of goals there is a local decision set $L_{i,j.} \in P(A)$, where A is the set of decision alternatives (see Def 1 a)) and P(A) the power set upon A. For conflicting goals, for instance, the following aggregation policy which deduces the appropriate decision set is given:

if (g_1 is in trade-off to g_2) and (g_1 is slightly more important than g_2) then $L_{1,2} := S_1 / D_2$.

In case of very similar goals (analogous or cooperative goals) the priority information even is not necessary:
if (g_1 cooperates with g_2) then $L_{1,2} := S_1 \cap S_2$ because $S_1 \cap S_2$ surely satisfies both goals.
if (g_1 is independent of g_2) then $L_{1,2} := S_1 \cup S_2$ because

ISBN: 978-989-95079-6-8

$S_1 \cup S_2$ surely do not interact neither positively nor negatively and we may and want to pursue both goals.

In this way for every pair of goals g_i and g_j, i,j ∈ {1,…,n} decision sets are aggregated. The importance of goals is expressed by the so called priorities. A priority of a goal g_i is a real number $P_i \in [0,1]$. The comparison of the priorities is modeled based on the linear ordering of the real interval [0,1]. The statements like g_i *slightly more important than* g_j are defined as linguistic labels that simply express the extend of the difference between P_i and P_j.

2.4 Multiple Goal Aggregation as Final Aggregation based on the Local Decision Sets

The next step of the aggregation process is the final aggregation. The final aggregation is performed based on a sorting procedure of all local decision sets $L_{i,j}$. Again the priority information is used to build a semi-linear hierarchy of the local decision sets by sorting them. The sorting process sorts the local decision sets with respect to the priorities of the goals. Subsequently an intersection set of all local decision sets is built. If this intersection set is empty then the intersection of all local decision sets except the last one in the hierarchy is built. If the resulting intersection set again is empty then the second last local decision set is excluded from the intersection process. The process iterates until the intersection is not empty (or more generally speaking until its fuzzy cardinality is big enough with respect to a given threshold). The first nonempty intersection in the iteration process is the final decision set and the membership values of this set give a ranking of the decision alternatives that is the result of the aggregation process (for more details see [2]).

2.5 Complexity Analysis of the Aggregation Process

As already discussed for instance in [4] the complexity of the aggregation process is $O((card(A))^2 * (card(G))^2)$ and the complexity of the information required for the description of both the positive and the negative impact functions is $O(card(A) * card(G))$.

3 Application of the Aggregation in Case of Reduced Preference Relations

In preference based decision making the input preference information has to be defined upon the power set of the decision alternatives [6]. This means that the complexity of the input information required is exponential with respect to the cardinality of the set of decision alternatives. This also means that the input information required is very difficult to obtain. Especially if the number of decision goals increases and the goals are partly conflicting the required preference information has to be multidimensional and the provider of the preference information has to express all the multidimensional interactions between the goals through the preference relation. With increasing number of goals and interactions between them the complexity of the required input preference relation possesses the same complexity as the decision problem itself. But, if the complexity of the required input is the same as the solution of the underlying decision problem itself then the subsequent aggregation of the input does not really help to solve the problem and is rather obsolete.

Therefore we propose to reduce the complexity of the input preference relation. Instead of requiring a preference relation defined upon the power set of the set of the decision alternatives which expresses the multidimensionality of the impacts of the decision alternatives on the goals for every single decision goal a linear preference ranking of the decision alternatives with respect to that goal is required. This means that for every goal a preference ranking defined on the set of decision alternatives is required instead of a ranking defined on the power set of the alternatives. The multidimensionality of the goals is then computed from all the single goal preference rankings using the concept of interactions between the goals as defined in section X.2.

In the sequent we extend the definition Def. 1. The extension defines how a single goal preference ranking defined on the set of the decision alternatives is transformed into positive and negative impact sets:

Def. 1c) Let A be a non-empty and finite set of decision alternatives, G a non-empty, finite set of goals as defined in Def. 1) a),

$A \cap G = \emptyset, a \in A, g \in G, \delta \in (0,1]$.

Let $>p_g$ be a preference ranking defined upon A with respect to *g* defining a total order upon A with respect to g, such that

$a_{i1} >p_g a_{i2} >p_g a_{i3} >p_g \ldots >p_g a_{im}$, where m=card(A) and $\forall a_{ij}, a_{ik} \in A, a_{ij} >p_g a_{ik} :\Leftrightarrow a_{ij}$ is preferred to a_{ik} with respect to the goal g. The preference relation $>p_g$ is called the *reduced single goal preference relation* of the goal g.

For simplicity, instead of $>p_g$ we also equivalently write RSPR of the goal g. All the RSPRs for all $g \in G$ are called the *reduced preference relation* RPR for the whole set of goals *G*. In order to avoid complete redundancy within the RPR we additionally define that the RSPRs of all the goals are different.

Let us assume that there is a decision situation with *n* decision goals where *n=card(G)* and *m* decision alternatives where *m=card(A)*. In the subsequent we propose an additional extension of the original definition Def. 1b with the aim to transform the single goal preference relations RSPR of every goals $g \in G$ into the positive and negative impact sets *Sg* and *Dg*:

Def. 1d) Let again A be a non-empty and finite set of decision alternatives, G a non-empty, finite set of goals as defined in Def. 1) a), $A \cap B \neq \emptyset$, $a_i \in A$, *m=card(*A*), g*∈G, i, c∈ {1, …, *m*}. For any goal g we obtain both the positive and the negative impact sets Sg and Dg by defining the values of δ according to Def. 1a) and 1b) as follows:

Def. 1d1) For the positive impact set:

$Sg(a_i)=\delta:=1/i$ iff $i \in [1,c-1]$,
$Sg(a_i)=\delta:=0$ iff $i \in [c,m]$.

Def. 1d2) For the negative impact set:

$Dg(a_i)=\delta:=0$ iff $i \in [1,c-1]$,
$Sg(a_i) = \delta:= 1/(m-i+1)$ iff $i \in [c, m]$.

Using the definition Def. 1d1 and Def. 1d2 for any goal g∈G we obtain a transformation of the RPR into positive

ISBN: 978-989-95079-6-8

and negative impact sets of all the goals and can evaluate the interactions of the goals using Def. 2 that are implied by the RPR.

Compared to classical preference based decision models this transformation helps to reduce the complexity of the input preference information required without losing modeling power for complex real world problems. The advantage is that using Def. 2. the interactions between goals that are implied by the RPR expose the incompatibilities and compatibilities that may be hidden in the RPR. The exposed incompatibilities and compatibilities are used adequately during the further calculation of the decision sets. Note that the exposition is calculated with a polynomial number of calculation steps with degree 2. The only additional information required from the decision maker is the priority information for each goal which has to be expressed as a weight with a value between 0 and 1. Many real world applications show that despite the reduced input complexity there is no substantial loss of decision quality [3].

*Statement 1***:** *In particular this means that it is not necessary to have classical preference relations defined upon the power set of the decision alternatives in order to handle complex decision problems with both positively and negatively interacting decision goals.*

Another interesting question is how the decision making based on interactions between decision goals is formally related to aggregation methods based on weighted sums. In order to investigate this we introduce the notion of r-consistency of RPRs and will consider the question under which conditions the weighted sum aggregation may be appropriate from the point of view of the application of the decision making based on interactions between goals if we have an RPR as input. For this we define the following:

Def.3) Given a discrete and finite set A of decision alternatives. Given a discrete and finite set G of goals. Let $r \in (0,1]$. The reduced preference relation RPR is called r-consistent :$\Leftrightarrow \exists\, c \in \{1, \ldots, m\}$, $m=card(A)$ such that $\forall\, (g_i, g_j) \in G \times G$, $i,j \in \{1, \ldots, n\}$, $n=card(G)$, (g_i *cooperates with* g_j) $\geq r$.

Let us now consider an important consequence of the interaction between goals using the notion of r-consistency. This notion will imply a condition under which an aggregation based on a weighted sum may lead to an appropriate final decision.

For this let us assume that the quite sophisticated final aggregation process of the iterative intersections as described in section 2.4 is replaced by the following rather intuitive straight forward consideration of how to obtain an optimal final decision.

Again let us identify the local decision sets $L_{i,j}$ obtained after the application of the local decision policies for each pair of goals (g_i, g_j) as the first type of decision subsets of the set of all decision alternatives A which, according to the decision model are expected to contain an optimal decision alternative a_k. Thus we define the set of sets

T1:= { $L_{i,j}$ | i ,j $\in$ {1, .. n}, $n=card(G)$} and expect that the optimal decision alternative has a positive membership in at least one of the sets of T1. Since we want to consider multiple goals we may also expect that an optimal decision alternative may have a positive membership in at least one of the intersections of pairs of the local $L_{i,j}$. Thus we define T2:= { $L_{k,l} \cap L_{p,q}$ | k,l,p,q $\in$ {1, .. n}, $n=card(G)$}. In order to simplify the subsequent explanation we concentrate on the crisp case and replace all membership values > 0 in all $L_{i,j}$, and $L_{k,l} \cap L_{p,q}$ by the membership value 1. Now we define the system of these crisp sets GDS as follows:

Def.4) GDS:={∅,T1,T2}, T1,T2 are the sets of crisp sets that we construct as described above by replacing all membership values > 0 in all $L_{i,j}$, and $L_{k,l} \cap L_{p,q}$ by the membership value 1.

With this definition we are able to formulate the following theorem that describes a property of the crisp GDS in the case that the underlying decision situation stems from a reduced preference relation RPR to which the decision model based on interactions between decision goals is applied. This property will enable us to relate this decision making to the calculation of optimal decisions by concepts based on weighted sums that are strongly connected with the notion of a matroid [7] and optimal decisions obtained by Greedy algorithms.

Theorem 1: If the reduced preference relation RPR is r-consistent then the system (P(A),GDS) is a matroid.

Sketch of the Proof:

The proof will show that the following matroid conditions [7] hold: 1. $\emptyset \in$ GDS, 2. $X \subseteq Y$, $Y \in$ GDS $\Rightarrow X \in$ GDS and 3. $X,Y \in$ GDS card(X)<card(Y) $\Rightarrow \exists x \in Y \backslash X$ and $X \cup \{x\} \in$ GDS.

Ad1.: Holds by definition of GDS. Ad2.: If X is one of the $L_{i,j}$s then Y must be one of the another $L_{k,l}$ with $i \neq k$, $j \neq l$. If X is one of the $L_{k,l} \cap L_{p,q}$ then Y must be $L_{k,l}$ or $L_{k,l}$. Ad3.: This condition holds because the RPR is r-consistent and therefore we have $\forall$ k,l,p,q $\in$ {1, .. n}, $n=card(G)$, $card(L_{i,j})=card(L_{k,l} \cap L_{p,q})=c$.

With this result we obtain the following corollary that relates decision making based on interactions between decision goals to calculating optimal decisions by concepts based on weighted sums.

Corollary 1: If the reduced preference relation RPR is r-consistent then there exists a greedy algorithm based on the maximization of the weighted sum $\Sigma_{i \in \{1,\ldots,n\}, k \in \{1,\ldots,m\}} P_i * S_i(a_k)$, *where* S_i *is defined as in Def. 1d1 and* $a_k \in A$ *being a decision alternative,* $n=card(G)$, $m=card(A)$. P_i *is the priority of the goal* g_i *as used for the calculation of both the local and the global decision sets (see sections 4.3 and 4.4).*

Sketch of the Proof:

The proof is an immediate result from the Theorem 1 and the theory of matroids used in the field of combinatorial optimization [7].

4 Discussion of the Consequences

As already mentioned, this corollary relates the decision making based on interactions between decision goals to the calculation of optimal decisions by concepts based on weighted sums. It shows that weighted sums both as aggregation and optimization concept are rather appropriate if the goals or criteria are cooperative e.g. if they interact

ISBN: 978-989-95079-6-8

positively (or at least do not interact at all being independent). In contrast to this the decision making based on interactions between decision goals is more general and it reflects both positive and negative interactions between the goals. This is important in the context of real world decision making and optimization problems which usually possess partly conflicting goal structures. If the aggregation is performed by weighted sums, like for instance Choquet integrals, then one can suppose that the aggregation will only work if the choices to be made are free of conflicts. The hypothesis formulated in [6] postulating that there exists a preference relation on subsets of decision alternatives which is compatible with a weak order of decision alternatives belonging to different subsets of decision alternatives seems not to be that natural as postulated (especially in the field of optimization problems where partly contradicting goals are daily business).

Even if for a given application field the postulated preference relation exists, in case the number of decision alternatives increases it becomes almost impossible to provide the exponentially increasing number of required preference statements. If additionally, as in many real world applications, the number of decision goals increases, it is almost impossible for a human decision maker to keep all the interactions between the goals in mind and supervise the consistency of the preference statements with respect to the interactions between these decision goals. We see that the adapted decision making approach based on interactions between goals helps reducing the complexity of the required preference information without reducing neither the quality of the decision results [3] nor the complexity of decision situations that are modeled [4]: The required input preference information is not defined upon the power set of the decision alternatives. Instead of defining the required preferences upon the power set of the decision alternatives with respect to all the goals, for every decision goal the preference of the decision alternatives for this goal is defined upon the set of decision alternatives by a simple linearly ordered preference ranking of the decision alternatives (for each particular goal). This means that it is not necessary to require from the decision maker that he or she keeps in mind all the interactions between the goals. We are able to calculate the interactions as implication of linearly ordered single goal preference rankings by applying Def. 2. accordingly as shown in section 3. The only additional information that we require from the decision maker are assertions about the importance of the goals, namely again only a simple (linear) ordering of the so-called priorities of goals which are real numbers from the interval [0,1] and express for every goal its importance in a particular decision situation (see sections 2.3 and 2.4).

5 Conclusions

In this paper it is discussed in which way a decision making approach based on interactions between goals is applied in order to use preference information that is not defined upon the power set of the set decision alternatives but simply on the decision set itself. Instead of defining the preferences upon the power set of the decision alternatives with respect to all the goals, for every decision goal the preference of the decision alternatives for this goal is defined upon the set of decision alternatives by a linear preference ranking of the decision alternatives. Although the input complexity is reduced, the decision quality is not. Therefore we conclude that classical preference relations defined upon the power set of the decision alternatives are not necessary in order to adequately handle complex decision problems. We have also discussed in which way the decision model based on interaction between decision goals can be related to decision making methods based on weighted sums only. Through a link to the theory of matroids it is shown that the model is more general in the sense that an adequate aggregation is possible even if the goals are partly conflicting whereas weighted sums are in this kind of situations rather restricted.

References

[1] Felix, R., "Relationships between goals in multiple attribute decision making", *Fuzzy Sets and Systems*, Vol.67, 47-52, 1994.

[2] Felix, R. (1998). “Decision-making with interacting goals”.In: *Handbook of Fuzzy Computation*, Ruspini, E., Bonissone, P.P, Pedrycz, W. (Eds.), IOP Publishing Ltd, 1998

[3] Felix, R. (2007). Real World Applications of a Fuzzy Decision Model Based on Relationships between Goals (DMRG). In: *Forging the New Frontiers, Fuzzy Pioneers I* (1965-2005), Springer Verlag in the series Studies in Fuzziness and Soft Computing, October 2007.

[4] Felix, R. (2008). Multicriterial Decision Making (MCDM): Management of Aggregation Complexity Through Fuzzy Interactions Between Goals or Criteria. In: *Proceedings of the 12th International IPMU Conference*, Málaga, Spain, 2008

[5] Modave, F., Dubois, D., Grabisch, M. and Prade, H., "A Choquet integral representation in multicriteria decision making", *AAAI Fall Symposium*, Boston, Ma, November 1997.

[6] Modave, F. and Grabisch, M., "Preference *representation by a Choquet integral: Commensurability hypothesis. In: Proceedings of the 7th International IPMU Conference*, Paris, France, 164-171, 1998.

[7] Oxley, J., “Matroid Theory”, Oxford University Press, 1992.

[8] Saaty, T. L., "The Analytic Hierarchy Process", Mc Graw-Hill, 1980.

[9] Torra, V., "Weighted OWA operators for synthesis of information", *Proceedings of the fifth IEEE International Conference on Fuzzy Systems*, New Orleans, USA, 966-971, 1996.

[10] Yager, R.R., "Families of OWA operators", *Fuzzy Sets and Systems*, Vol. 59, 125-148, 1993.

ISBN: 978-989-95079-6-8

A Fuzzy Real Option Valuation Approach to Corporate Investment Decisions

Elettra Agliardi[1] Maria Letizia Guerra[2] Luciano Stefanini[3]

1.Department of Economics, University of Bologna, Italy
2.Department Matemates, University of Bologna, Italy
3.Department of Economics and Quantitative Methods, University of Urbino, Italy
Email: elettra.agliardi@unibo.it, m.guerra@unibo.it, luciano.stefanini@uniurb.it

Abstract— Real options are a typical framework in economics that involves uncertainty. Very often, in fact, managers have vague ideas about the future expected cash flows, the cost of the project and many other variables that are fundamental in the process decision among many investments. The calculation of the value function of real options can take advantage of a model of uncertainty that include stochastic processes and fuzzy numbers.

A special version of the multiple population differential evolution algorithm is designed to compute the level-cuts of the fuzzy extension of the multidimensional real valued function of fuzzy numbers in the resulting optimization problems.

We perform some computational experiments about the option to defer investment, that is an American call option on the present value of the completed expected cash flows with the exercise price equal to the required outlay. We show that fuzziness may help for a more profitable decision.

Keywords— Fuzzy Numbers, Parametric Representation, Real Options, Sensitivity Analysis.

1 Introduction

Nowadays the most suitable valuation methodology for corporate investment decisions is the real options theory (ROT) because it takes into account management's flexibility to adapt ongoing projects in response to uncertain conditions. Since Myers' ([7]) innovative idea of viewing firm's future investment opportunities as real options – that is, the right but not the obligation to undertake some business decision at a cost during a certain period of time -, a vast literature has developed, which elaborate both theoretical and empirical methods for quantifying the values of various real (call or put) options embedded in investment opportunities. Dixit and Pindyck ([3]) develop a systematic treatment of ROT, providing the fundamentals of this method and also emphasizing the market implications of such valuation of investment decisions under uncertainty. Trigeorgis ([12]) provides a classification of real options that maps different categories of investments into the space of different types of financial options.

The value of real options depends on some basic variables:(i) the underlying asset, which is the current value of (gross) expected future operating cash flows, (ii) the exercise price, which is the cost of the project; (iii) the time to expiration of the option, that is the time up to which the project can be undertaken (either finite or infinite); (iv) the standard deviation of the value of the underlying risky asset; (v) the risk-free rate of interest over the life of the option. All the above-mentioned variables are uncertain and therefore various stochastic models have been introduced in ROT to deal with the uncertainty surrounding most corporate decisions. In general expected future cash flows are assumed to evolve according to a geometric Brownian motion, but it often happens that reality is more complex than a normal distribution.

The imprecision associated with the subjective judgement and estimation of future cash flows, which is typical of management's project decisions, needs to be incorporated in the treatment of uncertainty. By introducing fuzzy numbers, we are able to capture the somewhat vague and imprecise ideas the manager possesses about the future expected cash flows, the profitability of the project, the costs of the project, etc. To the best of our knowledge, such an approach has never been discussed in the literature, with the exception of Carlsson and Fuller [1], that interpret the possibility of making an investment decision in terms of a European option, while we use an American option. In addition, we elaborate a computing methodology which is more general and we can represent the shape of the value functions.

The paper is organized as follows. In section 2 we present some basic elements of fuzzy theory which will be used in the numerical implementation of real option models. In section 3 we describe the introduction of fuzziness in real options and section 4 collects some of the computational experiments that have been performed in order to capture how and how much fuzziness affects the decision. Finally, section 5 concludes.

2 Fuzzy Numbers and Parametric Representation

Fuzzy numbers (more details in [4]) are a very powerful and flexible way to describe uncertainty or possibilistic values for given variables for which a precise quantification is not possible or one is interested in evaluating the effects of variations around a specified value. In fact a fuzzy number models the different specifications of intervals around a given precise value; it is defined, informally, as a "cascade" of intervals, which start with a given number $a \in \mathbb{R}$ and grow increasing to a final interval which gives the most uncertain set of possible values. The levels of the cascade are usually parametrized by a parameter $\alpha \in [0,1]$ which represents the so called membership value (or possibilistic degree) of a given interval, with the convention that $\alpha = 1$ corresponds to the possibly exact value (the core of the fuzzy number) while $\alpha = 0$ corresponds to the highest uncertainty (the support of the fuzzy number). With the same convention on α we can say that $1 - \alpha$ is the level of uncertainty of the corresponding interval.

ISBN: 978-989-95079-6-8

A wide class of fuzzy numbers with the core at $a \in \mathbb{R}$ is obtained by considering its membership function $\mu : \mathbb{R} \longrightarrow [0,1]$ such that, denoting $[a^-, a^+]$ the interval representing the support,

$$\mu(x) = \begin{cases} L(x) & if \quad a^- \leq x \leq a \\ R(x) & if \quad a \leq x \leq a^+ \\ 0 & otherwise \end{cases} \quad \text{for } x \in \mathbb{R} \tag{1}$$

where $L(x)$ is the left branch, an increasing function with $L(a^-) = 0$, $L(a) = 1$ and $R(x)$ is the right branch, a decreasing function with $R(a) = 1$, $R(a^+) = 0$. A fuzzy number obtained by the form (1) is called LR-fuzzy number (with usual notation $u = \langle a^-, a, a^+ \rangle_{L,R}$).

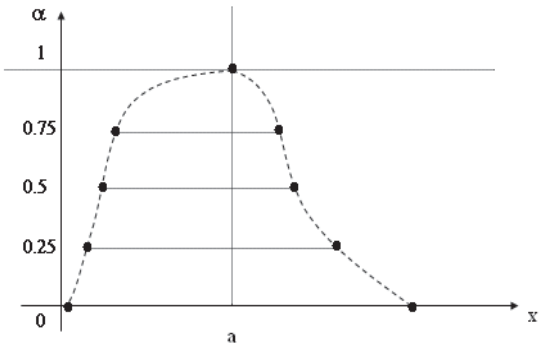


Figure 1: A fuzzy number as a "cascade" of intervals representing increasing uncertainty around the given value a.

For values of $\alpha \in]0,1]$, the $\alpha - cut$ is defined to be the compact interval $[u]_\alpha = \{x | \mu(x) \geq \alpha\}$ and the support is $[u]_0 = cl\{x | \mu(x) > 0\}$ ($cl(A)$ is the closure of set A). The *level-cuts* of a fuzzy number are "nested" closed intervals and this property is the basis for the LU representation (L for lower, U for upper).

Definition 1 *An LU-fuzzy quantity (number or interval) u is completely determined by any pair $u = (u^-, u^+)$ of functions $u^-, u^+ : [0,1] \longrightarrow \mathbb{R}$, defining the end-points of the $\alpha - cuts$, satisfying the three conditions:(i) $u^- : \alpha \longrightarrow u^-_\alpha \in \mathbb{R}$ is a bounded monotonic nondecreasing left-continuous function $\forall \alpha \in]0,1]$ and right-continuous for $\alpha = 0$;(ii) $u^+ : \alpha \longrightarrow u^+_\alpha \in \mathbb{R}$ is a bounded monotonic nonincreasing left-continuous function $\forall \alpha \in]0,1]$ and right-continuous for $\alpha = 0$;(iii) $u^-_\alpha \leq u^+_\alpha \ \forall \alpha \in [0,1]$. In particular, the $\alpha - cuts$ of a fuzzy number or interval are nonempty, compact intervals of the form $[u]_\alpha = [u^-_\alpha, u^+_\alpha] \subset \mathbb{R}$.*

The support of u is the interval $[u^-_0, u^+_0]$ and the core is $[u^-_1, u^+_1]$. We refer to the functions $u^-_{(.)}$ and $u^+_{(.)}$ as the lower and upper branches on u, respectively. The membership function can be written as $\mu_u(x) = \sup\{\alpha | x \in [u^-_\alpha, u^+_\alpha]\}$, where the left branch is the increasing inverse of $u^-_{(.)}$ on $[u^-_0, u^-_1]$ and the right is the decreasing inverse of $u^+_{(.)}$ on $[u^+_1, u^+_0]$.

To model the monotonic branches u^-_α and u^-_α we start with an increasing shape function p such that $p(0) = 0$ and $p(1) = 1$ and a decreasing function q such that $q(0) = 1$ and $q(1) = 0$, with the four numbers $u^-_0 \leq u^-_1 \leq u^+_1 \leq u^+_0$ defining the support $[u^-_0, u^+_0]$ and the core $[u^-_1, u^+_1]$ and we define

$$\begin{aligned} u^-_\alpha &= u^-_1 - (u^-_1 - u^-_0)p(\alpha) \text{ and} \\ u^+_\alpha &= u^+_1 - (u^+_1 - u^+_0)q(\alpha) \text{ for all } \alpha \in [0,1]\,. \end{aligned} \tag{2}$$

The two shape functions p and q, as suggested in [11], are selected in a family of parametrized monotonic functions where the parameters are related to the first derivatives of p and q in 0 and 1; there are many ways to define p and q as illustrated in [9]. For each decomposition we require (in the differentiable case) $4(N+1)$ parameters $u = (\alpha_i; u^-_i, \delta u^-_i, u^+_i, \delta u^+_i)_{i=0,1,...,N}$ satisfying the following conditions:

$$\begin{aligned} u^-_0 &\leq u^-_1 \leq ... \leq u^-_N \leq u^+_N \leq u^+_{N-1} \leq ... \leq u^+_0 \\ \delta u^-_i &\geq 0, \delta u^+_i \leq 0. \end{aligned} \tag{3}$$

and on each sub-interval $[\alpha_{i-1}, \alpha_i]$ we use the data $u^-_{i-1} \leq u^-_i \leq u^+_i \leq u^+_{i-1}$ and the slopes $\delta u^-_{i-1}, \delta u^-_i \geq 0$ and $\delta u^+_{i-1}, \delta u^+_i \leq 0$.

The simplest representation is obtained on the trivial decomposition of the interval $[0,1]$, with $N = 1$ (without internal points) and $\alpha_0 = 0, \alpha_1 = 1$. In this simple case, u can be represented by a vector of 8 components

$$u = (u^-_0, \delta u^-_0, u^+_0, \delta u^+_0; u^-_1, \delta u^-_1, u^+_1, \delta u^+_1). \tag{4}$$

In the search for the value of a real option, the fundamental step is the computation of fuzzy-valued functions. Given a function $y = f(x_1, x_2, ..., x_n)$ of n real (crisp) variables $x_1, x_2, ..., x_n$, its fuzzy extension is obtained to evaluate the effect of uncertainty on the x_j modelled by the corresponding fuzzy number u_j. If $v = f(u_1, u_2, ..., u_n)$ denotes the fuzzy extension of a continuous function f in n variables, then for each level α the resulting interval $[v^-_\alpha, v^+_\alpha]$ represents the propagation of uncertainty from x_j to y. In particular, if the uncertainty on the original variables is modelled by fuzzy numbers, the obtained v is yet a fuzzy number starting from a single value (at level $\alpha = 1$) to the most uncertain interval (at level $\alpha = 0$).

It is well known that the fuzzy extension of f to normal upper semicontinuous fuzzy intervals (with compact support) has the level-cutting property, i.e. the $\alpha - cuts$ $[v^-_\alpha, v^+_\alpha]$ of v are the images of the $\alpha - cuts$ of $(u_1, u_2, ..., u_n)$ and are obtained by solving the box-constrained optimization problems $(EP)_\alpha$:

$$\begin{cases} v^-_\alpha = \min \left\{ f(x_1, ..., x_n) | x_k \in [u^-_{k,\alpha}, u^+_{k,\alpha}],\ k = 1, ..., n \right\} \\ v^+_\alpha = \max \left\{ f(x_1, ..., x_n) | x_k \in [u^-_{k,\alpha}, u^+_{k,\alpha}],\ k = 1, ..., n \right\}. \end{cases} \tag{5}$$

With the exception of simple elementary cases for which the optimization problems above can be solved analytically, the direct application of *(EP)* may be difficult and computationally expensive. Usually, the $\alpha - cuts$ $[v^-_\alpha, v^+_\alpha]$ of v are computed at a prefixed given number of values α_j of interest (say from 10 to 100 points) and the membership function is approximated pointwise. As we will see, an advantage of the LU-parametrization is to obtain the extended fuzzy numbers $v = f(u_1, u_2, ..., u_n)$ in the same parametric form as for $u_1, u_2, ..., u_n$ with a possible important reduction in the computational effort and with a good approximation. As for the

ISBN: 978-989-95079-6-8

fuzzy numbers, we will consider the fuzzy extension of multivariate differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$.

For a vector of n fuzzy numbers $u = (u_1, u_2, ..., u_n)$ let

$$u_k = (u^-_{k,i}, \delta u^-_{k,i}, u^+_{k,i}, \delta u^+_{k,i})_{i=0,1,...,N} \text{ for } k = 1, 2, ..., n$$

be the LU representation of the $k-th$ component. Let $v = f(u_1, u_2, ..., u_n)$ and $v = (v^-_i, \delta v^-_i, v^+_i, \delta v^+_i)_{i=0,1,...,N}$ be its LU representation; the $\alpha - cuts$ of v are obtained by solving the box-constrained optimization problems (5). For each $\alpha = \alpha_i$, $i = 0, 1, ..., N$ the $\min$ and the $\max$ (5) can occur either at a point whose components $x_{k,i}$ are internal to the corresponding intervals $[u^-_{k,i}, u^+_{k,i}]$ or are coincident with one of the extreme values; denote by $\widehat{x}^-_i = (\widehat{x}^-_{1,i}, ..., \widehat{x}^-_{n,i})$ and $\widehat{x}^+_i = (\widehat{x}^+_{1,i}, ..., \widehat{x}^+_{n,i})$ the points where the $\min$ and the $\max$ take place; then

$$v^-_i = f(\widehat{x}^-_{1,i}, \widehat{x}^-_{2,i}, ..., \widehat{x}^-_{n,i}) \text{ and } v^+_i = f(\widehat{x}^+_{1,i}, \widehat{x}^+_{2,i}, ..., \widehat{x}^+_{n,i}) \tag{6}$$

and the slopes δv^-_i, δv^+_i are computed (as f is differentiable) by

$$\begin{aligned} \delta v^-_i &= \sum_{\substack{k=1 \\ \widehat{x}^-_{k,i} = u^-_{k,i}}}^{n} f'^-_k \delta u^-_{k,i} + \sum_{\substack{k=1 \\ \widehat{x}^-_{k,i} = u^0_{k,i}}}^{n} f'^-_k \delta u^+_{k,i} \\ \delta v^+_i &= \sum_{\substack{k=1 \\ \widehat{x}^0_{k,i} = u^-_{k,i}}}^{n} f'^+_k \delta u^-_{k,i} + \sum_{\substack{k=1 \\ \widehat{x}^0_{k,i} = u^0_{k,i}}}^{n} f'^+_k \delta u^+_{k,i} \end{aligned} \tag{7}$$

where $f'^-_k = \frac{\partial f(\widehat{x}^-_{1,i}, .., \widehat{x}^-_{n,i})}{\partial x_k}$ and $f'^+_k = \frac{\partial f(\widehat{x}^0_{1,i}, .., \widehat{x}^0_{n,i})}{\partial x_k}$.To solve the optimization problems (5), we use an implementation of a multiple population differential evolution algorithm extensively described and analyzed in [10].

- Dl ppleRi le fgkdfe kf PRSRh emRkd Rek

The *option to defer investment* is an American call option on the present value of the completed expected cash flows with the exercise price being equal to the required outlay. A project that can be postponed allows learning more about potential project outcomes before making a commitment. A seminal contribution on the option to defer is McDonald and Siegel [6] where the optimal time to invest and an explicit formula for the value of the option to invest are derived for an irreversible project whose net profits follow a geometric Brownian motion.

A firm is supposed to consider the following investment opportunity: at any time t the firm can pay some estimated cost K to install an investment project whose expected future net cash flows conditional on undertaking the project have an estimated present value Π. The installation of such project is irreversible. Let Π follow a geometric Brownian motion of the form:

$$d\Pi = \Pi(\mu dt + \sigma dW_t) \tag{8}$$

where $\mu < r$ is the appreciation rate, r is the risk-free interest rate and σ is the volatility ($\mu \in R, \sigma > 0$) and W is a standard Wiener process. For simplicity, let us assume that the time to expiration of this investment opportunity is infinite, which facilitates the derivation of a closed-form solution. If $V = V(\Pi)$ is the option value then it holds:

$$\frac{1}{2}\sigma^2 \Pi^2 V''(\Pi) + \mu \Pi V'(\Pi) - rV = 0$$

for $\Pi < \Pi^*$ with the initial condition $V(0) = 0$ and smooth-pasting $V(\Pi^*) = \Pi^* - K, V'(\Pi^*) = 1$. The solution is

$$\begin{cases} \Pi^* = K \frac{\phi}{\phi - 1} \\ V(\Pi) = (\Pi^* - K)\left(\frac{\Pi}{\Pi^*}\right)^{\phi} \end{cases} \tag{9}$$

with $\phi = \frac{1}{2} - \frac{\mu}{\sigma^2} + \left(\left(\frac{\mu}{\sigma^2} - \frac{1}{2}\right)^2 + \frac{2r}{\sigma^2}\right)^{\frac{1}{2}} > 1$.

In our methodology fuzziness is present in three steps. The estimated present value Π of future net cash flows of the project follows the stochastic differential equation (8) and we model the uncertainty of its parameters across intervals of values. The intervals are built with differentiated levels of uncertainty; given a crisp value, the levels produce a shape that can be characterized by asymmetries or nonlinearities depending on subjective beliefs and available information of the decision maker. It follows that fuzzy parameters play the lead role in a sensitivity analysis that starts gradually from a null variation to the greatest variation of the uncertainty consistent with data. In particular W_t remains a standard Brownian motion (the theory and applications of fuzzy set-valued stochastic differential equations has received recently several significant contributions from the seminal paper of Feng [5]).

Fuzziness comes out also in the valuation function of the option (obtained with the extension principle) that depends not only on Π, σ and μ but also on r and K, which we assume to be fuzzy too.

Finally, fuzziness affects the crucial value Π^*: as soon as Π reaches the threshold value Π^*, the firm finds it optimal to invest (case of the option to defer investment) or disinvest and liquidate (case of the option to abandon). Thus, the decision is based on the threshold value, which depends on all the parameters of the model. In the valuation method based on fuzzy variables, $\{\Pi_t, t \geq 0\}$ is assumed to be a fuzzy stochastic process, which is specified by the following membership function

$$\mu_{\Pi_t(\omega)}(x) = max\{1 - \mid (x - \widehat{\Pi}_t(\omega))/\beta_t(\omega) \mid, 0\},$$

that is, the fuzzy random variable Π_t is of the triangular type, with centre $\widehat{\Pi}_t(\omega)$, and left-width and right-width $\beta(\omega)$. The assumption of fuzziness is related to the manager's subjective belief about the future profitability of the project. The choice of a triangle-type shape is not restrictive at all and is introduced for simplicity only. Observe that the fuzziness in the process increases as $\beta(\omega)$ becomes bigger. The α-cuts of $\Pi_t(\omega)(x)$ are $\Pi^{\pm}_{t,\alpha}(\omega) = \left[\Pi^-_{t,\alpha}(\omega), \Pi^+_{t,\alpha}(\omega)\right] = \left[\widehat{\Pi}_t(\omega) - (1-\alpha)\beta(\omega), \widehat{\Pi}_t(\omega) + (1-\alpha)\beta(\omega)\right]$. It is also reasonable to assume that K is a fuzzy number. In the case of an option to defer, K is the estimated liquidation value of the firm's stock of capital and is affected by depreciation, fluctuating market evaluation and taxation regimes. In the case of an option to abandon, K denotes the investment cost and has many components which can change during the waiting period, due to various unpredictable circumstances.

ISBN: 978-989-95079-6-8

The extension principle is then applied to obtain the fuzzy Π^* and $V(\Pi^*)$ from the exact solutions given in equation (9). In the formulae (6)-(7) the vector $\widehat{x}_i$ is equal to $(\widehat{\mu}_i, \widehat{\sigma}_i, \widehat{r}_i, \widehat{K}_i)$ and some of the partial derivatives that define the slopes of the representation are nothing else than the first order Greeks, in particular, $\frac{\partial f(\widehat{\mu}_i,\widehat{\sigma}_i,\widehat{r}_i,\widehat{K}_i)}{\partial \sigma}$ is the Vega and $\frac{\partial f(\widehat{\mu}_i,\widehat{\sigma}_i,\widehat{r}_i,\widehat{K}_i)}{\partial r}$ is the Rho.

The degree of the uncertainty and the way in which it is spread from the model, play a central role in the analysis of the real option. The nonlinearities entering in the definition of $V(\Pi)$ in (9) are the main cause of such effects and they can propagate or contract uncertainty. It is very important to perceive the magnitude and the type of these effects. In particular we are interested in the analysis of how the various kinds of uncertainties inserted into the parameters will produce the corresponding uncertainties in Π^*, $V^* = V(\Pi^*)$, Π^{**} and $V^{**} = V(\Pi^{**})$.

As soon as information (on μ, σ, r, K) is modelled by fuzzy numbers, Π^* and V^* also become fuzzy and are represented by $\alpha - cuts$ $\left[\Pi_\alpha^{*^-}, \Pi_\alpha^{*^0}\right]$ and $\left[V_\alpha^{*^-}, V_\alpha^{*^0}\right]$ for all degrees of possibility α. The maximal uncertainty corresponds to the supports at $\alpha = 0$, given by the intervals $\left[\Pi_0^{*^-}, \Pi_0^{*^0}\right]$ and $\left[V_0^{*^-}, V_0^{*^0}\right]$ for Π^* and V^* respectively.

Due to the nonlinearity of Π^* and V^*, the $\alpha - cuts$ are not necessarily symmetric and, for a given uncertainty on the input values μ, σ, r and K, they have different left and right variations. Let $\widehat{\Pi}^*$ and $\widehat{V}^*$ denote the values of Π_α^* and V_α^* corresponding to $\alpha = 1$. It is immediate to argue that V^* is symmetric if and only if $\Delta V_\alpha^{*^0} = \Delta V_\alpha^{*^-}, \forall \alpha \in [0, 1[$ where

$$\Delta V_\alpha^{*^0} = V_\alpha^{*^0} - \widehat{V}^*, \quad \Delta V_\alpha^{*^-} = \widehat{V}^* - V_\alpha^{*^-}.$$

The quantity $\Delta V_\alpha^{*^0}$ represents the possible increase in $\widehat{V}^*$ due to uncertainty and analogously, $\Delta V_\alpha^{*^-}$ measures the possible decrease. The same argument can be applied to Π_α^* and $\widehat{\Pi}^*$, defining the quantities $\Delta\Pi_\alpha^{*^0} = \Pi_\alpha^{*^0} - \widehat{\Pi}^*$ and $\Delta\Pi_\alpha^{*^-} = \widehat{\Pi}^* - \Pi_\alpha^{*^-}$.

. 6 f d gl kHkdf e HcRngRhdl Re ki

We test the fuzziness effect in the option to defer investment by running several computational experiments; the fuzzy version of the indicated parameters (say θ) are obtained as triangular symmetric fuzzy numbers, centered at the crisp values and with the support being the interval $[\theta - 0.1\theta, \theta + 0.1\theta]$, corresponding to a symmetric uncertainty of 10% of the value of the parameter. To analyze the effect of the uncertainty on the output variables, we give the plots of their membership functions and the tables including the values for $\alpha - levels$ with $\alpha = 1$ (the crisp level), $\alpha = 0.75$, $\alpha = 0.5$, $\alpha = 0.25$, $\alpha = 0$ (corresponding to the uncertainty of 10% in the parameters).

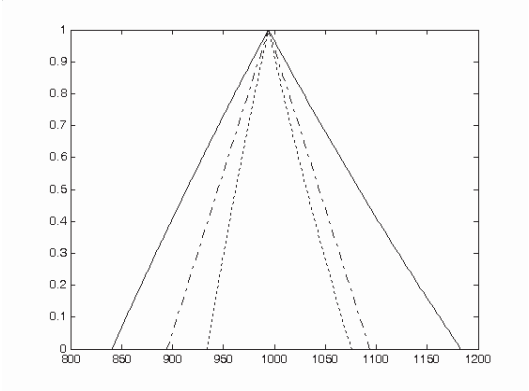
The robustness of the fuzzy model for the option to defer investment is tested with three sets of real data that we call, for short, Data1, Data2 and Data3, referring to three different industrial sectors. *Data1* refers to an investment decision in the human genome sciences project (HGSI) whose data are taken from the Human Genome project database (details in http://www.ornl.gov/sci/ techresources/ Human_Genome/ home.shtml). *Data2* refers to an investment decision in a big infrastructure, that is the Eurotunnel project (details in [2]). *Data3* deals with the case of an investment in new capacity in the public-utility sector, i.e. the electricity market (details in [8]). The values of the parameters μ, σ, r and K are the following:

	: HkH(	: HkH)	: HkH-
μ	0.01	0.025	0.03
σ	0.048	0.183	0.173
r	0.044	0.06	0.08
K	704.9	8865	600

In figures concerning the behavior of Π^* we report the three different cases that we will denote as: *Allfuzzy* (straight line) when the parameters μ, σ, r and K are fuzzy, *Kcrisp* (dotted line) when μ, σ, r are fuzzy and K is crisp and finally *Kfuzzy* (dashed line) when μ, σ, r are crisp and K is the unique source of uncertainty.

4.0.1 Results for Data1

Figure 2 shows that, as expected, the greatest uncertainty in Π^* occurs in the Allfuzzy case, when all the fuzzy quantities are considered to be fuzzy; but it is interesting to observe that in the Kcrisp case (dotted line) the generated uncertainty is less then in the Kfuzzy case (dashed line), i.e. the uncertainty in the values of only K produces more uncertainty on Π^* then the uncertainty in the values of μ, σ and r.



D5l1 hR) 4 Π^* for Data1

Table 1, Table 2 and Table 3 report values of the $\alpha - cut$ for Data1 in the Allfuzzy, Kcrisp and Kfuzzy case respectively.

Table 1

α	Π^-	Π^+
1.0	994.28	994.28
0.75	953.16	1037.56
0.5	913.97	1083.26
0.25	876.50	1131.69
0	840.58	1183.25

Table 2

Π^-	Π^+
994.28	994.28
977.60	1012.25
962.07	1031.67
947.56	1052.74
933.98	1075.68

Table 3

α	Π^-	Π^+
1.0	994.28	994.28
0.75	969.43	1019.14
0.5	944.57	1043.99
0.25	919.71	1068.86
0	894.86	1093.71

ISBN: 978-989-95079-6-8

Observe that in the Kfuzzy case (Table 3) the threshold value Π^* displays a symmetric shape in all analyzed projects because Π^* depends linearly on K. In the Allfuzzy and Kcrisp cases, instead, we can observe an asymmetric pattern, due to the nonlinear dependence of Π^* with respect to the other variables.

At level 0.5 the average values are 998.615 in Allfuzzy and 996.87 in Kcrisp, which are larger than the crisp value 994.28. Since on average the fuzzy threshold value is larger then without fuzziness, just considering the crisp value the decision to invest would be too early. Figure 3 shows the graphical behavior of the fuzzy function $V(\Pi)$ in the Allfuzzy case; the little crosses point the optimal values of Π corresponding to the levels of Π^* for $\alpha = 0, 0.25, 0.5, 0.75, 1$.

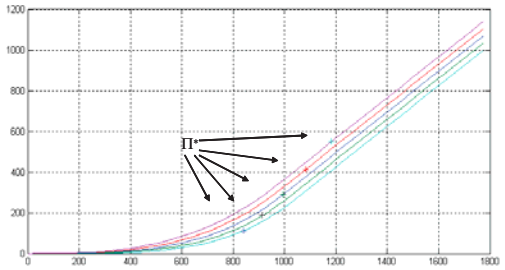


Figure 4: Data1 when parameters μ, σ, r and K are fuzzy.

It is evident that fuzziness implies a certain degree of freedom in the choice of Π^*. Figure 4 illustrates $V(\Pi)$ as a fuzzy function (a sequence of fuzzy numbers).

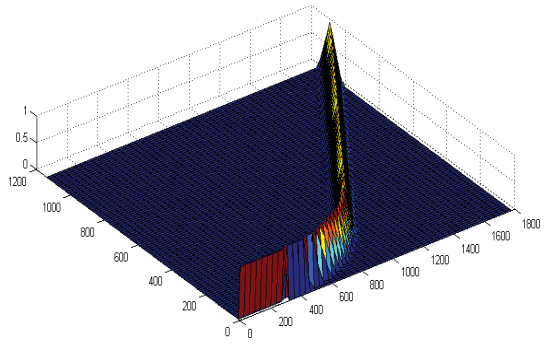


Figure 5: $V(\Pi)$ in the Allfuzzy case

4.0.2 *Results for Data2*

Figure 5 shows the behavior of Π^* in the three different cases: again the biggest uncertainty occurs in the Allfuzzy case, when all the quantities are fuzzy; but we observe that in the Kcrisp case (dotted line), when μ, σ, r are the sources of uncertainty and K is the unique crisp value, the uncertainty in Π^* is bigger then in the Kfuzzy case, i.e. the same level of uncertainty in K produces less uncertainty on Π^* then the uncertainty in the other parameters. With respect to Data1, there is here an inversion.

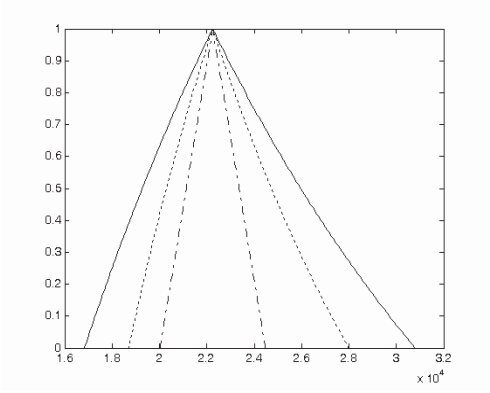


Figure 6: Π^* for Test 2

Table 4 and Table 5 report the values of the $\alpha - cut$ of Π^* in the Allfuzzy and Kcrisp case for Data2.

Table 4:

α	Π^-	Π^+
1.0	22249.62	22249.62
0.75	20686.18	23997.55
0.5	19277.60	25967.74
0.25	18000.33	28209.16
0	16835.43	30786.48

Table 5:

α	Π^-	Π^+
1.0	22249.62	22249.62
0.75	21216.59	23412.24
0.5	20292.21	24731.18
0.25	19459.82	26241.08
0	18706.04	27987.71

If we compute again the average values at level 0.5, they are 22622.67 in Allfuzzy and 22511.695 in Kcrisp, which are larger than the crisp value 22249.62. It follows that in the Data2 project it is confirmed the suggestion to wait for the decision to invest.

4.0.3 *Results for Data3*

The last project we consider for an option to defer investment is Data3; the relative values of Π^* are reported in Figure 6:

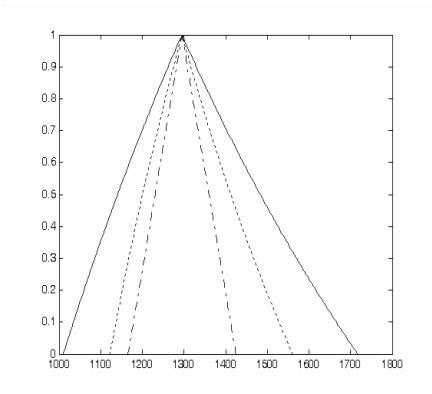


Figure 7: Π^* for Test4

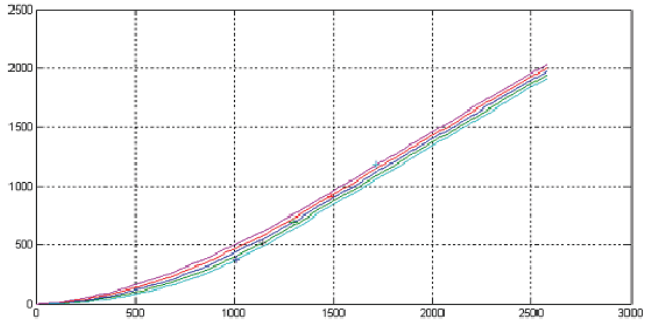
ISBN: 978-989-95079-6-8

Table 6, Table 7 report values of the $\alpha - cut$ for Data3 in the Allfuzzy and Kcrisp case respectively.

	GHIdR2	
α	Π^-	Π^+
1.0	1295.31	1295.31
0.75	1214.47	1384.31
0.5	1140.61	1482.98
0.25	1072.75	1593.18
0	1010.11	1717.32

GHIdR3	
Π^-	Π^+
1295.31	1295.31
1245.61	1350.55
1200.64	1412.36
1159.73	1482.03
1122.34	1561.20

The graphical representation of $V(\Pi)$ in Allfuzzy case is in Figure 7.



D5l1 hR34 Data3 in the Allfuzzy case

Some further considerations concerning the $\alpha - cut$ values in all the data set enable us to state that our model allows us to describe how the investment decision is actually affected by a perceived increase in "fuzziness". For a pessimistic (optimistic) firm an increase in fuzziness decreases (increases) the perceived value of the project in comparison with the crisp value. On average - for most decision makers- an increase in fuzziness has a positive impact on the investment opportunity, i.e. it increases the perceived value of the project. As a consequence, the decision to invest is delayed in comparison with the absence of fuzziness. However, for pessimistic decision-makers imprecise information about the project value becomes available over time, which makes waiting with investment less valuable. Thus, for pessimistic firms higher fuzziness erodes the subjective value of the investment opportunity. Notice that this result is in keeping with the literature on real options and ambiguity aversion it contrasts with the impact of volatility in the standard real option theory.

1 6feNlilfei

The powerful contribution of fuzzy modelling can be shown in many fields and, especially in human sciences like economics, it can provide rigorous models (a detailed motivation of its use is in Zadeh [13]). In tis paper, we model the uncertainty involved in real options theory through fuzzy numbers represented in the LU model; when including fuzziness, the decision rule moves away from the original one and the choice to delay or not the investment becomes a key feature of the fuzzy model.

Our model allows us to quantify how the threshold values change with fuzzy parameters. Practically, we solve the investment/disinvestment decision in terms of a problem of fuzzy stopping time - that is, a fuzzification of the classical optimal stopping time - and our computation experiments provide the sensitivity analysis of the decision variable with respect to the relevant parameters. Our main results are that the decision to invest is delayed in comparison with the case of absence of fuzziness, the decision to disinvest is anticipated in a fuzzy environment. The fuzzy model we use provides the set of values within which the decision is taken by the manager. Our empirical validation, which is based on case studies from real world, seems to confirm the validity of the model and to open up further ways of research in finance in a fuzzy environment.

F RSRhRe NRt

[1] C. Carlsson, R. Fullér, A Fuzzy Approach to Real Option Valuation, *Fuzzy Sets and Systems*, 139, (2003), 297-312.

[2] A. Damodaran, The Promise and Peril of Real Options, Working Paper, Stern School of Business, 1999.

[3] A. Dixit, R. Pindyck , *Investment under Uncertainty*, Princeton University Press, (1994), Princeton.

[4] D. Dubois, H. Prade (ed), *Fundamentals of Fuzzy Sets*, Kluwer, Boston, The Handbooks of Fuzzy Sets Series, 2000.

[5] Y. Feng, Fuzzy stochastic differential systems. *Fuzzy Sets and Systems*, 115, 2000, 351–363.

[6] R. McDonald, D. Siegel, The Value of Waiting to Invest, *The Quarterly Journal of Economics*, 101, 4 1986, 707-728.

[7] S. C. Myers, Determinants of Corporate Borrowing, *Journal of Financial Economics*, 1977, 147-76.

[8] F.A. Roques, N.S. Savva, Price Cap Regulation and Investment Incentives under Demand Uncertainty, Cambridge Working Papers in Economics 0636, 2006.

[9] L. Stefanini, L.Sorini, M.L.Guerra, Parametric Representations of Fuzzy Numbers and Applications to Fuzzy Calculus, *Fuzzy Sets and Systems*, 157/18, 2006, 2423-2455.

[10] L. Stefanini, Differential Evolution Methods for the Fuzzy Extension of Functions, Working Papers 0705, University of Urbino Carlo Bo, Department of Economics, 2007.

[11] L. Stefanini, L. Sorini, M.L. Guerra, Fuzzy Numbers and Fuzzy Arithmetics, in W. Pedrycz, A. Skowron, and V. Kreinovich (Eds.): *Handbook of Granular Computing*, John Wiley & Sons, 2008, 249-283.

[12] L. Trigeorgis, Real options and Interactions with Financial Flexibility, *Financial Management*, 1993, 202-24.

[13] L.A. Zadeh, Is there a need for fuzzy logic?, *Information Sciences*, 178, 2008, 2751-2779.

ISBN: 978-989-95079-6-8

A dynamic classification method for the discrimination of evolving data

Laurent Hartert Moamar Sayed Mouchaweh Patrice Billaudel

Université de Reims Champagne-Ardenne, CReSTIC,
Moulin de la Housse, BP 1039, 51687 Reims Cedex, France
Email: laurent.hartert@univ-reims.fr, moamar.sayed-mouchaweh@univ-reims.fr, patrice.billaudel@univ-reims.fr

Abstract—*Classes issued of evolving systems are dynamic and their characteristics vary over the time. Assigning a pattern to a class is achieved using a classifier. Therefore, the classifier parameters must be adapted online in order to take into account the temporal changes in the classes' characteristics. This adaptation is based only on the recent and useful information carried out by the new incoming classified patterns. In this paper, we propose to develop the classification method Fuzzy Pattern Matching (FPM) to be operant in the case of dynamic classes. This development is based on the use of an incremental algorithm to follow the accumulated gradual temporal changes of classes' characteristics after the classification of each new pattern. When these changes reach a suitable predefined threshold, the classifier parameters are adapted online using the recent and representative patterns.*

Keywords— Classification, dynamic patterns, evolving systems, pattern recognition.

1 Introduction

Dynamic systems evolve between different normal or faulting functioning modes in the course of time. In statistical Pattern Recognition (PR) [12, 9], observations about system functioning modes are divided into groups of similar patterns, called classes, using an unsupervised learning method [10, 4] or human experience. These patterns, with their class assignments, constitute the learning set. They are represented by a set of *d* features, or attributes, so they can be viewed as *d*-dimensional features vectors, or points, in the feature space. A supervised learning method [9, 16] uses the learning set to build a classifier that best separates the different classes in order to minimize the misclassification error. The model of each class can be represented by a membership function which determines the membership value of a pattern to a class. Then, new incoming patterns are assigned to the class for which they have the maximum membership value. In supervised learning methods, the membership function can be generated using Probability Density Function (PDF) estimation based methods or heuristic based ones.

Patterns describing the system functioning can be static or dynamic. A static pattern is represented by a point in the feature space while a dynamic pattern is represented by a multidimensional trajectory. In this case, the feature space has an added dimension which is the time [3]. Classes can also be static or dynamic. Static classes are represented by restricted areas formed by similar static patterns in the feature space. Hence, the way in which patterns occur is irrelevant to their membership values. Therefore, the classifier's parameters remain unchanged with the time. However, data issued from evolving processes are non stationary. In this case, classes become dynamic and their characteristics change in the course of time. Thus, the classes' membership functions must be adapted to take into account these temporal changes. This requires an adaptive classifier with a mechanism for adjusting its parameters over the time. Hence, some of the new incoming points reinforce and confirm the information contained in the previous data, but the other ones could bring new information (creation, drift, fusion, splitting of classes, etc.). This new information could concern a change in operating conditions, development of a fault or simply more significant changes in the process dynamic.

The general principle of dynamic PR methods [14, 3, 7, 13] is to observe the change of some statistical properties of classes, in order to decide in which state the system is: unchanged, gradually changed or completely changed. Thus, the classifier parameters, i.e. the membership functions, will be respectively unchanged, slightly adapted, or relearned from scratch.

In the literature, the membership functions are generally adapted using two approaches [2, 3, 14]. The first one acts directly on the classifier parameters, by substituting or adding some recent and representative patterns to the learning set [3, 14], according to the state in which the system is. This adaptation is based only on the most recent batch of patterns selected by one of the two following methods. The first method uses a time window, with a fixed or a variable size, which permits to reduce or limit the growing size of the database by accepting the *n* most recent patterns [14, 17]. The size of the time window must be well-chosen to obtain a compromise between a fast adaptation and a sufficient number of representative patterns. The second method is based on the use of a template containing a fixed number of selected patterns according to their age and usefulness [11]. Nevertheless, it is subjective and difficult to estimate the utility of patterns. The second approach to adapt the membership functions is based on the use of evolving neural networks [1, 2, 7]. In [2], a potential function based on the distance between data points is defined for the new points. The first data point potential is considered as equal to 1 and it establishes the first neuron (or rule) which is considered as the prototype (or centre) of the first cluster. Then, the next new data points may possess a potential close or greater than the one of the prototype neuron. This point can reinforce or confirm the information contained in the previous ones, or if the point is more informative than the data used as prototype, a new neuron (new rule) is added. In [1], the neural network is based on a multiprototype Gaussian modeling of non convex classes. The activation

ISBN: 978-989-95079-6-8

function of each hidden neuron determines the membership degree of an observation to one prototype of a class. With the first acquisition, the network is initialized there is creation of the first prototype constituting the first class. The prototype is parameterized by its centre and an initial covariance matrix. Then, according to the membership degree of new acquisitions, the prototype (the hidden neuron) can be adapted, eliminated or a new prototype can be created.

Dynamic classification methods are used to solve several real problems. The application of [3] concerns the credit-scoring which aims to decide whether a new customer is a good or a bad risk according to changes in his consumption. In [1], the authors aim to detect and to follow up the progressive evolution of the functioning mode from normal to faulty one of a thermal regulator. This evolution is due to the age of the system's components or to other temporal factors in its environment. In [13], the author looks to estimate a ratio between the number of good parts in a completed batch of microelectronic chips and the number of parts obtained from the same batch if there is no default. Due to high complexity and variability of modern microelectronics manufacturing industry, this ratio is affected daily by hundreds of material-related, equipment-related and human-related factors. In [7], dynamic traffic data streams are treated in order to reduce the waiting time of drivers at the road intersections.

We use the classification method Incremental Fuzzy Pattern Matching (IFPM) [15] which is an incremental version of FPM [6]. The membership functions are based on the estimation of the marginal class probability density functions using histograms. Then, the membership functions are adapted incrementally after the classification of each new pattern. IFPM provides good result in the case of static classes based on static patterns. This method is simple and has a low and constant classification time according to the size of the database. However, IFPM has no mechanism to forget the no more useful patterns.

In this paper, we propose to develop IFPM to be operant in the case of dynamic classes. This development is called Dynamic FPM (DFPM). The paper is structured as follows. The two first parts are devoted respectively to the functioning of IFPM and DFPM. In the third part, the limits of IFPM and results of DFPM are presented in the case of dynamic classes. The last part concludes the paper and presents the perspectives of our future work.

2 Incremental Fuzzy Pattern Matching

2.1 Principle

The functioning of Incremental FPM (IFPM) involves the learning, the classification, and the incremental update phases.

2.1.1 Learning phase

Let X_i be the set of learning points belonging to the class C_i. Each class C_i contains N_i points, or patterns, in the feature space, $\Re^d$, formed by d attributes. X is the learning set, containing N points x belonging to C classes, so $X = \bigcup_{i=1}^{C} X_i$. The learning phase consists in building, based on X, a decision rule characterized by a set of membership functions. These latter are based on histograms allowing to estimate the conditional probability density of each class C_i according to each attribute j. The number h of bins b_k, $k \in \{1,2,..., h\}$ of a histogram is determined experimentally. The choice of h conditions the performances of IFPM. The lower $\min^j$ and upper $\max^j$ borders of each attribute are determined manually by an expert. In the case of dynamic classes, these borders must be well defined to contain all patterns of the classes' evolutions. This constitutes a shortcome of IFPM when the classes are dynamic. The width Δ^j of a bin according to the attribute j is defined by:

$$\Delta^j = \frac{(\max^j - \min^j)}{h} \tag{1}$$

The histogram or the distribution of probability $\{p_i^j(b_{ik}^j), i \in \{1,2,...,c\}, j \in \{1,2,...,d\}, k \in \{1,2,...,h\}\}$ for the class C_i according to the attribute j is determined by calculating the probability $p_i^j(b_{ik}^j)$ of each bin b_{ik}^j:

$$p_i^j(b_{ik}^j) = \frac{n_{ik}^j}{N_i} \tag{2}$$

where n_{ik}^j is the number of learning points of the class C_i which are located in the bin b_{ik}^j and N_i is the total number of points of the class C_i. Then, the resulting distribution of probability is transformed into a distribution of possibility $\{\pi_i^j(b_{ik}^j), i \in \{1,2,...,c\},\ j \in \{1,2,...,d\}, k \in \{1,2,...,h\}\}$ by using the transformation of Dubois and Prade [8]:

$$\pi_i^j(b_{ik}^j) = \sum_{z=1}^{h} \min(\ p_i^j(b_{iz}^j),\ p_i^j(b_{ik}^j)) \tag{3}$$

The possibility measure has the advantage to take into account the imprecision and the uncertainty contained in the data [8]. Finally, the density of possibility Π_i^j of the class C_i according to the attribute j is obtained by a linear interpolation of the bins centers of the histogram of possibility.

2.1.2 Classification phase

A pattern x is assigned to a known class C_i using the following three steps:

- determination of the membership possibility value π_i^j of x^j to each class C_i according to each attribute j. This possibility is obtained by projection of x^j on the density Π_i^j of each class C_i according to each attribute j (Fig. 1).

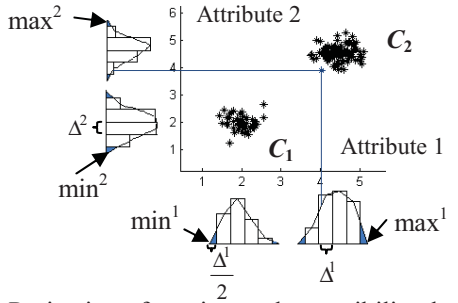


Figure 1: Projection of a point on the possibility densities. Δ^1 and Δ^2 are the bins' widths according respectively to attributes 1 and 2.

ISBN: 978-989-95079-6-8

- fusion, for each class C_i, of all membership possibility values $\pi_i^1, \pi_i^2, ..., \pi_i^d$ by the aggregation operator "minimum". The result of this fusion represents the membership possibility π_i of x to each class C_i,
- classification of the point x to the class for which it has the highest membership value.

2.1.3 *Incremental update phase*

Each new classified pattern carries information about the temporal changes of the class' characteristics. In order to take into account this new information, the membership functions must be relearned after the classification of each new pattern. An incremental learning algorithm is proposed in [15] to relearn the membership functions based on the most recently classified pattern in such a way that the previous learned information is reused. When a new point is classified in the class C_i, the number of points belonging to this class becomes $N_i + 1$, so the probability of each bin changes. If the point is located in the bin b_{ik}^j, the new probability $p_i'^j(b_{ik}^j)$ of this bin is [15]:

$$p_i'^j(b_{ik}^j) = \frac{n_{ik}^j}{N_i} \times \frac{N_i}{N_i+1} + \frac{1}{N_i+1} = p_i^j(b_{ik}^j) \times \frac{N_i}{N_i+1} + \frac{1}{N_i+1} \quad (4)$$

For each other bin, the new probability is:

$$p_i'^j(b_{iz}^j) = \frac{n_{iz}^j}{N_i} \times \frac{N_i}{N_i+1} = p_i^j(b_{iz}^j) \times \frac{N_i}{N_i+1}, z=1...h \ \& \ z \neq k \quad (5)$$

Using (4) and (5), the probability histograms can be updated after the classification of a new point without the need to recalculate them. We just need to know in which bin the new point is located. The histogram of possibility is then calculated by (3). IFPM has been developed to deal with data represented in a feature space of correlated attributes as the case of XOR database [5]. In the next part, the improvements realized on IFPM are presented.

3 Dynamic Fuzzy Pattern Matching

We have called Dynamic Fuzzy Pattern Matching (DFPM) the improved version of IFPM. DFPM integrates a mechanism to adjust the classifier's parameters when serious changes in the classes' characteristics are detected during a growing time window. The size of the time window depends of the application's dynamic. The next subsections detail the two phases of DFPM's functioning.

3.1 *Detection phase*

In the detection phase, three indicators are used to monitor the temporal changes of a system. The first indicator is a measure evaluating the bins' usefulness. This indicator is based on the accumulated temporal changes which have occurred in the probability distributions during a variable time sliding window. It considers the number of new points located in each bin of the probability histograms, as well as its initial probability before the assignment of the new pattern(s). Let $\Delta T = t_w - t_s$ be the growing size of the time window at the present time t_w. This time window starts at the instant t_s. Let $\{p_i^j(b_{ik}^j), k \in \{1,2,..., h\}\}$ be the probability histogram for the class C_i according to the attribute j at the instant t_s. The probability of each bin is calculated by (2). Let ΔN_i be the number of new classified points in the class C_i and let Δn_{ik}^j be the number of new points located in the bin b_{ik}^j at the present time t_w. The new probability of this bin is calculated by:

$$p_i'^j(b_{ik}^j) = \frac{n_{ik}^j + \Delta n_{ik}^j}{N_i + \Delta N_i} \quad (6)$$

where n_{ik}^j and N_i are respectively the number of points of the class C_i located in the bin b_{ik}^j and the total number of points of the class C_i before the classification of the new pattern(s). The accumulated temporal changes for each bin are evaluated as the difference between the bin's probabilities at the present time $p_i'^j(b_{ik}^j)$ and at the beginning $p_i^j(b_{ik}^j)$ of the growing time window:

$$\Delta p_i^j(b_{ik}^j) = p_i'^j(b_{ik}^j) - p_i^j(b_{ik}^j) = \frac{N_i.\Delta n_{ik}^j - n_{ik}^j.\Delta N_i}{N_i.(N_i + \Delta N_i)} \quad (7)$$

We define the usefulness measure as a gradual measure, $I_i^j(b_{ik}^j) : \{1,2,...,h\} \rightarrow [-1,1]$ for each bin b_{ik}^j of each class C_i according to each attribute j. This measure evaluates the usefulness of each bin based on the accumulated temporal changes in the probability distributions defined by (7):

$$I_i^j(b_{ik}^j) = \begin{cases} \dfrac{p_i'^j(b_{ik}^j) - p_i^j(b_{ik}^j)}{p_i^j(b_{ik}^j)} \in [-1,0] \ if \ \Delta p_i^j(b_{ik}^j) < 0 \\ \dfrac{p_i'^j(b_{ik}^j) - p_i^j(b_{ik}^j)}{p_i'^j(b_{ik}^j)} \in [0,1] \ if \ \Delta p_i^j(b_{ik}^j) > 0 \end{cases} \quad (8)$$

This measure is calculated after the classification of each new classified point. When the value of this indicator is negative for one bin, it indicates a decrease of its usefulness. This decrease is due to the fact that no new classified points, or just a few ones, are located in this bin. On the contrary, the positive values indicate an increase of the bin's usefulness due to the assignment of many new points in the bin, according to its initial value. It is interesting to take into account the old or initial probabilities. Indeed, if there is no important change in the probability distribution, it is normal that the bin with the highest probability receives more new points than the others. If most of the new classified points are located in the bins which possess a small initial probability, this indicates a serious change in the probability distribution and thus their usefulness indicators must have a more important positive value.

To detect an evolution, a second indicator, which is a residual, is calculated based on the usefulness' measure of all histograms' bins according to all attributes. The goal of this residual is to provide one indicator for each class, which evaluates the temporal accumulative changes in the probability distributions of this class. This residual is sensitive for these changes. Two thresholds *th*1 and *th*2 predefined by an expert according to the system dynamic are used to determine the size ΔT of the time window. The value of this residual is updated and compared after each pattern assignment to *th*1 and *th*2. If the residual value is

ISBN: 978-989-95079-6-8

inferior to *th*1, then no serious change in the class' characteristics is yet started. As long as the residual is inferior to *th*1, the measure of usefulness is based on the difference between the current bin probability and the one of the previous classified pattern. If the residual value is superior to *th*1 and inferior to *th*2, a serious change has begun to occur. The time window starts at this moment, $t=t_s$, and patterns are stored in a block, called the evolution block. During this window, the measure of usefulness is calculated using (8). Patterns are stored in the evolution block until the residual value reaches *th*2. At this moment the growing time window reaches its final size $t_w=t_f$. This residual is defined for each class C_i as follows:

$$0 \le R_i = \max\left(\frac{\sum_{k=1}^{h}\left|I_i^j(b_{ik}^j)\right|}{h}\right) \le 1, j \in \{1,..,d\} \quad (9)$$

When the residual value reaches *th2*, a serious change can be decided in a class so the classifier parameters (densities of possibility) must be adapted using only the patterns stored in the evolution block. The old patterns which have become obsolete are forgotten and an online update of the classifier's parameters can then be performed.

3.2 Adaptation phase

This phase consists in the adaptation of the probability histograms and the update of the lower and upper attributes' borders as well as the number of bins *h*. Indeed, in IFPM, these latter are defined initially using (1) and they remain unchanged. For dynamic classes, it is not realistic to define static attributes' borders. Thus, in DFPM these borders will be updated online with the classification of new points. If a new classified point involves the creation of a bin for a histogram, then a bin is added to the histogram according to this attribute so that *h* is incremented and the histograms' borders are updated:

$$\begin{cases} x^j < \min^j - \frac{\Delta^j}{2} \Rightarrow \min'^j = \min^j - \Delta^j \\ x^j > \max^j + \frac{\Delta^j}{2} \Rightarrow \max'^j = \max^j + \Delta^j \end{cases} \quad (10)$$

$$h' = h + 1 \quad (11)$$

The value of Δ^j is defined initially by (1). With this update *h* can be different according to each attribute. The adaptation of probability histograms is achieved using an incremental approach. Let $\left\{p_i^j(b_{ik}^j), k \in \{1,2,...,h\}\right\}$ be the probability histogram for the class C_i according to the attribute *j*. When the residual value of a class reaches *th*2, old patterns are deleted and only the new patterns of the evolution block are considered. So, the value of probability for each bin is calculated incrementally by:

$$p_i''^j(b_{ik}^j) = \frac{\Delta n_{ik}^j}{\Delta N_i} = \left(\frac{n_{ik}^j + \Delta n_{ik}^j}{N_i + \Delta N_i} - \frac{n_{ik}^j}{N_i} \times \frac{N_i}{N_i + \Delta N_i}\right) \times \frac{N_i + \Delta N_i}{\Delta N_i}$$

$$p_i''^j(b_{ik}^j) = \left(p_i'^j(b_{ik}^j) - (p_i^j(b_{ik}^j) \times \frac{N_i}{N_i + \Delta N_i})\right) \times \frac{N_i + \Delta N_i}{\Delta N_i}, k \in \{1,..h\} \quad (12)$$

This incremental adaptation permits to follow the evolution of classes online with a constant and low classification time. Then, the histogram of possibility is calculated using (3).
Fig. 2 illustrates the functioning phases of DFPM.

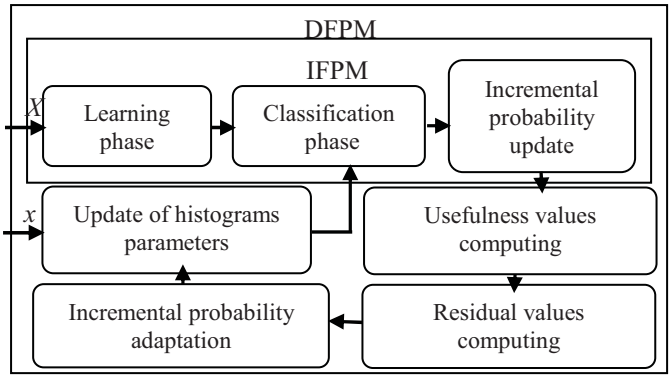


Figure 2: Functioning phases of DFPM.

4 Results with dynamic classes

Classes become dynamic if at least one of their characteristics changes with the time. For the sake of simplicity we have used a set of simulations described in a two dimension space. The obtained results can be extended for higher dimensions. The generated datasets follow a normal distribution of two independent variables (mean and standard deviation). We suppose that one unit of time corresponds to the classification of one point. This hypothesis is verified since DFPM has a constant and low classification time.

4.1 Class drift

Figure 3 shows a drift of a class in the course of time. The drift database has been created as follows:
- t=0: the possibility densities are already learned using 100 labeled patterns. The mean values of these data points are μ^1=3 and μ^2=3 (Fig. 3.a).
- t=1 to 100: 100 new incoming patterns, possessing the same initial class' characteristics (mean and standard deviation) occur in the class; thus there is no evolution or drift.
- t=101 to 200: a sudden change takes place in the class' mean values μ^j according to each attribute *j*, $j \in \{1,2\}$. This sudden change is followed by a slow drift of the mean value according to each attribute (Fig. 3.b):

$$\begin{cases} \mu^{1'}(t) = \mu^1 + 2 + \frac{4 \times (t-100)}{100} \\ \mu^{2'}(t) = \mu^2 + 2 + \frac{2 \times (t-100)}{100} \end{cases} \quad 101 \le t \le 200 \quad (13)$$

- t=201 to 300: 100 new patterns with the same characteristics as the ones of the drifted class occur. Fig 3.c presents the 0.1 membership level curve expected for the final drifted class. This curve contains all the points which have a membership possibility value to the class greater or equal to 0.1.

ISBN: 978-989-95079-6-8

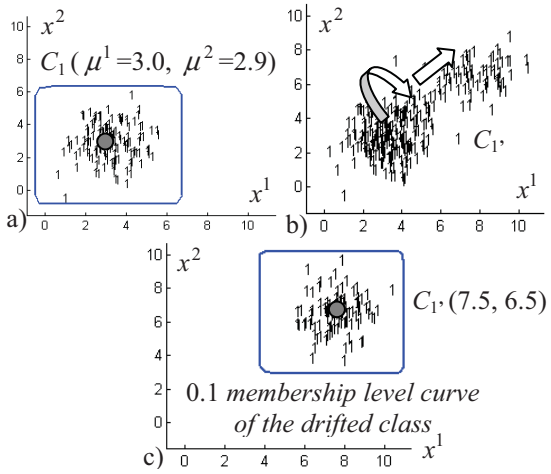


Figure 3: Evolution of the drift for the class C_1. The mean values of the class are presented before and after evolution between parentheses.

Fig. 4 presents the 0.1 membership level curve obtained by IFPM at the end of the class' drift ($t = 200$). We suppose here that each attribute borders were correctly determined in order to include the class drift.

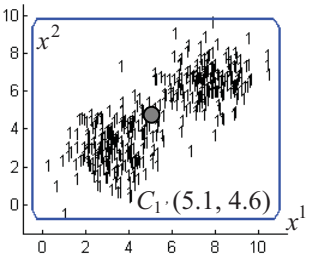


Figure 4: Classification of all patterns of the drifted class by IFPM.

All the patterns classified by IFPM are considered representative. Thus, the membership function, represented by the 0.1 membership level curve in Fig. 4, is not adapted to the drifted class with its new means values (see Fig. 3.c). This is due to the fact that all patterns were considered without eliminating any old and useless pattern.

The DFPM method is evaluated on the same drift of a class generated by (14). The residual values obtained progressively after the classification of each pattern in the class are represented in Fig. 5.

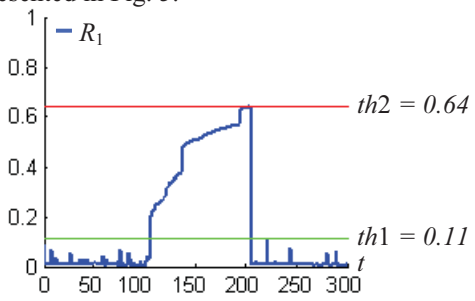


Figure 5: Residual values obtained by DFPM for the class drift.

Based on Fig. 5, we can conclude:

- between t=1 and t=100, the class' evolution has not yet begun.
- at t=105, the evolution of the class is detected. To detect this evolution the *th1* must be equal to 0.11.
- at t=204, the class drift is finished. To detect the end of this class drift the second threshold *th*2 must be equal to 0.64.
- finally, between t=205 and t=300 no evolution is observed so the residual values are small for the latest patterns and the class is stable. These patterns are classified in the class as usual.

Fig. 6 permits to see the membership level curve obtained by DFPM just before the evolution (Fig. 6.a) and the final result of classification after evolution when the adaptation of the histograms was realized (Fig. 6.b).

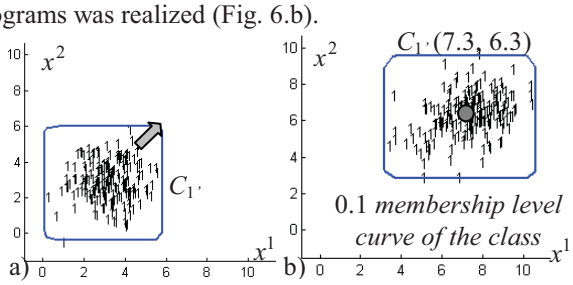


Figure 6: 0.1 membership level curve obtained by DFPM before evolution and after classification of all patterns.

The drift of the class has been well detected for both attributes. The classifier has been adapted with the representative patterns. However, the thresholds *th*1 and *th*2 must be well determined. A too big or too small *th*1 value leads to lose some representative patterns or to keep some obsolete ones, while a too big or a too small *th*2 value leads to detect too late or too early the class evolution.

The mean values of the class obtained by DFPM are close to those expected, contrary to those obtained by IFPM (Fig. 4). However, some patterns which represent the connection between the old class and the evolved one are still in the final class obtained by DFPM.

4.2 Rotation of class

The rotation of a class is presented in Fig. 7. The standard deviations of the class according to the two attributes are initiated at σ_1 =4 and σ_2 =0.25. The possibility densities are already learned using 150 learning patterns of the class (Fig. 7.a). Then, new incoming points evolve, between t=1 and t=150, to achieve a rotation of the class (Fig. 7.b).

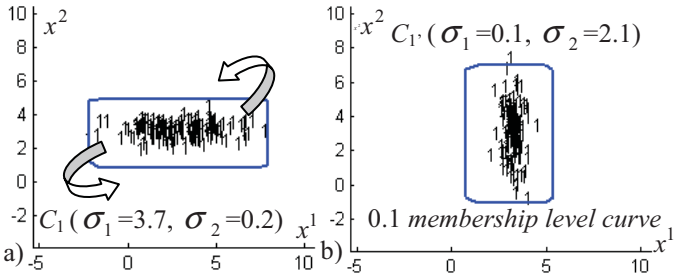


Figure 7: Rotation of the class C_1.

These changes are generated for the class according to each attribute j, $j \in \{1,2\}$, by:

$$\left.\begin{aligned} \mu_1^{1'} &= \mu_1^1 \times \sqrt{\frac{\sigma_2 \times t}{150}} \\ \mu_1^{2'} &= \mu_1^2 \times \sqrt{\frac{\sigma_1 \times t}{150}} \end{aligned}\right\} \quad 1 \le t \le 150 \qquad (14)$$

Fig. 8 presents the 0.1 membership level curve obtained by IFPM after the rotation of the class (t=150).

ISBN: 978-989-95079-6-8

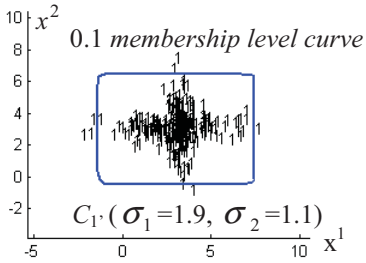


Figure 8: Final classification result obtained by IFPM for the patterns of the rotation's case.

All patterns have been classified in the class, but IFPM has not detected the rotation of the class. The membership level function is not adapted and useless patterns are still in the class. With DFPM, the residual values obtained after the classification of each pattern in the class are represented in Fig. 9.

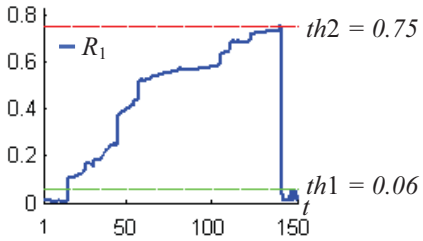


Figure 9: Residual values obtained by DFPM for the rotation of the class.

Based on Fig. 9, we can conclude that the use of $th1$=0.06 and $th2$=0.75 leads to well detect the class rotation.
Fig. 10 presents the membership level curve obtained by DFPM after classification of the evolving patterns when the adaptation of the histograms was realized.

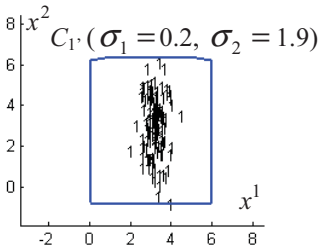


Figure 10: Classification result obtained by DFPM after adaptation of the class.

The classifier has been adapted with all representative patterns detected since the start of the class rotation. The new conserved useful patterns are the only patterns used to calculate the membership function. The resulting class corresponds well to the expected one, contrary to the class obtained by IFPM (Fig. 8).

5 Conclusions

In this paper, the Incremental version of the classification method Fuzzy Pattern Matching (IFPM) has been developed to discriminate dynamic classes. This development, called Dynamic FPM (DFPM), follows online the accumulated gradual changes in classes' probability distributions after the classification of each new pattern. Bins' usefulness values are calculated for each class and attribute. Based on these usefulness values, a residual value is calculated for each class in order to follow the evolution of each class. Then, the classes histograms are adapted in an incremental manner using the recent and useful patterns. This adaptation is achieved when the residual value reaches a suitable threshold according to the system dynamic. The DFPM method will be applied on two real applications. The first one concerns the following of a treatment's evolution for hemiplegic's patient according to its response to a medical treatment. The second application is a folding metal system which evolves from a normal mode to a faulty one according to temporal changes in the characteristics of the system and to the wear of its tools.

References

[1] H. Amadou Boubacar, *et al.,* AUDyC Neural Network using a new Gaussian Densities Merge Mechanism. *ICANNGA05*, pp. 155-158. Coimbra, Portugal, 2005.

[2] P.P. Angelov, A fuzzy controller with evolving structure. *Information Sciences* 161 (1-2), pp.21-35, 2004.

[3] L. Angstenberger, Dynamic Fuzzy Pattern Recognition. Dissertation, *Fakultät für Wirtschaftswissenschaften der Rheinisch-Westfälischen Technischen Hochschule*, Aachen, Germany, 2000.

[4] J.C. Bezdek, Pattern Recognition with fuzzy objective function algorithms, *Plenum Press,* 1981.

[5] M. S. Bouguelid, *et al.*, New results on diagnosis by pattern recognition, 4th International Conference on Informatics in Control, Automation and Robotic, pp. 167-172, Angers, France, 2007.

[6] M. Cayrol, *et al.,* Fuzzy Pattern Matching, Kybernetes, 11, pp. 103-116, 1982.

[7] L. Cohen, *et al.,* Incremental Info-Fuzzy Algorithm for Real Time Data Mining of Non-Stationary Data Streams, *TDM Workshop*, Brighton UK, 2004.

[8] D. Dubois, H. Prade, On possibility/probability transformations. *Fuzzy Logic*, pp. 103-112, 1993.

[9] R.O. Duda, *et al*., Pattern Classification 2nd edition. *Wiley-Interscience*, 2001.

[10] H. Frigui, R. Krishnapuram, Clustering by competitive agglomeration, *Pattern Recognition 30* (7), pp. 1109-1119, 1997.

[11] W.J. Gibb, *et al.*, Adaptive classification of myocardial electrogram waveforms, *IEEE Transactions on Biomedical Engineering*, N° 41, pp. 804-808, 1994.

[12] A.K. Jain, *et al*., Statistical Pattern Recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22 (1), pp. 4-37, 2000.

[13] M. Last, Online classification on nonstationary data streams, *Intelligent Data Analysis*, vol. 6 (2), pp. 129-147, 2002.

[14] G. Nakhaeizadeh, *et al.,* Dynamic Supervised Learning. Some Basic Issues and Application Aspects. *Classification and Knowledge Organization*. Springer Verlag, Berlin, pp. 123-135, 1997.

[15] M. Sayed Mouchaweh, *et al*., Incremental Learning in fuzzy pattern matching. *Fuzzy Sets and Systems* 132 (1), pp. 49-62, 2002.

[16] C.W. Therrien, Decision Estimation and Classification: An Introduction to Pattern Recognition and Related Topics. *John Wiley & Sons*, New York, 1989.

[17] G. Widmer, M. Kubat, Learning in the Presence of Context Drift and Hidden Contexts, *Machine Learning 23*, pp. 69-101, 1996.

ISBN: 978-989-95079-6-8

Possibility theory and formal concept analysis in information systems

Didier Dubois Henri Prade

IRIT, University of Toulouse III
118 Route de Narbonne, 31062 Toulouse Cedex 9, France
Email: dubois@irit.fr, prade@irit.fr

Abstract— *The setting of formal concept analysis presupposes the existence of a relation between objects and properties. Knowing that an unspecified object has a given property induces a formal possibility distribution that models the set of objects known to possess this property. This view expressed in a recent work by the authors of the present paper, has led to introduce the set-valued counterpart to the four set functions evaluating potential or actual, possibility or necessity that underlie bipolar possibility theory, and to study associated notions. This framework puts formal concept analysis in a new, enlarged perspective, further explored in this article. The "actual (or guaranteed) possibility" function induces the usual Galois connexion that defines the notion of a concept as the pair of its extent and its intent. A new Galois connexion, based on the necessity measure, partitions the relation in "orthogonal" subsets of objects having distinct properties. Besides, the formal similarity between the notion of division in relational algebra and the "actual possibility" function leads to define the fuzzy set of objects having* most *properties in a set, and other related notions induced by fuzzy extensions of division. Generally speaking, the possibilistic view of formal concept analysis still applies when properties are a matter of degree, as discussed in the paper. Lastly, cases where the object / property relation is incomplete due to missing information, or more generally pervaded with possibilistic uncertainty is also discussed.*

Keywords— possibility theory; formal concept analysis.

1 Introduction

Formal concept analysis [1] exploits the duality between objects and properties in a lattice theory setting, which has led to an original and practical view of the notion of a formal concept with application in data mining. A concept is then a pair made of a set of objects and a set of properties that are in mutual correspondence. These two sets are the extent and of the intent of the concept respectively. In this framework, properties are *binary*, and *complete* information is assumed about the relation linking objects and properties.

Fuzzy set theory [2] has emphasized the idea that properties are not always all-or-nothing notions, but are rather often a matter of degree. This has led to an extension of the original formal concept analysis setting by allowing intermediate truth values for the propositions "object x has property y" [3]. However, complete information is still assumed. Namely, for any pair (object, property), it it is known to what degree the object has the property. Besides, fuzzy sets have also been the starting point for the development of a new approach for the representation of uncertainty, named possibility theory [4, 5]. Fuzzy sets then have a disjunctive reading and represent states of incomplete information, i.e. (soft) restrictions on the mutually exclusive possible values of a single-valued variable.

The authors of the present paper have recently advocated the interest of a possibilistic reading of formal concept analysis where the possibility theory set-functions are shown to be meaningful in formal concept analysis [6]. Under this view, the set of objects known to possess a given property plays the role of a formal possibility distribution that restricts the possible value (identity) of an unspecified object only described as having the given property. This leads to an enlarged setting that we continue to investigate in this paper. Besides, this enlarged setting can be itself extended either by allowing properties to be non-Boolean, or by considering that the relation between objects and properties may be incompletely known.

The next section provides the background on a possibility theory-inspired view of formal concept analysis. Four basic operators that are the counterparts of the four basic set-functions in bipolar possibility theory are introduced, leading to consider another Galois connexion distinct from the usual one that gives birth to the notion of concept. Section 3 briefly describes the extension of the enlarged setting to fuzzy properties, i.e., when the relation linking objects and properties becomes fuzzy. Section 4 relates the operator underlying classical formal concept analysis and the notions of division (and quotient) in relational algebra. This provides a basis for defining the fuzzy set of objects having $most$ properties in a set (where $most$ is a fuzzy quantifier) and other related notions. Section 5 deals with the situation where information is incomplete or uncertain. Due to the lack of space, the paper only outlines new ideas, without developing them.

2 A possibility theory view of concept analysis

A formal information system is viewed here as a binary relation R between a set Obj of objects and a set $Prop$ of Boolean properties. Some authors speak of 'attribute' instead of 'property'. As we shall see in Section 5, this distinction only matters for attributes with non-binary domains. R is called *context* in formal concept analysis. If $X \subseteq Obj$, $\overline{X}$ is its complement $Obj \setminus X$. The notation $(x, y) \in R$ means that object x has property y. $R(x) = \{y \in Prop | (x, y) \in R\}$ is the set of properties of object x. Similarly, $R^{-1}(y) = \{x \in Obj | (x, y) \in R\}$ is the set of objects having property y. Its characteristic function induces a two-valued possibility distribution π:

$$\forall x \in Obj, \quad \pi(x) = \begin{cases} 1 & \text{if } x \in R^{-1}(y) \\ 0 & \text{otherwise,} \end{cases}$$

Intuitively speaking, if all we know about an unknown object is that it has property y then this object may be any x such that $\pi(x) = 1$ in context R. Thus, the relation R for a particular property y plays the role of a possibility distribution π encoding a set of possible values for x.

ISBN: 978-989-95079-6-8

2.1 Possibility theory

In possibility theory [5], two "measures" are associated with a possibility distribution π defined on a universe U as the characteristic (membership) function of a fuzzy set E representing the available information (in the above case $U = Obj$, and $E = R^{-1}(y)$ is an ordinary subset of U). Namely

i) a possibility measure Π (or "potential possibility"):

$$\Pi(A) = \max_{x \in A} \pi(x).$$

It estimates to what extent event A is *consistent with* the information represented by π. $\Pi(A \cup B) = \max(\Pi(A), \Pi(B))$ is the characteristic property of possibility measures [4].

ii) a dual measure of necessity N, expressing that an event is all the more necessarily (certainly) true as the opposite event is more impossible. N thus reflects an "actual necessity":

$$N(A) = 1 - \Pi(\overline{A}) = 1 - \max_{x \notin A} \pi(x),$$

where $\overline{A} = U \setminus A$. $N(A)$ estimates to what extent event A is *implied* by the information E represented by π (inasmuch as this information entails that any realization of $\overline{A}$ is more or less impossible). Necessity measures are characterized by the decomposability property $N(A \cap B) = \min(N(A), N(B))$.

Π and N are based on the $maximum$ of π over A and $\overline{A}$ respectively; two other set-functions [5] use the $minimum$:

iii) a measure of "actual (or guaranteed) possibility"

$$\Delta(A) = \min_{x \in A} \pi(x),$$

which estimates to what extent all elements in A are possible. Δ can be also termed "$sufficiency\ measure$" since $\Delta(A) = 1$ is enough for ensuring that all realizations of A are actually possible. Clearly, $\Delta \leq \Pi$. Note also that $\Delta(A)$ and $N(A)$ are unrelated. $\Delta(A \cup B) = \min(\Delta(A), \Delta(B))$ is the characteristic property of guaranteed possibility measures.

iv) a dual measure of "potential necessity or certainty"

$$\nabla(A) = 1 - \Delta(\overline{A}) = 1 - \min_{x \notin A}(\pi(x))$$

which estimates to what extent there exists at least one value in the complement of A that has a zero (or more generally a low) degree of possibility. This is clearly a $necessary$ condition for having "$x \in A$" somewhat certain. Property $\nabla(A \cap B) = \max(\nabla(A), \nabla(B))$ characterizes these measures.

2.2 An enlarged formal concept analysis setting

These four set functions make sense in the formal concept analysis. Namely, four remarkable sets can be defined:

$$R^{\Pi}(X) = \{y \in Prop | R^{-1}(y) \cap X \neq \emptyset\}$$
$$R^{N}(X) = \{y \in Prop | R^{-1}(y) \subseteq X\}$$
$$R^{\Delta}(X) = \{y \in Prop | R^{-1}(y) \supseteq X\}$$
$$R^{\nabla}(X) = \{y \in Prop | R^{-1}(y) \cup X \neq Obj\}$$

whose respective characteristic functions are $\Pi(X)$, $N(X)$, $\Delta(X)$, and $\nabla(X)$. Their meanings are as follows w. r. t. a subset of objects X in context R.

- $R^{\Pi}(X)$ is the set of properties that are associated with at least one object in X. Formally, we have

$$R^{\Pi}(X) = \cup_{x \in X} R(x).$$

$R^{\Pi}(X)$ is such that any object that satisfies one of them is possibly in X. In other words, if an object has none of the properties in $R^{\Pi}(X)$ then it cannot belong to X. Moreover, we have $R^{\Pi}(X_1 \cup X_2) = R^{\Pi}(X_1) \cup R^{\Pi}(X_2)$.

- $R^{N}(X)$ is the set of properties s. t. any object that satisfies one of them is necessarily in X. Having any property in $R^{N}(X)$ is a sufficient condition for belonging to X, and $R^{N}(X) = \overline{R^{\Pi}(\overline{X})} = Prop \setminus R^{\Pi}(\overline{X})$. Thus,

$$R^{N}(X) = \cap_{x \notin X} \overline{R(x)}.$$

and $R^{N}(X_1 \cap X_2) = R^{N}(X_1) \cap R^{N}(X_2)$.

- $R^{\Delta}(X)$, set of properties shared by all objects in X is

$$R^{\Delta}(X) = \cap_{x \in X} R(x).$$

In other words, satisfying all properties in $R^{\Delta}(X)$ is a necessary condition for an object to belong to X. $R^{\Delta}(X)$ is a partial conceptual characterization of objects in X: objects in X should have all the properties of $R^{\Delta}(X)$ and may have some others (that are not shared by all objects in X). It is worth noticing that $\overline{R^{\Pi}(X)}$ provides a negative conceptual characterization of objects in X since it gathers all the properties that are never satisfied by any object in X. Besides, it can be checked that $R^{N}(X) \cap R^{\Delta}(X)$ is the set of properties possessed by all objects in X and only by them. Moreover, we have $R^{\Delta}(X_1 \cup X_2) = R^{\Delta}(X_1) \cap R^{\Delta}(X_2)$.

- Note that $R^{\nabla}(X) = \overline{R^{\Delta}(\overline{X})} = Prop \setminus R^{\Delta}(\overline{X})$. Thus $R^{\nabla}(X)$ is the set of properties in $Prop$ that are not satisfied by at least one object in $\overline{X}$, i.e. $R^{\nabla}(X)$ is the set of properties that some object in $\overline{X}$ misses. In other words, in context R, for any property in $R^{\nabla}(X)$, there exists at least one object outside X that misses it. We have

$$R^{\nabla}(X) = \cup_{x \notin X} \overline{R(x)}.$$

and the following decomposability property holds $R^{\nabla}(X_1 \cap X_2) = R^{\nabla}(X_1) \cup R^{\nabla}(X_2)$.

Note that $R^{\Pi}(X)$ and $R^{N}(X)$ get larger when X increases, while $R^{\Delta}(X)$ and $R^{\nabla}(X)$ get smaller. The four modal-like operators R^{Π}, R^{N}, R^{Δ}, and R^{∇} have been considered by Düntsch and Orlowska [7] in the Boolean algebra setting, where R^{Δ} is called *sufficiency* operator, and its representation capabilities are studied. Taking inspiration as the previous authors from rough sets [8], Yao [9] also lays bare these four subsets. In both cases, the four operators were introduced without any mention of possibility theory.

Results in possibility theory have their counterparts in the enlarged formal concept analysis setting, as, e.g., [6]: If R^{-1} is s. t. $\forall y \in Prop, R^{-1}(y) \neq \emptyset$ and $R^{-1}(y) \neq Obj$, then

$$\forall X \subseteq Obj, \quad R^{N}(X) \cup R^{\Delta}(X) \subseteq R^{\Pi}(X) \cap R^{\nabla}(X). \quad (1)$$

Assuming that the property y is non trivial with respect to the set of objects Obj, i.e. $R^{-1}(y) \neq \emptyset$ (at least one object has property y) and $R^{-1}(y) \neq Obj$ (at least one object has not property y), then the four sets $R^{\Pi}(X)$, $R^{\Delta}(X)$, $R^{N}(X)$,

ISBN: 978-989-95079-6-8

Situation	$y \in R^{\Pi}(X)$	$y \in R^{\Delta}(X)$	$y \in R^{N}(X)$	$y \in R^{\nabla}(X)$
1. $X = \overline{R^{-1}(y)}$	0	0	0	0
2. $X \subset \overline{R^{-1}(y)}$	0	0	0	1
3. $\overline{R^{-1}(y)} \subset X$	1	0	0	0
4. $R^{-1}(y) \cap X \neq \emptyset$, $R^{-1}(y) \cap \overline{X} \neq \emptyset$	1	0	0	1
5. $R^{-1}(y) \subset X$	1	0	1	1
6. $X \subset R^{-1}(y)$	1	1	0	1
7. $R^{-1}(y) = X$	1	1	1	1

Figure 1: The seven possible positions of X and $R^{-1}(y)$

	objects							
	1	2	3	4	5	6	7	8
a					×	×	×	×
b					×	×		
c						×	×	×
d					×	×	×	×
e							×	
f					×	×		×
g	×	×	×	×				
h		×	×	×				
i				×				

Figure 2: R: a relation objects/properties a, b, c, ..., i

$R^{\nabla}(X)$ are necessary and sufficient for describing (and distinguishing between) the seven relative possible positions of X and $R^{-1}(y)$, as shown in Table 1, where 1 (resp. 0) stands for $y \in A$ (resp $y \notin A$) where A is the set $R^{*}(X)$ associated to the column (and $*$ is Π, Δ, N, or ∇). Note that the 9 = 16 - 7 remaining binary 4-tuples are ruled out by the constraints induced by (1), namely $R^{N}(X) \subseteq R^{\Pi}(X)$, $R^{\Delta}(X) \subseteq R^{\Pi}(X)$, $R^{\Delta}(X) \subseteq R^{\nabla}(X)$, $R^{N}(X) \subseteq R^{\nabla}(X)$. For instance, the "trivial" cases $R^{-1}(y) = \emptyset$ and $R^{-1}(y) = Obj$ (ruled out by the constraints) would be captured by distinct 4-tuples in Table 1, namely $(0 \quad 0 \quad 1 \quad 1)$ and $(1 \quad 1 \quad 0 \quad 0)$ respectively.

The above characterization of remarkable sets of properties w. r. t. a set of objects can be easily adapted for defining the corresponding sets of objects associated to a set of properties $Y \in Prop$: namely $R^{-1\Pi}(Y)$, $R^{-1N}(Y)$, $R^{-1\Delta}(Y)$, and $R^{-1\nabla}(Y)$. Their definitions can be easily obtained by swapping R and R^{-1} and exchanging the roles of the sets Obj and $Prop$. Namely,

$$R^{-1\Pi}(Y) = \{x \in Obj | R(x) \cap Y \neq \emptyset\} = \cup_{y \in Y} R^{-1}(y)$$
$$R^{-1N}(Y) = \{x \in Obj | R(x) \subseteq Y\} = \cap_{y \notin Y} \overline{R^{-1}(y)}$$
$$R^{-1\Delta}(Y) = \{x \in Obj | R(x) \supseteq Y\} = \cap_{y \in Y} R^{-1}(y)$$
$$R^{-1\nabla}(Y) = \{x \in Obj | R(x) \cup Y \neq Obj\} = \cup_{y \notin Y} \overline{R^{-1}(y)}.$$

Remark The above operators can be combined together. For instance, consider an object x_0. Let $R(x_0)$ be the set of its (known) properties. Compute $R^{-1\Delta}(R(x_0))$, the set of objects that share these properties. Then get $R^{\Pi}(R^{-1\Delta}(R(x_0)))$, which is the set of properties that are associated with at least one object sharing the properties of x_0. Viewing the table $Obj \times Prop$ as the information pertaining to a repertory of cases, and x_0 as a partially known extra object (not in Obj) for which one tries to guess other properties, the expression $R^{\Pi}(R^{-1\Delta}(R(x_0)))$ may be viewed as the result of a *case-based reasoning* procedure, i.e. a set of potential properties that x_0 may also have. Besides, its subset $R^{N}(R^{-1\Delta}(R(x_0)))$ is the set of properties that alone characterize the objects sharing the properties of x_0. Thus, if one of the properties in $R^{N}(R^{-1\Delta}(R(x_0)))$ is not already among the known properties of x_0, it may be considered as a serious candidate property for x_0.

2.3 Galois connexions

In formal concept analysis, the pair of set valued functions R^{Δ} and $R^{-1\Delta}$ induces a Galois connexion [10] between 2^{Obj} and 2^{Prop}. Then, a *formal concept* is a pair (X, Y) such that $X = \{x \in Obj | R(x) \supseteq Y\}$ and $Y = \{y \in Prop | R^{-1}(y) \supseteq X\}$, i.e. such that $X = R^{-1\Delta}(Y)$ and $Y = R^{\Delta}(X)$, X is called its *extent* and Y its *intent*. In other words, in a formal concept (X, Y), Y is the set of properties shared by all the objects in X, and X is the set of objects that possess all the properties in Y. Then $X \times Y \subseteq R$, i.e. $\forall x \in X, \forall y \in Y, (x, y) \in R$. A formal concept is a maximal pair that satisfies the latter condition (where maximality is taken in the sense of set inclusion).

Putting formal concept analysis in the perspective of possibility theory, it becomes then natural to also consider
- the pairs (X, Y) s. t. $X = R^{-1\Pi}(Y)$ and $Y = R^{\Pi}(X)$;
- the pairs (X, Y) s. t. t $X = R^{-1N}(Y)$ and $Y = R^{N}(X)$;
- the pairs (X, Y) s. t. $X = R^{-1\nabla}(Y)$ and $Y = R^{\nabla}(X)$.

First, observe that $X = R^{-1\nabla}(Y)$ and $Y = R^{\nabla}(X)$ holds if and only if $\overline{X} = R^{-1\Delta}(\overline{Y})$ and $\overline{Y} = R^{\Delta}(\overline{X})$ holds, i.e., if $(\overline{X}, \overline{Y})$ is a formal concept, due to the duality between operators R^{Δ} and R^{∇}. Similarly, $X = R^{-1\Pi}(Y)$ and $Y = R^{\Pi}(X)$ holds if and only if $\overline{X} = R^{-1N}(\overline{Y})$ and $\overline{Y} = R^{N}(\overline{X})$ holds. But, it can be easily seen that a pair (X, Y) such that $X = R^{-1N}(Y)$ and $Y = R^{N}(X)$, i.e. such that $X = \{x \in Obj | R(x) \subseteq Y\}$ and $Y = \{y \in Prop | R^{-1}(y) \subseteq X\}$ is not generallly a formal concept, as now exemplified. This Galois connexion has been introduced by [11] on a formal basis, but its practical meaning was apparently not really discussed.

Example 1 *We consider an example of relation R described by the table of Figure 2. This relation defines the links between eight objects $Obj = \{1, 2, 3, 4, 5, 6, 7, 8\}$ and nine properties $Prop = \{a, b, c, d, e, f, g, h, i\}$. There is a "×" in the cell corresponding to an object x and to a property y if the object x has property y, in other words the "×"s describe the relation R (or context). An empty cell corresponds to the fact that $(x, y) \notin R$, i.e., it is known that object x has not property y. It can be checked that the pairs $(\{1, 2, 3, 4\}, \{g, h, i\})$, $(\{5, 6, 7, 8\}, \{a, b, c, d, e, f\})$ are pairs (X, Y) such that $X = R^{-1N}(Y)$ and $Y = R^{N}(X)$. These two pairs are not formal concepts. They are disjoint w. r. t. both Obj and $Prop$. Examples of formal concepts are $(\{2, 3, 4\}, \{g, h\})$, $(\{6, 7, 8\}, \{a, c, d\})$, or $(\{5, 6, 7, 8\}, \{a, d\})$. Note that these latter examples are here obtained by considering appropriate subsets in the previous pairs.*

A pair (X, Y) that satisfies $X = R^{-1N}(Y)$ and $Y = R^{N}(X)$

ISBN: 978-989-95079-6-8

properties \ objects	1	2	3	4	5	6	7	8
a					×	×	×	×
b					×	×		
c						×	×	×
d				×	×	×	×	×
e							×	
f					×	×		×
g	×	×	×	×				
h		×	×	×				
i				×				

Figure 3: R': relation R modified

is such that all the objects in X possess *at least one* property in Y and the properties in Y are only (possibly) possessed by the objects in X. While the intent of a formal concept is a *conjunction* of properties, the pairs (X, Y) forming an "N-block" correspond to sets of objects defined through *disjunctions* of properties. Finding such pairs, which may not exist, aims at decomposing the relation R into independent blocks without object or property in common, as in the Figure 2 example. When such a decomposition no longer holds, as in Figure 3, pairs (X, Y) such that $X = R^{-1N}(Y)$ and $Y = R^N(X)$ no longer exist, except for the trivial pair (Obj, $Prop$), as shown in the next example.

Example 2 *Let us now consider a modified version of relation R, say R′, depicted in the table of Figure 3, where object 4 has also the additional property d now. Then it can be checked that we still have* $R'^N(\{1,2,3,4\}) = \{g,h,i\}$*, but* $R'^{-1N}(\{g,h,i\}) = \{1,2,3\}$*, since* $R'(4) = \{d,g,h,i\} \not\subseteq \{g,h,i\}$*. Similary,* $R'^{-1N}(\{a,b,c,d,e,f\}) = \{5,6,7,8\}$*, but* $R'^N(\{5,6,7,8\}) = \{a,b,c,e,f\}$ *since* $R'^{-1}(d) = \{4,5,6,7,8\} \not\subseteq \{5,6,7,8\}$*. Thus the pairs* $(\{1,2,3,4\},\{g,h,i\})$*,* $(\{5,6,7,8\},\{a,b,c,d,e,f\})$ *are no longer pairs* (X, Y) *such that* $X = R'^{-1N}(Y)$ *and* $Y = R'^N(X)$ *in the new context R'.*

3 Handling fuzzy properties in the new setting

In the previous section, properties were supposed to be Boolean. Hence, when an object satisfies a property, it fully satisfies it: there is no intermediary degree of satisfaction since the property is not gradual. Thus, the relation linking objects and properties was all-or-nothing. When properties become a matter of intensity, i.e., when an object may have a property to some degree, the relation R between objects and properties becomes fuzzy. However, when relaxing the Booleanity assumption, we still assume that we have complete information. Namely, it is known to what extent α object x has property y for any pair (x, y), which is denoted $\mu_R(x, y) = \alpha$. Then, $\mu_{R^{-1}(y)}(x) = \alpha$ denotes the fact that object x satisfies property y at degree α where $\mu_{R^{-1}(y)}$ is the membership function of the fuzzy set of objects that constitutes the extension of $R^{-1}(y)$. Such an extension to standard concept analysis (based on the operator called here Δ) has been studied by Belohlavek [3]; see also [12].

Then, the four operators introduced in the previous section easily extend to the case where relation R is fuzzy. Namely,

$$\begin{array}{llll} \mu_{R^\Pi(X)}(y) & = \Pi_y(X) & = \max_{x\in X} & \mu_{R^{-1}(y)}(x) \\ \mu_{R^N(X)}(y) & = N_y(X) & = \min_{x\notin X} & 1-\mu_{R^{-1}(y)}(x) \\ \mu_{R^\Delta(X)}(y) & = \Delta_y(X) & = \min_{x\in X} & \mu_{R^{-1}(y)}(x) \\ \mu_{R^\nabla(X)}(y) & = \nabla_y(X) & = \max_{x\notin X} & 1-\mu_{R^{-1}(y)}(x) \end{array}$$

where Π_y, N_y, Δ_y and ∇_y are respectively potential possibility, actual necessity, actual possibility, and potential necessity measures, based on the gradual possibility distribution $\pi = \mu_{R^{-1}(y)}$. They thus enjoy the corresponding characteristic decomposability properties of these respective measures.

The following results are straightforward:

- $\mu_{R^N(X)}(y) = 1 - \mu_{R^\Pi(\overline{X})}(y)$;
- $\mu_{R^\nabla(X)}(y) = 1 - \mu_{R^\Delta(\overline{X})}(y)$
- if $\mu_{R^\Pi(X)}(y) = \alpha$ then $\exists x \in X, \mu_R(x, y) = \alpha$ and $\forall x \in X, \mu_R(x, y) \leq \alpha$
- if $\mu_{R^N(X)}(y) = \alpha$ then $\mu_R(x, y) > 1 - \alpha \Rightarrow x \in X$
- if $\mu_{R^\Delta(X)}(y) = \alpha$ then $x \in X \Rightarrow \mu_R(x, y) \geq \alpha$
- if $\mu_{R^\nabla(X)}(y) = \alpha$ then $\exists x \notin X, \mu_R(x, y) = 1 - \alpha$ and $\forall x \notin X, \mu_R(x, y) \geq 1 - \alpha$

The first two results extend duality relations to the graded case. The other ones express the meaning of each fuzzy set. Thus a property belongs to $R^\Pi(X)$ to degree α inasmuch as objects in X possess this property to at most degree α. Then $h(R^\Pi(X)) = \max_{y\in Y} \mu_{R^\Pi(X)}(y) = 0$ means that no object in X possesses a property in Y to any extent. A property belongs to $R^N(X)$ to degree α if any object possessing this property to a degree greater than $1 - \alpha$ necessarily belongs to X. In particular, any object possessing this property to some positive degree belongs to X, if $\alpha = 1$. A property belongs all the more to $R^\Delta(X)$ as any object in X possesses this property to a greater degree. Lastly, $\hbar(R^\nabla(X)) = \min_{y\in Y} \mu_{R^\nabla(X)}(y) = \alpha$ means that for any property y in Y objects outside X possess this property to at most to degree $1 - \alpha$. In particular, if $\hbar(R^\nabla(X)) = 1$, for any property in Y there exists an object outside X that misses it.

Moreover, we have the following counterpart to (1):

Proposition 1 *If* $R^{-1}(y)$ *is such that for* $y \in Prop$*,* $h(\mu_{R^{-1}(y)}) = 1$ *and* $\hbar(\mu_{R^{-1}(y)}) = 0$*, then*

$$\forall X \subseteq Obj, \quad \max(N_y(X), \Delta_y(X)) \leq \min(\Pi_y(X), \nabla_y(X)).$$

Lastly, it is clear that when nesting the operators as in the Remark at the end of Section 2.2, or when extending Galois connexions when R is fuzzy, we are led to extend again the above definitions to cases where X and Y become fuzzy sets themselves. From a possibility theory point of view, this means defining generalized measures for fuzzy events. Depending on the properties we want to preserve, several choices are possible here (see on this point [13], especially pages 51–64), which would lead to different extensions for the four operators. Choosing the appropriate extensions depends on the intended use and interpretation of the definition we want to generalize. We leave these questions for further research.

ISBN: 978-989-95079-6-8

4 Generalized quotient and fuzzy quantifiers

In a relational database, given an ordered set of attributes $\mathbf{A} = \{\mathcal{A}^1, ..., \mathcal{A}^n\}$, information is stored in a relational table $\mathcal{R}$ where each column corresponds to an attribute, and a row to an object belonging to *Obj*. Thus a cell in such a relational table corresponds to the value of an attribute for an object. Any row in the relational table $\mathcal{R}$ is also called a 'tuple'. A 'tuple' is thus an ordered set of attribute values pertaining to an object. Let $\mathcal{T}u(\mathcal{R})$ denote the set of tuples in $\mathcal{R}$. Quotients are relational algebra operations that aim at finding out the subrelational table $\mathcal{R} \div \mathcal{S}$ of a finite relational table $\mathcal{R}$, containing subtuples of $\mathcal{R}$ that have for complements in $\mathcal{R}$ *all* the tuples of a relational table $\mathcal{S}$. The quotient operation is defined by

Definition 1 *Relational quotient.*

$$\mathcal{R} \div \mathcal{S} = \{t, \forall s \in \mathcal{T}u(\mathcal{S}), (t, s) \in \mathcal{T}u(\mathcal{R})\}$$

where s denotes a tuple of $\mathcal{S}$ and t a subtuple of $\mathcal{R}$ such that (t, s) is a tuple of $\mathcal{R}$.

The definition of $R^{\Delta}(X)$ can be viewed as a particular case of such a division. Indeed $R^{\Delta}(X) = \{y \in Prop | R^{-1}(y) \supseteq X\} = \{y \in Prop | \forall x \in X, (x, y) \in R\}$. Besides, a relation $R \subseteq Prop \times Obj$ can be viewed as equivalent to a 2-attribute relational table $\mathcal{R}$ with $\mathbf{A} = \{Object - name, Property - name\}$, where (x, y) (resp. (y, x)) is a tuple in $\mathcal{R}$ (resp. $\mathcal{R}^{-1}$) if and only if $(x, y) \in R$. Then, it becomes clear that $R^{\Delta}(X) = \mathcal{R}^{-1} \div \mathcal{X}$, where $\mathcal{X}$ is the one-attribute relational table containing the object names. Similarly, $R^{-1\Delta}(Y) = \mathcal{R} \div \mathcal{Y}$ ($\mathcal{Y}$ is the one-attribute relational table associated to Y).

The fuzzy extensions of the basic operationsR^{Δ}, $R^{-1\Delta}$, which underly formal concept analysis, can thus be related to fuzzy division operations in fuzzy relational databases where tuples are weighted [14, 15]. As we shall see, this also provides a way for introducing fuzzily quantified conjunctions in order to require that tuples in R are associated with "at least k", or more generally '*most*' objects in X , rather than *all* elements in X as in formal concept analysis basic operations. In the following, we discuss these extensions in the setting of formal concept analysis.

First, the definition of $R^{\Delta}(X)$ extended to the case where R becomes fuzzy, as given in the previous section, namely $\mu_{R^{\Delta}(X)}(y) = min_{x \in X} \mu_{R^{-1}(y)}(x)$, is the exact counterpart to the fuzzy division $\mu_{\mathcal{R}^{-1} \div \mathcal{X}}(y) = \min_{x \in \mathcal{T}u(\mathcal{X})} \mu_R(x, y)$, for all pair (y, x) in $\mathcal{T}u(\mathcal{R}^{-1})$, when R is a fuzzy relation and X remains a classical set. The definition of a fuzzy division in fuzzy relation databases includes the more general case where X is also a fuzzy set. This can be done as well here, choosing an appropriate type of inclusion between fuzzy sets, i.e. an implication connective $\rightarrow$ in the expression:

$$\mu_{R^{\Delta}(X)}(y) = \min_{x \in Obj} \mu_X(x) \rightarrow \mu_{R^{-1}(y)}(x)$$

in relation with the intended meaning of having X fuzzy. For instance, taking Gödel implication ($a \rightarrow b = 1$ if $a \leq b$, and $a \rightarrow b = b$ if $a > b$) amounts to seeing $\mu_X(x)$'s as a significance threshold to which $\mu_R(x, y)$ is compared, while using Dienes implication ($a \rightarrow b = \max(1 - a, b)$) would be more in agreement with the idea of viewing $\mu_X(x)$ as a level of priority of x, just requiring the inclusion of important objects in $R^{-1}(y)$ (indeed the less important x, the greater $(1 - \mu_X(x))$, and the smaller the impact of x on the global evaluation, even when x totally fails to have property y (see [14, 15] for details).

Having both a fuzzy relation R and a fuzzy set X of objects may sound unrealistic in practice. In formal concept analysis, it is the starting point for natural weakening of the quantifier 'for all' into "at least k", or even into 'most'. The idea is to require that "at least k" (or more generally 'most') objects in X are the important objects that are the most in relation R with property y.

Let I be a fuzzy constraint on integers, defined by a membership function of the form: $\mu_I(0) = 1$and$\mu_I(i) \geq \mu_I(i+1)$. For instance, "at least k objects are important" is represented by $\mu_I(i) = 1$ if $0 \leq i \leq k$ and $\mu_I(i) = 0$ for $i \geq k + 1, n$ where $n = |X|$. Let us reorder the $\mu_{R^{-1}(y)}(x_k)$'s decreasingly, so that objects (x_k)'s that are more in relation R with y are the most important ones:

$$\mu_{R^{-1}(y)}(x_{\sigma(1)}) \geq \mu_{R^{-1}(y)}(x_{\sigma(2)}) \geq ... \geq \mu_{R^{-1}(y)}(x_{\sigma(n)}).$$

Then the extent to which property y is possessed by 'at least k' objects in X can be computed as

$$\mu_{R^{\Delta}(X),I}(y) = \min_i \ \max(\mu_{R^{-1}(y)}(x_{\sigma(i)}), 1 - \mu_I(i)),$$

I being defined as above. This expression, which involves an absolute fuzzy quantifier, may be easily modified in order to introduce relative quantifiers Q like 'most', having an increasing membership function in $[0, 1]$, by changing $1 - \mu_I(i + 1)$ into $\mu_Q(i/n)$ for $i = 0, n - 1$ and $\mu_Q(1) = 1$. It gives

$$\mu_{R^{\Delta}(X),Q}(y) = \min_i \ \max(\mu_{R^{-1}(y)}(x_{\sigma(i)}), \mu_Q(\frac{i-1}{n})).$$

Clearly, if Q means 'all', $\mu_Q(t) = 0$ for $t < 1$, then the above expression reduces to $\mu_{R^{\Delta}(X)}(y) = \min_{x \in X} \ \mu_{R^{-1}(y)}(x)$.

5 Incomplete and uncertain information

Until now, it has been assumed that we have *complete* information about the existing links between properties in $Prop$ and objects in Obj. Namely, $(x, y) \in R$ means that object x satisfies property y and $(x, y) \notin R$ means that object x does not satisfy property y, rather than "we do not know if $(x, y) \in R$ or not". Clearly, this assumption may be relaxed, while Boolean properties are still assumed: One may consider that there are pairs (x, y) for which it is not known at all if x has property y or not. This case has been considered in [16]. Information may be also uncertain, i.e., we are certain at level α that x has property y, or at level β that x has not property y.

In the most general case, properties are non Boolean (i.e. $\mu_R(x, y)$ is supposed to belong to $[0, 1]$), but the extent $\mu_R(x, y)$ to which an object x has a property y may be only fuzzily known under the form of a possibility distribution $\pi_{\mu_R}^{(x,y)}$ on $[0, 1]$ that restricts its possible values. Then one may not be even sure in general that some property y is possessed by an object x at least at some degree α). Since the information about $\mu_R(x, y)$ is now represented by a fuzzy set (on $[0, 1]$), and the four measures introduced in Section 3

ISBN: 978-989-95079-6-8

$\mu_{R^{\Pi}(X)}(y) = \Pi_y(X)$, $\mu_{R^{N}(X)}(y) = N_y(X)$, $\mu_{R^{\Delta}(X)}(y) = \Delta_y(X)$, $\mu_{R^{\nabla}(X)}(y) = \nabla_y(X)$ can themselves only be known under the form of induced possibility distributions (using the fuzzy set extension principle [2]). Let us take the example of $\mu_{R^{\Pi}(X)}(y) = \Pi_y(X) = \max_{x \in X} \mu_R(x, y)$. The induced possibility distribution is given by

$\pi_{\Pi_y(X)}(t) = \max_{i:\max_i t_i = t} \min_{x_i \in X} \pi_{\mu_R}^{(x_i,y)}(t_i)$ i.e.
$\pi_{\Pi_y(X)}(t) =$
$\max_i \min(\pi_{\mu_R}^{(x_i,y)}(t), \min_{j \neq i}(\max_{t_j \leq t} \pi_{\mu_R}^{(x_j,y)}(t_j))$.
Such a computation may be heavy in practice, but one may at least compute an upper bound of the possibility that a property y is associated to an object x with a degree equal to α, as $\pi^*_{\Pi_y(X)}(\alpha) = \max_{x \in X} \pi_{\mu_R}^{(x,y)}(\alpha)$. One may also compute the degree of membership of property y to the fuzzy set of properties that are possibly associated with at least one object in X at least to a degree ρ as $\Pi_{y,\rho}(X) = \max_{x \in X} \max_{\{t | t \geq \rho\}} \pi_{\mu_R}^{(x,y)}(t)$.

A maybe more promising approach for dealing with incomplete information in formal concept analysis is to first slightly modify the setting we start with by accommodating directly many-valued attributes instead of binary ones. We now outline this idea. Indeed when attribute domains are two-valued, they only give birth to a binary property (and its negation), while any non empty subset of a many-valued attribute domain (different from the domain itself) defines a non-trivial property. Take the example of the *color* attribute with domain $\{black, red, yellow, blue, green, ...\}$, $red\ or\ green$, or $red\ or\ yellow\ or\ blue$ are (imprecise) properties, beside the basic colors $black$, red, etc. Let Y now denote a set of attributes y, $dom(y)$ be the domain of y, P_y denote a non-empty subset of $dom(y)$. Let $\Gamma_y(x)$ represent the available information about the value of attribute y for object x. It is assumed that $\Gamma_y(x) \subseteq dom(y)$. Information is imprecise if $\Gamma_y(x)$ is not a singleton. For the moment, we suppose that information may be incomplete but not uncertain. $\Gamma_y(x) = \varnothing$ means that y does not apply to x, and $\Gamma_y(x) = dom(y)$ means that the value of y is unknown for x. $\Gamma_y(x) \neq \varnothing$ is now assumed.

Then $R^{-1N}(P_y) = \{x | \Gamma_y(x) \subseteq P_y\}$ is the set of objects that (certainly) have property P_y w. r. t. attribute y. Three other similar sets can be defined by reversing $\subseteq$, or replacing it by non-empty intersection, or non-covering union conditions, in the spirit of the basic definitions of Section 2. Let $R^{-1N\bigstar}$ be the relation that expresses that *certainly* objects have some (maybe imprecise) properties; it is defined on $X \times \cup_{y \in Y} \mathcal{P}(y)$, where $\mathcal{P}(y)$ denote the power set of $dom(y)$. Since $(\mathcal{P}(y), \subseteq)$ is a Boolean lattice, $(x, P_y) \in R^{-1N\bigstar}$ entails $(x, P'_y) \in R^{-1N\bigstar}$ as soon as $P_y \subseteq P'_y$ (if an object is red, it is also $red\ or\ green$). Clearly, $\{x | (x, red\ or\ green) \in R^{-1N\bigstar}\} = \{x | \Gamma_{color}(x) = red\} \cup \{x | \Gamma_{color}(x) = green\} \cup \{x | \Gamma_{color}(x) = red\ or\ green\}$. Moreover, we can also find out if there are only *possibly* $red\ or\ green$ objects, e.g. those that are known to be $red\ or\ blue$. If not, it means that there is no completion of the knowledge that can alter the extension of the set of $red\ or\ green$ objects in our example. More generally, we can look for concepts associated with sufficiently imprecise properties that remain stable under any knowledge completion. This can be extended to gradual uncertain knowledge by working with the α-cuts of the $\Gamma_y(x)$'s, i.e. pieces of information that are $(1-\alpha)$-certain. These are lines for further research.

6 Concluding remarks

Starting with a possibility-theoretic reading of concept analysis, we have reintroduced four operators, that enable us to describe all the different possible relations between a set of objects and a set of properties. Apart from retrieving the Galois connexion defining formal concepts, another Galois connexion based on the "actual necessity" operator is laid bare for decomposing the relation into independent blocks. The proposed setting extends to graded properties, leading to two kinds of Galois connected pair of fuzzy sets, whose meaning must be laid bare. Besides, the formal similarity between the actual possibility operator and relational algebra division operation, suggests a relaxation of the definition of concepts, computing the extent to which a property is *highly* possessed by *most* objects in a set. Lastly, extensions of the formal concept analysis setting to incomplete or uncertain information have been outlined. It is clear that many pending issues remain, such as e.g. the use of rough set reducts in this setting.

References

[1] B. Ganter and R. Wille. *Formal Concept Analysis, Mathematical Foundations*. Springer-Verlag, 1999.

[2] L.A. Zadeh. Fuzzy sets. *Inform. and Control*, 8, 338–353. 1965.

[3] R. Belohlavek. *Fuzzy Relational Systems: Foundations and Principles*, volume 20. Kluwer Academic/Plenum Press, 2002.

[4] L.A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.

[5] D.Dubois H.Prade Possibility theory: qualitative and quantitative aspects In *Quantified Representation of Uncertainty and Imprecision*, *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, vol.1, 169–226. Kluwer Acad. Pu. 1998.

[6] D. Dubois, F. Dupin de Saint-Cyr and H. Prade. A possibility-theoretic view of formal concept analysis. *Fudamentae Informaticae*, 75: 195–213, 2007.

[7] I.Düntsch, E.Orlowska Mixing modal and sufficiency operators *Bull. Section of Logic, Polish Acad. of Sci.*, 28(2):99–106, 1999.

[8] Z. Pawlak. *Rough Sets. Theoretical Aspects of. Reasoning about Data*. Kluwer Acad. Publ., Dordrecht, 1991.

[9] Y.Y. Yao. A comparative study of formal concept analysis and rough set theory in data analysis. In *Rough Sets and Current Trends in Computing, 4th Int. Conf., RSCTC 2004*, Uppsala

[10] O.Ore Galois connexions *Tr. Ame. Mat. Soc.* 55:493-513, 1944

[11] A. Popescu. A general approach to fuzzy concepts. *Math. Log. Quart.*, 50(3):265–280, 2004.

[12] A. Burusco and R. Fuentes-Gonzalez. Construction of the l-fuzzy concept lattice. *Fuzzy Sets and Syst.*, 97,109–114, 1998.

[13] B. Bouchon-Meunier, D. Dubois, L. Godo, H. Prade. Fuzzy sets and possibility theory in approximate and plausible reasoning. In *Fuzzy Sets in Approximate Reasoning and Information Systems*, (J. C. Bezdek et al. eds.), Kluwer, 15-190, 1999.

[14] D. Dubois M. Nakata H. Prade Extended divisions for flexible queries in relational databases *Knowledge Management in Fuzzy Databases* O. Pons et al. eds., Physica-Verlag 105-21, 2000

[15] P. Bosc, O. Pivert, D. Rocacher Characterizing the result of the division of fuzzy relations *Int. J. Appr. Reas.*, 45, 511-30, 2007

[16] P. Burmeister and R. Holzer. Treating incomplete knowledge in formal concept analysis. *Formal Concept Analysis*, (B. Ganter et al. eds.), Springer Verlag, LNAI 3626, pp. 114-126, 2005.

ISBN: 978-989-95079-6-8

On the use of Measures of Separability of Classes to characterise the Domains of Competence of a Fuzzy Rule Based Classification System

Julián Luengo, Francisco Herrera[1]

1. Dept. of Computer Science and Artificial Intelligence,
CITIC-UGR (Research Center on Information and Communications Technology)
University of Granada, Granada, 18071, Spain
Email: {julianlm,herrera}@decsai.ugr.es

***Abstract*—**

In this work we study the behaviour of a Fuzzy Rule Based Classification System, and its relationship to a certain data complexity measures family. As Fuzzy Rule Based Classification System we have selected a recent proposal called Positive Definite Fuzzy Classifier, which is a Fuzzy System that uses Support Vector Machines for its training, obtaining accurate results and a low number of rules.

We have examined several data complexity metrics of separability of classes *over a wide range of data sets built from real data, and try to extract behaviour patterns from the results for this learning method. Using these data complexity measures and the accuracy results of the Positive Definite Fuzzy Classifier, we have built a rule set which describes both good or bad behaviours of this Fuzzy Rule Based Classification System.*

These rules use different values of such data complexity measures as antecedents, so we aim to predict the behaviour of the method from the data set complexity metrics prior to its application. Therefore, the rule set could characterise the domains of competence of this particular Fuzzy Rule Based Classification System.

***Keywords*—** Classification, Data complexity, Fuzzy Rule Based Systems, Support Vector Machines

1 Introduction

Fuzzy Rule Based Classification Systems (FRBCSs) [12, 14] are a very useful tool in the ambit of Data Mining, since they are capable of building a linguistic model clearly interpretable by human beings. There is a vast literature in the field of FRBCSs [14], which is very active at this time [9, 1, 16, 13].

The prediction capabilities of classifiers are strongly dependent on the problem's characteristics. An emergent field, that uses a set of complexity measures applied to the problem to describe its difficulty, has recently arisen. These measures quantify particular aspects of the problem which are considered complicated to the classification task [11]. Studies of data complexity metrics applied to particular classification's algorithms can be found in [11, 4, 3, 19].

In this work we are interested in analysing the relationship between FRBCSs and the complexity measures, considering a case of study using the Positive Definite Fuzzy Classifier (PDFC) proposed by Chen and Wang [5]. In particular we consider one type of data complexity measures based on the separability of classes.

To perform this study, we have created several binary classification data sets from real world problems, 438 ones, and computed the value of 3 metrics proposed by Ho and Basu [10]. We have analysed the intervals of the complexity measures values related to the created data sets, in which PDFC method performs well or badly, and then formulated a rule for such intervals. The rules try to describe the ranges where some information and conclusions about the behaviour of PDFC method can be stated.

This contribution is organised as follows. In Section 2 we describe the FRBCS we have used. In Section 3 the considered complexity measures are described. In Section 4 we include the experimental set-up, the results obtained and the rules extracted, along with their analysis. Finally, in Section 5 some concluding remarks are made.

2 Preliminaries: Fuzzy Rule Based Classification System

Any classification problem consists of l training patterns $x_p = (x_{p1}, \ldots, x_{pn})$, $p = 1, 2, \ldots, l$ from M classes where x_{pi} is the ith attribute value ($i = 1, 2, \ldots, n$) of the p-th training pattern.

Let us consider a two-class classification problem of assigning class label $y \in \{+1, -1\}$ to input feature vector x_p. In this work we consider the additive FRBCSs with constant `THEN`-parts.

As learning method we use the PDFC method [5], which uses a Support Vector Machine (SVM) approach to build up the model. In the following two Subsections, first we include the fuzzy inference model and a complete description of the PDFC algorithm.

2.1 Fuzzy Inference Model: PDFC Method

Consider a fuzzy model with m fuzzy rules of the form:

$$\text{Rule } j: \text{ If } A_j^1 \text{ AND } A_j^2 \text{ AND } \ldots \text{ AND } A_j^n \text{ THEN } b_j \qquad (1)$$

where A_j^k is a fuzzy set with membership function $a_j^k : \mathbb{R} \to [0, 1]$, $j = 1, \ldots, m$, $k = 1, \ldots, n$, $b_j \in \mathbb{R}$. If we choose product as the fuzzy conjunction operator, addition for fuzzy rule aggregation, and center of area defuzzification, then the model becomes a special form of the Takagi-Sugeno fuzzy model, and the input output mapping, $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}$, of the model is defined as

$$\mathcal{F}(x_p) = \frac{\sum_{j=1}^{m} b_j \prod_{k=1}^{n} a_j^k(x_k)}{\sum_{j=1}^{m} \prod_{k=1}^{n} a_j^k(x_k)}. \qquad (2)$$

Equation (2) could not be well-defined on $\mathbb{R}$ if $\sum_{j=1}^{m} \prod_{k=1}^{n} a_j^k(x_k) = 0$ for some $x_p \in \mathbb{R}^n$, which could happen if the input space is not wholly covered by fuzzy rule "patches". To fix this problem, we add a fuzzy rule so that the

ISBN: 978-989-95079-6-8

denominator $\sum_{j=1}^{m}\prod_{k=1}^{n}a_j^k(x_k) > 0$ for all $x_p \in \mathbb{R}$. Thus the following rule is added:

$$\text{Rule } 0: \text{ If } A_j^1 \text{ AND } A_j^2 \text{ AND } \ldots \text{ AND } A_j^n \text{ THEN } b_0 \quad (3)$$

where $b_0 \in \mathbb{R}$, the membership functions $a_0^k(x_k) \equiv 1$ for $k = 1, \ldots, n$ and any $x_p \in \mathbb{R}^n$. Consequently, the input output mapping becomes

$$\mathcal{F}(x_p) = \frac{b_0 + \sum_{j=1}^{m} b_j \prod_{k=1}^{n} a_j^k(x_k)}{1 + \sum_{j=1}^{m}\prod_{k=1}^{n} a_j^k(x_k)}. \quad (4)$$

A classifier associates class labels with input features, i.e., it is essentially a mapping from the input space to the set of class labels. In binary case, thresholding is one of the simplest ways to transform $\mathcal{F}(x_p)$ to class labels $+1$ or -1.

Considering the FRBCS with $m+1$ fuzzy rules where Rule 0 is given by (3), then the system induces a binary fuzzy classifier, f , with decision rule

$$f(x_p) = \text{sign}(\mathcal{F}(x_p) + t) \quad (5)$$

where $t \in \mathbb{R}$ is a threshold. We can assume $t = 0$ without loss of generality.

The membership functions for a binary fuzzy classifier defined above could be any function from $\mathbb{R}$ to $[0, 1]$. We narrow our interests to a class of membership functions, $a_j^k : \mathbb{R} \to [0, 1]$, $j = 1, \ldots, m$, which are generated from a reference function a^k through location transformation [8], and the classifiers defined on them. In [5] well-known types of reference functions can be found, like the symmetric triangle and the gaussian function.

As consequence of the presented formulation, the decision rule of our binary fuzzy classifier can be written as:

$$f(x_p) = \text{sign}\left(\sum_{j=1}^{m} b_j K(x_p, z_j) + b_0\right) \quad (6)$$

where $z_j = [z_j^1, z_j^2, \ldots, z_j^n]^T \in \mathbb{R}$ contains the location parameters of a_j^k. $K : \mathbb{R}^n \times \mathbb{R}^n \to [0, 1]$ is a translation invariant kernel defined as

$$K(x_p, z_j) = \prod_{k=1}^{n} a^k(x_p^k - z_j^k) \quad (7)$$

which is actually a Mercer Kernel [6], if it has nonnegative Fourier transform. Again, from [5] some Mercer kernels can be built using the reference functions mentioned above.

Thus, the decision rule of a binary fuzzy classifier is

$$f(x_p) = \text{sign}\left(b_0 + \sum_{j=1}^{m} b_j \prod_{k=1}^{n} a_j^k(x_p^k)\right) \quad (8)$$

2.2 Learning method: SVM approach to build PDFC

Here, we assume that the reference functions are predetermined. So the remaining question is how to find a set of fuzzy rules ($\{z_1, \ldots, z_m\}$ and $\{b_0, \ldots, b_m\}$) from the given training so that the PDFC method has good generalization.

As given in (7), for a PDFC, a Mercer kernel can be constructed from the positive definite reference functions. The kernel implicitly defines a nonlinear mapping Φ that maps $\mathbb{X}$ into a kernel-induced feature space $\mathbb{F}$. Theorem 3.12 in [5] states that the decision rule of a PDFC can be viewed as a hyperplane in $\mathbb{F}$. It is well-known that the SVM algorithm finds a separating hyperplane with good generalization by reducing the empirical risk and, at the same time, controlling the hyperplane margin [21]. Thus we can use the SVM algorithm to find an optimal hyperplane in $\mathbb{F}$. Once we get such a hyperplane, fuzzy rules can be easily extracted. The whole procedure is described by the following algorithm 1.

Algorithm 1 SVM learning for PDFC

INPUTS: Positive definite reference functions $a^k(x_p)$ associated with n input variables, and a set of training samples
OUTPUTS: A set of fuzzy rules parameterized by z_j, b_j, and m. z_j contains the location parameters of the `IF`-part membership functions of the jth fuzzy rule, b_j is the `THEN`-part constant of the jth fuzzy rule, and $m+1$ is the number of fuzzy rules.
Steps:
1 Construct a Mercer kernel, K, from the given positivedefinite reference functions according to (7).
2 Construct an SVM to get a decision rule of the form

$$f(x) = sign\left(\sum_{i \in S} y_i \alpha_i K(x, x_i) + b\right),$$

being S the index set of the support vectors:
2.1 Assign some positive number to the cost C, and solve the quadratic program defined by the proper SVM to get the Lagrange multipliers α_i.
2.2 Find b (details can be found in, for example, [18]).
2.3 Extracting fuzzy rules from the decision rule of the SVM:
$b0 \leftarrow b$
$j \leftarrow 1$
for $i = 1$ **to** l **do**
  if $\alpha_i > 0$ **then**
    $z_j \leftarrow x_i$
    $b_j \leftarrow y_i \alpha_i$
    $j \leftarrow j + 1$
  end if
end for
$m \leftarrow j - 1$

In our study, we have considered the parameters values recommended by the authors. They are summarized as follows:

- $C = 100$ (weight of the classification error)
- $d = 0.25$ (parameter used by the reference functions)
- Type of reference functions: Gaussian

3 Data Complexity Measures Based on the Separability of Classes

In this section we describe the three metrics we have used in this contribution, with their correspondent acronym.

For our study, we will examine three measures of separability of classes from [10] which offer information for the PDFC method. They are described next.

- **N1**: fraction of points on class boundary. This method constructs a class-blind minimum spanning tree over the entire data set, and counts the number of points incident to an edge going across the two classes. The fraction of

ISBN: 978-989-95079-6-8

such points over all points in the data set is used as a measure. For two heavily interleaved classes, a majority of points are located next to the class boundary. However, the same can be true for a sparsely sampled linearly separable problem with margins narrower than the distances between points of the same class.

- **N2**: ratio of average intra/inter class Nearest Neighbour (NN) distance. For each input instance x_p, we calculate the distance to its nearest neighbour within the class (intraDist(x_p)) and the distance to nearest neighbour of any other class (interDist(x_p)). Then, the result is the ratio of the sum of the intra-class distances to the sum of the inter-class distances for each input example, i.e.,

$$N2 = \frac{\sum_{i=0}^{m} intraDist(x_i)}{\sum_{i=0}^{m} interDist(x_i)},$$

where m is the number of examples in the data set. This metric compares the within-class spread with the distances to the nearest neighbours of other classes. Low values of this metric suggest that the examples of the same class lay closely in the feature space. Large values indicate that the examples of the same class are disperse. It is sensitive to the classes of the closest neighbours to a point, and also to the difference in magnitude of the between-class distances and that of the within-class distances.

- **N3**: error rate of 1-NN classifier. This is simply the error rate of a nearest-neighbour classifier measured with the training set. The error rate is estimated by the leave-one-out method. The measure denotes how close the examples of different classes are. Low values of this metric indicate that there is a large gap in the class boundary.

4 Experimental Study: Analysis of the PDFC with Data Complexity Measures

In this Section we analyse the obtained results for the PDFC method. First, in Subsection 4.1 we present the experimental framework, with the data sets generation method, accuracy validation scheme, and the global average results of the PDFC method. Next we determine several rules based on PDFC's behaviour in Subsection 4.2. Finally we analyse the collective evaluation of the set of rules in Subsection 4.3.

4.1 Experimental Framework: Data Sets Generation

We evaluate the PDFC method on a set of 438 binary classification problems. These problems are generated from pairwise combinations of the classes of 20 problems from the University of California, Irvine (UCI) repository [2]. These are *iris, wine, new-thyroid, solar-flare, led7digit, zoo, yeast, tae, balanced, car, contraceptive, ecoli, hayes-roth, shuttle, australian, pima, monks, bupa, glass, haberman* and *vehicle*.

In order to do that, first we take each data set and extract the examples belonging to each class. Then we construct a new data set with the combination of the examples from two different classes. This will result in a new data set with only 2 classes and the examples which have two such classes as output. We perform this process for every possible pairwise combination of classes. However, if an obtained data set with this procedure proves to be linearly-separable, we discard it (since we could classify it with a linear classifier with no error). The complexity measure L1 from [10] indicates if a problem is linearly-separable if its value is zero, so every data set with a L1 value of zero will be discarded.

This method for generating binary data sets is limited by the proper combinatorics, and we can only obtain over 200 new data sets with the original 20 data sets with this first approach. In order to obtain more data sets, we group the classes two by two, that is, we create a new binary data set, and each of its two classes are the combination of two original classes each. For this second approach we have used *ecoli*, *glass* and *flare* data sets, since they have a high number of class labels. Again, those data sets with a L1 value of zero are discarded.

In order to measure the PDFC performance, we have applied a 10-fcv validation scheme. In Table 1 we show the global Training and Test accuracy obtained by the PDFC method.

Table 1: Global Average PDFC Training and Test Accuracy

PDFC Global % Accuracy Training	94.06%
PDFC Global % Accuracy Test	91.22%

4.2 Determination of Rules Based on the PDFC Behaviour

In the following we present the results of the execution over the 438 data sets summarized in Figures 1 to 3.

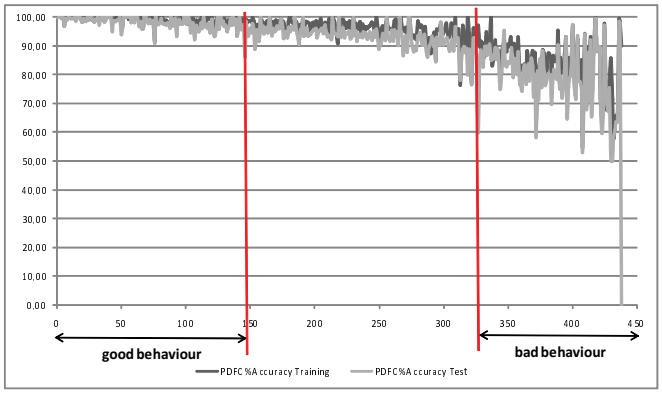


Figure 1: PDFC accuracy in Training/Test sorted by N1

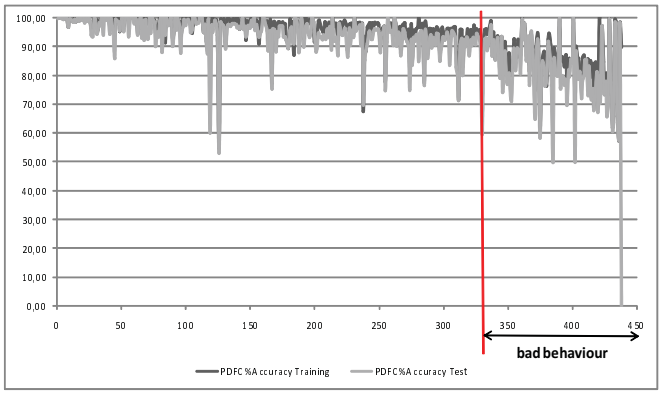


Figure 2: PDFC accuracy in Training/Test sorted by N2

For each complexity measure (N1, N2 and N3), the data sets are sorted by the ascending value of the corresponding complexity measure, and put altogether in a Figure. In the X

ISBN: 978-989-95079-6-8

Table 3: Rules with one metric obtained from the intervals

Id.	Rule	Support	%Training	Training Diff.	% Test	Test Diff.
R1+	If N1[X] $<$ 0.089 then *good behaviour*	33.11%	99.09%	4.91%	98.15%	6.55%
R2+	If 0 $<$ N3[X] $<$ 0.047 then *good behaviour*	28.31%	98.92%	4.86%	97.70%	6.48%
R1-	If N1[X] $\geq$ 0.25 then *bad behaviour*	25.57%	84.64%	-9.42%	79.01%	-12.21%
R2-	If N2[X] $>$ 0.5196 then *bad behaviour*	24.89%	86.14%	-7.92%	80.88%	-10.34%
R3-	If N3[X] $>$ 0.175 then *bad behaviour*	19.41%	82.64%	-11.42%	76.77%	-14.45%

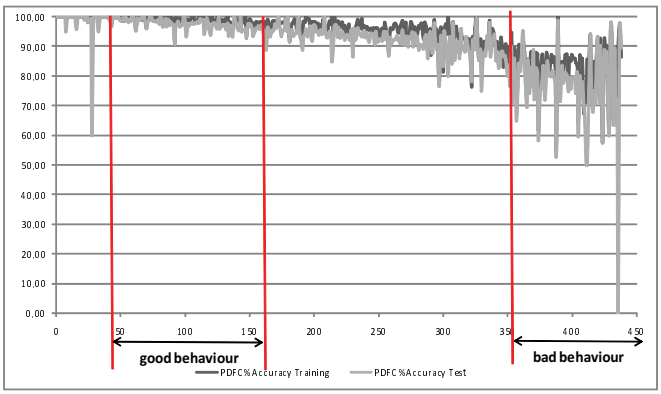


Figure 3: PDFC accuracy in Training/Test sorted by N3

axis we represent the data sets, not the complexity measure value, and the Y axis depicts the accuracy obtained both in training and test. The reason to do so is to give each data set the same space in the graphic representation. For those measures where we can find different *ad-hoc* intervals which present *good* or *bad behaviour* of the PDFC, we use a vertical line to delimit the interval of the region of interest.

- We understand for *good behaviour* an average high test accuracy in the interval, as well as the absence of over-fitting.
- By *bad behaviour* we refer to the presence of over-fitting and/or average low test accuracy in the interval.

In Table 2 we have summarized the intervals found ad-hoc from Figures 1 to 3.

Table 2: Significant intervals

Interval	PDFC Behaviour
$N1 < 0.089$	*good behaviour*
$0 < N3 < 0.047$	*good behaviour*
$N1 \geq 0.25$	*bad behaviour*
$N2 > 0.5196$	*bad behaviour*
$N3 > 0.175$	*bad behaviour*

From these ad-hoc intervals we construct several rules that model the performance of the FRBCS we have used. In Table 3 we have summarized the rules derived from Table 2. Given a particular data set X, we get the complexity measure of X with the notation $CM[X]$. Table 3 is organised with the following columns.

- The first column corresponds to the identifier of the rule for further references.
- The "Rule'" column presents the rule itself.
- The third column "Support" presents the percentage of data sets which verifies the antecedent of the rule.
- The column "% Training" shows the average accuracy in training of all the examples which are covered by the rule.
- The column "Training Diff." contains the difference between the training accuracy of the rule and the training accuracy across all 438 data sets.
- The column "% Test" shows the average accuracy in test of all the examples which are covered by the rule.
- The column "Test Diff." contains the difference between the test accuracy of the rule and the test accuracy across all 438 data sets.

As we can see in Table 3, the positive rules (denoted with a "+" symbol in their identifier) always show a positive difference with the global average, both in training and test accuracy. The negative ones (with a "-" symbol in their identifier) verify the opposite case. The support of the rules shows us that we can characterize a wide range of data sets and obtain significant differences in accuracy.

From this set of rules we can state that a low N1 value results in a good behaviour of the PDFC method. A low N3 value obtains the same results. In the other hand, a high value in the N1 metric produces a bad behaviour of the PDFC considered in our analysis. A high N3 value will also produce a bad behaviour of the PDFC method. With similar outcome, if N2 presents a high value, the PDFC method will obtain bad behaviour.

Although we have obtained some interesting rules, we can extend our study by considering the combination of these complexity metrics in order to obtain more precise and descriptive rules.

4.3 Collective Evaluation of the Set of Rules

The objective of this section is to analyse the good rules jointly, and the bad rules together as well. Thus we can arrive at a more general description, with wider support, of the behaviour of the PDFC with these joint rules. We perform the disjunctive combination of all the positive rules to obtain a single rule, and all the negative ones, so we obtain another rule.

ISBN: 978-989-95079-6-8

Table 4: Disjunction Rules from all simple rules

Id.	Rule	Support	%Training	Training Diff.	% Test	Test Diff.
PRD	If R1+ or R2+ then *good behaviour*	37.44%	98.99%	4.93%	97.90%	6.68%
NRD	If R1- or R2- or R3- then *bad behaviour*	31.51%	86.23%	-7.83%	81.24%	-9.98%
not characterised	If not PRD and not NRD then *good behaviour*	31.05%	96.06%	2.00%	93.27%	2.05%

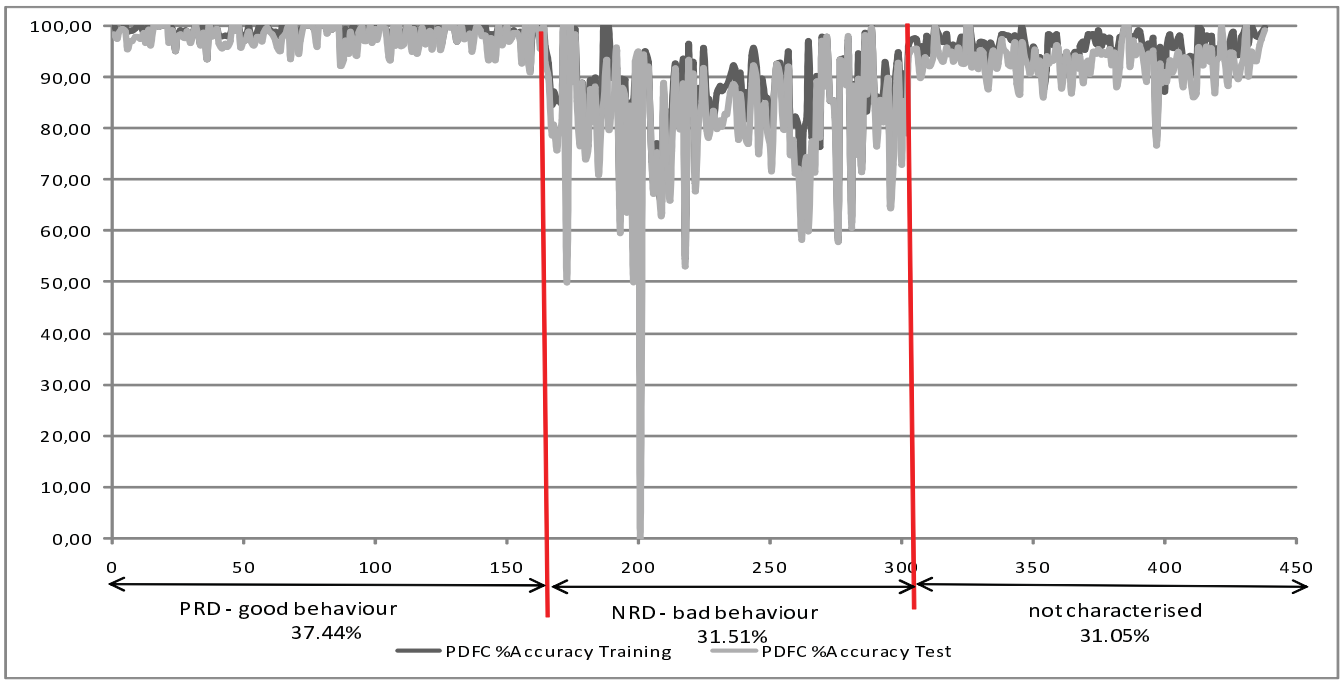


Figure 4: Three blocks representation for PRD, NRD and not covered data sets for PDFC

The new disjunctive rule will be activated if any of the component rules' antecedents are verified. In Table 4 we summarize both disjunctions, and a third rule representing those data sets which are not charaterised by either disjunction rules.

From the collective rules we can observe that the support has been increased from the single rules both for the Positive Rule Disjunction (PRD) and Negative Rule Disjunction (NRD). In the other hand, the Test and Training Accuracy Differences are similar to the single rules from Table 3. Since there are no data sets in PRD and NRD simultaneously, we can consider three blocks of data sets with their respective support, as depicted in Figure 4 (with no particular data set order within each block):

- The first block (the left-side one) represents the data sets covered by the PRD rule. They are the data sets recognized as being those in which the PDFC has good accuracy.
- The second block (the middle one) plots the data sets for the rule NRD, which are bad data sets for the PDFC method considered.
- The third and last block (the right-side one) contains the unclassified data sets by the previous two rules.

We can see that almost the 70% of the analysed data sets are covered by these two rules, and hence the *good behaviour* and *bad behaviour* consequents represent well the accuracy of PDFC methods.

5 Concluding Remarks

We have performed a study over a set of binary data sets with the PDFC method. We have computed some data complexity measures for the data sets in order to obtain intervals of such metrics in which the method's performance is significantly good or bad. We have constructed descriptive rules, and studied the interaction between the intervals and the proper rules.

We have obtained two rules which are simple and precise to describe both good and bad performance of the PDFC. Furthermore, we present the possibility of determining which data sets PDFC would performs well or badly prior to their execution, using the Data Complexity measures.

We must point out that this is a particular study for one specific method, the PDFC. On the other hand, this work presents a new challenge that could be extended to other FRBCSs, to analyse their domains of competence, and to develop new measures which could give more information on the behaviours of FRBCSs for pattern recognition.

Acknowledgment

This work has been supported by the Spanish Ministry of Science and Technology under Project TIN2008-06681-C06-01.

ISBN: 978-989-95079-6-8

J. Luengo holds a FPU scholarship from Spanish Ministry of Innovation and Science.

References

[1] P. Angelov, E. Lughofer and X. Zhou. Evolving fuzzy classifiers using different model architectures. *Fuzzy Sets and Systems*, 159(23):3160–3182, 2008.

[2] A. Asuncion and D.J. Newman. UCI Machine Learning Repository [http://www.ics.uci.edu/ mlearn/MLRepository.html]. Irvine, CA: University of California, School of Information and Computer Science, 2007.

[3] R. Baumgartner, R.L. Somorjai. Data complexity assessment in undersampled classification. *Pattern Recognition Letters*, 27:1383–1389, 2006.

[4] E. Bernadó-Mansilla, T.K. Ho. Domain of Competence of XCS Classifier System in Complexity Measurement Space. *IEEE Transactions on Evolutionary Computation*, 9(1):82–104, 2005.

[5] Y. Chen and J.Z. Wang. Support Vector Learning for Fuzzy Rule-Based Classification Systems. *IEEE Transactions on Fuzzy Systems*, 11(6):716–728,2003.

[6] N. Cristianini and J. Shawe-Taylor. *An introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[7] M. Dong, R. Kothari. Feature subset selection using a new definition of classificabilty. *Pattern Recognition Letters*, 24:1215–1225, 2004.

[8] D. Dubois and H. Prade. Operations on fuzzy numbers. *Int. J. Syst. Sci.*, 9(6):613–626, 1978.

[9] A. Fernández, S. García, M.J. del Jesus and F. Herrera. A study of the behaviour of linguistic fuzzy rule based classification systems in the framework of imbalanced data-sets. *Fuzzy Sets and Systems*, 159(18):2378–2398, 2008.

[10] T.K. Ho, M. Basu. Complexity Measures of Supervised Classification Problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(3):289–300, 2002.

[11] M. Basu, T.K. Ho. *Data Complexity in Pattern Recognition*. Springer, 2006.

[12] H. Ishibuchi, T. Nakashima, M. Nii. *Classification and modeling with linguistic information granules: advanced approaches to linguistic data mining*. Springer, Berlin, 2004.

[13] M.Z. Jahromi and M. Taheri. A proposed method for learning rule weights in fuzzy rule-based classification systems. *Fuzzy Sets and Systems*, 159(4):449-459, 2008

[14] L. Kuncheva L. *Fuzzy classifier design*. Springer, Berlin, 2000.

[15] Y. Li, M. Dong, R. Kothari. Classifiability-Based Omnivariate Decision Trees. *IEEE Transactions on Neural Networks*, 16(6):1547–1560, 2005.

[16] E.G. Mansoori, M.J. Zolghadri and S.D. Katebi. SGERD: A steady-state genetic algorithm for extracting fuzzy classification rules from data. *IEEE Transactions on Fuzzy Systems*, 16(4):1061–1071, 2008.

[17] R.A. Mollineda, J.S. Sánchez, J.M. Sotoca. Data Characterization for Effective Prototype Selection. *First edition of the Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA 2005). Lecture Notes in Computer Science 3523, Springer-Verlag*, 27–34, 2005.

[18] J. Platt. Machines using Sequential Minimal Optimization. In *B. Schoelkopf, C. Burges and A. Smola (Eds.), Advances in Kernel Methods - Support Vector Learning*, 1998.

[19] J.S. Sánchez, R.A. Mollineda, J.M. Sotoca. An analysis of how training data complexity affects the nearest neighbor classifiers. *Pattern Analysis and Applications*, 10(3):189–201, 2007.

[20] S. Singh. Multiresolution Estimates of Classification Complexity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1534–1539, 2003.

[21] V. Vapnik. *Statistical Learning Theory*. New York: Wiley, 1998.

ISBN: 978-989-95079-6-8

Learning Concurrently Granularity, Membership Function Parameters and Rules of Mamdani Fuzzy Rule-based Systems

Michela Antonelli, Pietro Ducange, Beatrice Lazzerini, Francesco Marcelloni

Dipartimento di Ingegneria dell'Informazione: Elettronica, Informatica, Telecomunicazioni
University of Pisa
Via Diotisalvi 2, 56122 Pisa
Email: {michela.antonelli, p.ducange, b.lazzerini, f.marcelloni}@iet.unipi.it

Abstract—*In this paper we tackle the issue of generating Mamdani fuzzy rule-based systems with optimal trade-offs between complexity and accuracy by using a multi-objective genetic algorithm, which concurrently learns rule base, granularity of the input and output partitions and membership function parameters. To this aim, we exploit a chromosome composed of three parts, which codify, respectively, the rule base, and, for each variable, the number of fuzzy sets and the parameters of a piecewise linear transformation of the membership functions. We show the encouraging results obtained on a real world regression problem.*

Keywords— Accuracy-Interpretability Trade-off, Granularity Learning, Mamdani Fuzzy-Rule-Based Systems, Multi-objective Evolutionary Algorithms, Piecewise Linear Transformation.

1 Introduction

In the last years, the problem of finding the right trade-off between interpretability and accuracy of Mamdani fuzzy rule-based systems (MFRBSs) [1] has arisen a growing interest in the fuzzy community [2]. To this aim, Multi-Objective Evolutionary Algorithms (MOEAs) have been extensively used for tuning or learning the data base (DB) and the rule base (RB) of the MFRBSs [3][4]. All these approaches generate approximated Pareto fronts with non-dominated solutions in the interpretability-accuracy space. For example, in [5], authors use a predefined DB and evolve only the RB, while in [6] the membership function (MF) parameters are adapted to a specific context exploiting a predefined RB. Furthermore, in [7] and [8] the tuning of the DB is performed together with a rule selection.

The ideal approach would be to learn concurrently DB and RB. So far, only approximations of this ideal approach have been proposed such as to learn simultaneously the overall RB and, for the DB, only the MF parameters [9][10] or the granularities of the uniform partitions defined on the input and output variables [11][12].

On the other hand, both granularity and MF parameters are a critical factor in MFRBS generation [6][13], since both affect accuracy and interpretability. Indeed, the former fixes an upper bound to the number of rules in the MFRBS and the latter adapts the meaning of the linguistic values to the specific application context.

In this paper, we propose a multi-objective evolutionary approach to generate MFRBSs with different trade-offs between complexity and accuracy. The main novelty of our approach is that the RB, the granularity of the input and output partitions and the MF parameters of each fuzzy set are learnt concurrently during the evolutionary process. To deal with RBs defined on different granularities, we exploit the concept of virtual RBs introduced in [12]. Further, the learning of the MF parameters is performed by using a piecewise linear transformation [14][15], which allows us to obtain a high modelling capability with a limited number of parameters.

Chromosomes are composed of three parts, which codify the RB, and, for each linguistic variable, the number of fuzzy sets and the parameters of the piecewise linear transformation, respectively. In the evolutionary process, we adopt the mating operators proposed in [5] for the fist part of the chromosome, and standard crossover and mutation operators for the second and third parts, respectively.

Our approach has been tested on a real world regression problem, with nine input variables, and has provided Pareto fronts with solutions characterized by better trade-offs between accuracy and complexity than solutions belonging to Pareto fronts generated by learning only the rule base with a fixed DB, and concurrently the rule base and the MF parameters with fixed granularities, respectively.

2 Mamdani Fuzzy Systems

Let $\mathbf{X} = \{X_1, ..., X_f, ..., X_F\}$ be the set of input variables and X_{F+1} be the output variable. Let $U_f (f = 1, ..., F+1)$ be the universe of the f^{th} variable. Let $P_f = \{A_{f,1}, ..., A_{f,T_f}\}$ be a fuzzy partition of T_f fuzzy sets on variable X_f. An MFRBS is composed of *M* rules expressed as:

$$R_m: \text{IF } X_1 \textbf{ is } A_{1,j_{m,1}} \text{ AND } \ldots \text{ AND } X_F \textbf{ is } A_{F,j_{m,F}} \text{ THEN } X_{F+1} \textbf{ is } A_{F+1,j_{m,F+1}} \quad (m = 1, ..., M) \qquad (1)$$

where $j_{m,f} \in [1, T_f]$ identifies the index of the fuzzy set (among the T_f fuzzy sets of partition P_f), which has been selected for X_f in rule R_m.

To take the "don't care" condition into account [16] a new fuzzy set $A_{f,0}$ $(f = 1, ..., F)$ is added to all the F input partitions P_f. This fuzzy set is characterized by a membership function equal to 1 on the overall universe.

The terms $A_{f,0}$ allow generating rules which contain only a

ISBN: 978-989-95079-6-8

subset of the input variables. It follows that $j_{m,f} \in [0, T_f]$, $f = 1,..., F$, and $j_{m,F+1} \in [1, T_{F+1}]$.

An MFRBS can be completely described by the following matrix $J \in \mathbb{N}^{M \times (F+1)}$

$$J = \begin{bmatrix} j_{1,1} & \cdots & j_{1,F} & j_{1,F+1} \\ \cdots & \cdots & \cdots & \cdots \\ j_{m,1} & \cdots & j_{m,F} & j_{m,F+1} \\ \cdots & \cdots & \cdots & \cdots \\ j_{M,1} & \cdots & j_{M,F} & j_{M,F+1} \end{bmatrix}$$

where the generic element (m, f) indicates that fuzzy set $A_{f,j_{m,f}}$ has been selected for variable X_f in rule R_m.

We adopt the product and the weighted average method as AND logical operator and defuzzification method, respectively.

Given a set of *N* input observations $\mathbf{x}_n = [x_{n,1},..., x_{n,F}]$, with $x_{n,f} \in \Re$, and the set of the corresponding outputs $x_{n,F+1} \in \Re$, $n = 1,..., N$, we apply a multi-objective evolutionary algorithm which produces a set of MFRBSs with different trade-offs between accuracy and complexity by learning simultaneously the RB, the granularity of the partitions of each variable and the MF parameters. The choice of the appropriate trade-off between accuracy and complexity depends on the particular application.

3 Granularity and MF Parameter Learning

3.1 Granularity Learning

To determine simultaneously the granularity of partitions and the RB in the evolutionary process is not an easy task since the RB depends on the number of fuzzy sets used to partition the variables. In [12] we have introduced the concept of *virtual RB*. A virtual RB is composed of rules defined by considering the variables partitioned with a user-defined maximum number $T_{\max}$ of fuzzy sets (*virtual partition*). All the mating operators are applied to virtual RBs: the actual granularity is used only in the computation of the fitness. In practice, we generate virtual RBs and assess their quality using each time different "lens" depending on the actual number of fuzzy sets used to partition the single variables. Thus, we do not worry about the actual granularity in applying crossover and mutation operators.

To map the virtual RB defined on variables uniformly partitioned with $T_{\max}$ fuzzy sets into a concrete RB defined on variables uniformly partitioned with T_f fuzzy sets, we adopt a mapping strategy. Let X_f **is** $\hat{A}_{f,h}$, $h \in [0, T_{\max}]$, be a generic fuzzy proposition defined in a rule of the virtual RB. Then, the proposition will be mapped to X_f **is** $\tilde{A}_{f,s}$, with $s \in [0, T_f]$, where $\tilde{A}_{f,s}$ is the fuzzy set more similar to $\hat{A}_{f,h}$ among the T_f fuzzy sets $\hat{A}_{f,j}$ defined on X_f. For the sake of simplicity, we have trivially considered as similarity measure the distance between the centroids of the two fuzzy sets. If there are two fuzzy sets in $\tilde{P}_f = \{\tilde{A}_{f,1},..., \tilde{A}_{f,T_f}\}$ with centroids at the same distance from the centroid of $\hat{A}_{f,h}$, we choose randomly one of the two fuzzy sets.

Note that different rules of the virtual RB can be mapped to equal rules in the concrete RB. This occurs because distinct fuzzy sets defined on the partitions used in the virtual RB can be mapped to the same fuzzy set defined on the partitions used in the concrete RB. In the case of equal rules, only one of these rules is considered in the concrete RB. The original different rules are, however, maintained in the virtual RB. Indeed, when the virtual RB will be interpreted by using different "lens", all these rules can again be meaningful and contribute to increase the accuracy of the MFRBS. Thus, the concept of virtual RB allows us to explore the search space and concurrently exploiting the optimal solutions achieved during the evolutionary process.

3.2 MF Parameter Learning

We approach the problem of learning the MF parameters by using a piecewise linear transformation [14][15]. The transformation is described in Fig. 1 for a generic variable X_f. In the following, we assume that the interval ranges of the original and transformed variables are identical. Further, we consider triangular fuzzy sets $A_{f,j}$ defined by the tuple $(a_{f,j}, b_{f,j}, c_{f,j})$, where $a_{f,j}$ and $c_{f,j}$ correspond to the left and right extremes of the support of $A_{f,j}$, and $b_{f,j}$ to the core. Finally, given a generic partition $P_f = \{A_{f,1},..., A_{f,T_f}\}$, we assume that, for $j = 2 \ldots T_f - 1$, $b_{f,j} = c_{f,j-1}$ and $b_{f,j} = a_{f,j+1}$, and $a_{f,1} = b_{f,1}$ and $b_{f,T_f} = c_{f,T_f}$.

Before passing the input value x_f to the MFRBS, we apply the transformation $t(x_f)$. Thus, we have that:

$$A_{f,j}(x_f) = \tilde{A}_{f,j}(t(x_f)) = \tilde{A}_{f,j}(\tilde{x}_f)$$

where $\tilde{A}_{f,j}$ and $A_{f,j}$ are two generic fuzzy sets from the uniform and non-uniform fuzzy partitions, respectively. In those regions where *t* has a high value of the derivative (high slope of the lines), the fuzzy sets $A_{f,j}$ are narrower; otherwise, the fuzzy sets $A_{f,j}$ are wider.

As observed in [14], the transformation *t* must be non-decreasing. Further, the fuzzy sets $A_{f,j}$ of the non-uniform fuzzy partition induced by the transformation *t* are not necessarily of triangular shape. This could bring to fuzzy partitions difficultly interpretable. To preserve the shape of the MFs, we force the change of slopes in *t* to coincide with the cores of the fuzzy sets in the partitions.

Let $b_{f,1},..., b_{f,T_f}$ and $\tilde{b}_{f,1},..., \tilde{b}_{f,T_f}$ be the cores of $A_{f,1},..., A_{f,T_f}$ and $\tilde{A}_{f,1},..., \tilde{A}_{f,T_f}$, respectively. Transformation *t* can be defined as:

$$t(x_f) = \frac{\tilde{b}_{f,j} - \tilde{b}_{f,j-1}}{b_{f,j} - b_{f,j-1}}(x_f - b_{f,j-1}) + \tilde{b}_{f,j-1}, \quad b_{f,j-1} \le x_f < b_{f,j},$$

with *j*=2…T_f.

Once fixed the granularity of the partitions, $\tilde{b}_{f,1},..., \tilde{b}_{f,T_f}$ are fixed and therefore known. Further, $b_{f,1}$ and b_{f,T_f} coincide

ISBN: 978-989-95079-6-8

with the extremes of the universe U_f of X_f. Thus, $t(x_f)$ depends on T_f –2 parameters, that is, $t(x_f; b_{f,2},...,b_{f,T_f-1})$. Once fixed $b_{f,2},...,b_{f,T_f-1}$, the partition $P_f = \{A_{f,1},...,A_{f,T_f}\}$ can be obtained simply by transforming the three points $(\tilde{a}_{f,j}, \tilde{b}_{f,j}, \tilde{c}_{f,j})$, which describe the generic fuzzy set $\tilde{A}_{f,j}$, into $(a_{f,j}, b_{f,j}, c_{f,j})$ applying $t^{-1}(\tilde{x}_f)$.

We define the piecewise linear transformation on the maximum granularity. When we reduce the granularity, to maintain the original shape of the MFs, we do not apply the piecewise linear transformation to all the points of the universe, but only to the three points, which define the generic triangular MF. In practice, we transform $(\tilde{a}_{f,j}, \tilde{b}_{f,j}, \tilde{c}_{f,j})$ into $(a_{f,j}, b_{f,j}, c_{f,j})$ by applying $t^{-1}(\tilde{x}_f)$. Fig. 2 shows an example of this transformation for granularity 5 by using the piecewise linear transformation in Fig. 1, defined with granularity 7.

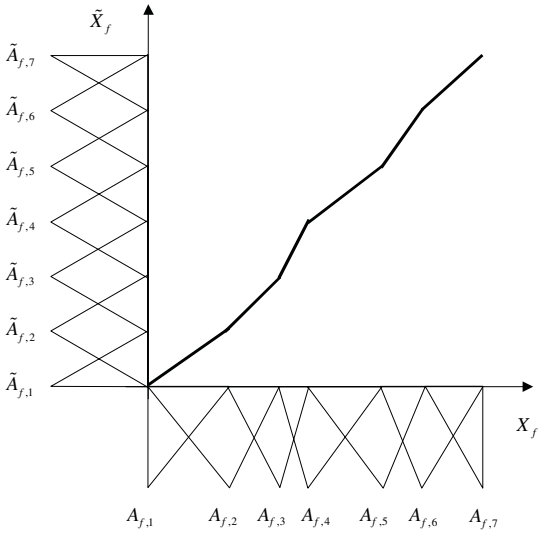


Figure 1: An example of piecewise linear transformation.

4 The Multi-Objective Evolutionary Approach

4.1 Chromosome coding

Each solution is codified by a chromosome C composed of three parts (C_1, C_2, C_3), which define the virtual RB, and the granularities and the piecewise linear transformations of all the variables, respectively. In particular, C_1 encodes the virtual RB by considering that each variable X_f is uniformly partitioned by using $T_{\max}$ fuzzy sets, that is, $P_f = \{A_{f,1},...,A_{f,T_{\max}}\}$. As described in [5], C_1 is composed of $M \cdot (F+1)$ natural numbers where M is the number of rules currently present in the virtual RB. The RB (defined as *concrete RB*) used to compute the fitness is obtained by means of the mapping strategy using the actual granularities fixed by C_2.

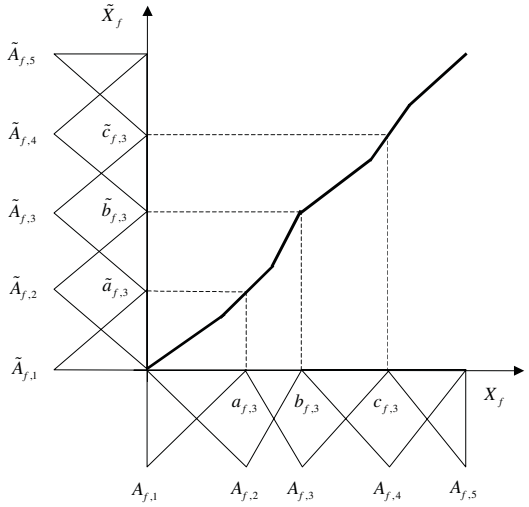


Figure 2: An example of piecewise linear transformation with granularity different from $T_{\max}$.

C_2 is a vector containing $F+1$ natural numbers: the f^{th} element of the vector contains the number $T_f \in [2, T_{\max}]$ of fuzzy sets which partition the linguistic variable X_f. $T_{\max}$ is fixed by the user and is the same for all the linguistic variables.

C_3 is a vector containing $F+1$ vectors of $T_{\max}-2$ real numbers: the f^{th} vector contains the $[b_{f,2},...,b_{f,T_{\max}-1}]$ points which define the piecewise linear transformation for the linguistic variable X_f. To preclude that the piecewise linear transformation can become decreasing, we force $b_{f,j}$ to vary in $\left[\tilde{b}_{f,j} - \frac{\tilde{b}_{f,j} - \tilde{b}_{f,j-1}}{2}, \tilde{b}_{f,j} + \frac{\tilde{b}_{f,j+1} - \tilde{b}_{f,j}}{2}\right]$, $\forall j \in [2, T_{\max}-1]$.

Each chromosome is associated with a bi-dimensional objective vector. The first element of the vector measures the complexity of the MFRBS as the number of propositions used in the antecedents of the rules contained in the concrete RB (the number of antecedents may be different, as explained in the next subsection, between the virtual and concrete RBs). The second element assesses the accuracy as the half of the mean square error (MSE) between the output of the MFRBS and the expected output.

4.2 Genetic operators

In order to generate the offspring populations, we exploit both crossover and mutation. We apply separately the one-point crossover to C_1 and C_2 and the BLX-α crossover, with α = 0.5, to C_3. Let s_1 and s_2 be two selected parent chromosomes. The common gene for C_1 is chosen by extracting randomly a number in $[M_{\min}, \rho_{\min}]$, where $M_{\min}$ is the minimum number of rules, which must be present in a rule base, and $\rho_{\min}$ is the minimum number of rules in s_1 and s_2. The common gene for C_2 is extracted randomly in $[1, F+1]$.

ISBN: 978-989-95079-6-8

As regards mutation, we apply two mutation operators for C_1. The first operator adds γ rules to the virtual RB, where γ is randomly chosen in $[1, \gamma_{\max}]$. The upper bound $\gamma_{\max}$ is fixed by the user. If $\gamma + M > M_{\max}$, then $\gamma = M_{\max} - M$. For each rule R_m added to the chromosome, we generate a random number $v \in [1, F]$, which indicates the number of input variables used in the antecedent of the rule. Then, we generate v natural random numbers between 1 and F to determine the input variables which compose the antecedent part of the rule. Finally, for each selected input variable f, we generate a random natural number $j_{m,f}$ between 1 and $T_{\max}$, which determines the fuzzy set $A_{f,j_{m,f}}$ to be used in the antecedent of rule R_m in the virtual RB. To select the consequent fuzzy set $A_{F+1,j_{m,F+1}}$, a random number between 1 and $T_{\max}$ is generated.

The second mutation operator randomly changes δ elements of the matrix J associated with the virtual RB. The number δ is randomly generated in $[1, \delta_{\max}]$. The upper bound $\delta_{\max}$ is fixed by the user. For each element to be modified, a number is randomly generated in [0, $T_{\max}$].

The mutation applied to C_2 randomly chooses a gene $f \in [1, F+1]$ and changes the value of this gene by randomly adding or subtracting 1. If the new value is lower than 2 and larger than $T_{\max}$, then mutation is not applied.

The mutation applied to C_3 first chooses randomly a variable $f \in [1, F+1]$, then extracts a random value $j \in [2, T_{\max} - 1]$ and changes the value of $b_{f,j}$ to a random value in the allowed interval $\left[\tilde{b}_{f,j} - \frac{\tilde{b}_{f,j} - \tilde{b}_{f,j-1}}{2}, \tilde{b}_{f,j} + \frac{\tilde{b}_{f,j+1} - \tilde{b}_{f,j}}{2}\right]$.

We experimentally verified that these mating operators ensure a good balancing between exploration and exploitation, thus allowing the multi-objective evolutionary algorithm described in the next subsection to create good approximations of the Pareto fronts.

4.3 Multi-objective evolutionary algorithm

We adopted the (2+2)M-PAES proposed in [5]. Unlike classical (2+2)PAES, which uses only mutation to generate new candidate solutions, (2+2)M-PAES exploits both crossover and mutation. Further, in (2+2)M-PAES, current solutions are randomly extracted at each iteration rather than maintained until they are not replaced by solutions with particular characteristics.

Fig. 3 shows a pseudo-code which describes the application scheme of the different operators to generate the offspring solutions o_1 and o_2 from the selected parents s_1 and s_2. Note that P_{c1}, P_{c2} and P_{c3} represent the probabilities of applying the crossover operator on C_1, C_2 and C_3, respectively. Similarly, P_{m1}, P_{m2} and P_{m3} represent the probabilities of applying the mutation operator on C_1, C_2 and C_3, respectively, and P_{add} and P_{inc} represent the probabilities of adding rules and of increasing the granularity of the selected variable, respectively. In Fig. 3 we also report the values of the probabilities we used in our experimentations.

At the beginning, we generate two solutions s_1 and s_2. The genes of C_2 in both the solutions are set to the maximum value $T_{\max}$, while the genes of C_1 and C_3 are randomly generated. At each iteration, the application of crossover and mutation operators produces two new candidate solutions from the current solutions s_1 and s_2. These candidate solutions are added to the archive only if they are dominated by no solution contained in the archive; possible solutions in the archive dominated by the candidate solutions are removed. Typically, the size of the archive is fixed at the beginning of the execution of the (2+2)M-PAES. In this case, when the archive is full and a new solution z has to be added to the archive, if z dominates no solution in the archive, then we insert z into the archive and remove the solution (possibly z itself) that belongs to the region with the highest crowding degree [5]. If the region contains more than one solution, then, the solution to be removed is randomly chosen.

```
P_c1 = P_c2 = P_c3 = 0.5;
P_m2 = P_m3 = 0.2;
P_add = 0.55;
P_inc = 0.85;
…
//Generate two new solutions
[s_1, s_2] = random_selection(archive)
o_1 = s_1
o_2 = s_2
if (rand() < P_c1)
    [o_1.C_1,o_2.C_1]=crossover_C_1 (s_1.C_1,s_2.C_1);
    P_m1 = 0.01;
else
    P_m1 = 1;
endif
if (rand() < P_c2)
    [o_1.C_2,o_2.C_2]=crossover_C_2 (s_1.C_2,s_2.C_2);
endif
if (rand() < P_c3)
    [o_1.C_3,o_2.C_3]=crossover_C_3 (s_1.C_3,s_2.C_3);
endif
loop i=1,2
    if (rand() < P_m1)
        if (rand() <P_add)
          o_i.C_1 = first_mutation_operator();
        else
          o_i.C_1 =second_mutation_operator();
        endif
    endif
    if (rand() < P_m2)
        if (rand <P_inc)
          o_i.C_2 = increase_granularity();
        else
          o_i.C_2 = decrease_granularity();
        endif
    endif
    if (rand() < P_m3)
       o_i.C_3 = transf_function_mutation();
    endif
endloop
```

Figure 3: Genetic operators application scheme.

ISBN: 978-989-95079-6-8

5 Experimental results

We tested our approach on a real world regression problem provided by the Bilkent University Function Approximation Repository [17]. The dataset contains the weather data of Ankara, a Turkish town, collected from 01/01/1994 to 28/05/1998. Here, the goal is to predict the mean temperature from the values of nine weather features, namely *max temperature*, *min temperature*, *dewpoint*, *precipitation*, *sea level pressure*, *standard pressure visibility*, *wind speed*, and *max wind speed*. The dataset contains 1609 instances of these features and corresponding mean temperatures. In order to asses the reliability of our approach, we performed a five-fold cross-validation and executed six trials, with different seeds, for each fold. We set $T_{\max}=5$, $M_{\min}=5$, $M_{\max}=30$, $\gamma_{\max}=5$, $\delta_{\max}=5$, the maximum number of evaluations to 300000 and the archive size to 64.

To assess the advantages of learning concurrently RB, granularities of partitions and MF parameters, we compared the results achieved by our approach with the ones obtained by applying the (2+2)M-PAES to learn only rules, that is, using only the first part of the chromosome and fixing the number of fuzzy sets in all partitions to $T_{\max}$, and to learn concurrently the RB and the MF parameters, using only the first and third parts of the chromosome with fixed granularities. We denote these two approaches as PAES-RB and PAES-SF, respectively, while the proposed approach is denoted as PAES-SFG.

Figures 4 and 5 show the average Pareto fronts achieved by the three algorithms on the training and test sets, respectively. The average Pareto front consists of the average values of the twenty most accurate solutions of each of the thirty Pareto front approximations (at least twenty solutions were always generated in all the trials for both datasets). In practice, the average Pareto fronts are obtained as follows. First, the solutions in the Pareto front approximations produced in each of the thirty trials are ordered for increasing MSE values. Then, only the twenty solutions with the lowest MSEs are retained for each Pareto front approximation. Finally, the average values, on the thirty Pareto front approximations, of complexity and MSE for these twenty solutions are computed. The choice of considering only the twenty solutions with the lowest MSEs was motivated by the observation that the other solutions are in general characterized by quite high MSEs which make these solutions impractical.

We note that the average Pareto front generated by PAES-SFG outperforms the other two average Pareto fronts both on the training and on the test sets. This proves that learning concurrently the RB, the granularities of the input and output partitions and the membership parameters allows generating FRBSs with better trade-offs between accuracy and complexity than the other two MOEAs used for comparison.

In Table 1 we show the average results corresponding to three representative points of the average Pareto fronts: the first (the most accurate), the median and the last (the least accurate) point. We refer to these average values as First, Median and Last, respectively. It is interesting to observe that the average MSEs of the most accurate solutions generated by PAES-SFG are 47% and 68%, and 33% and 56% lower than the corresponding solutions in the Pareto fronts generated by PAES-SF and PAES-RB with comparable complexity on the training and test sets, respectively.

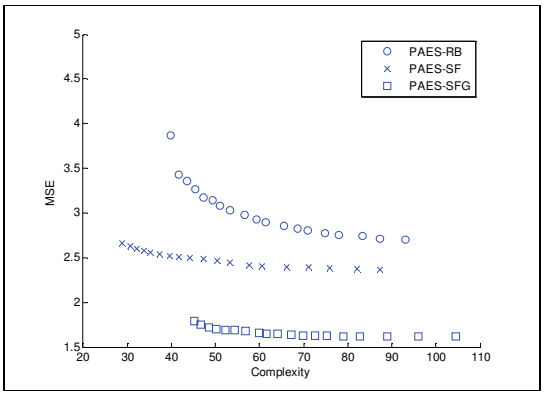


Figure 4. Average Pareto fronts on training set.

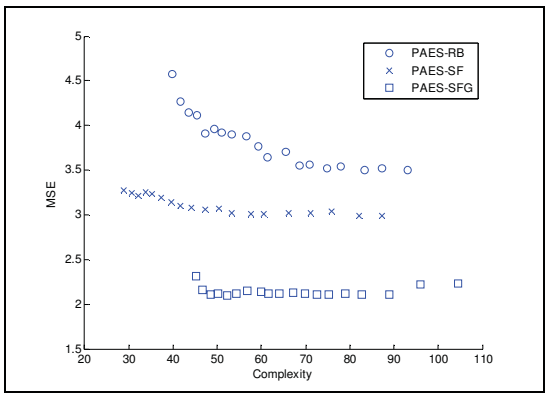


Figure 5. Average Pareto fronts on test set.

Table 1. Results obtained on three representative points of the average Pareto fronts

	MSE_{Tr}	MSE_{Ts}	Complexity	# Rules
	PAES-RB			
First	2.70±1.27	3.50±1.90	93.20±19.99	28.30±2.65
Median	2.93±1.40	3.77±2.17	59.33±17.02	22.80±4.48
Last	3.87±3.08	4.58±3.43	39.80±16.36	17.67±5.00
	PAES-SF			
First	2.36±1.60	2.99±2.18	87.23±17.33	24.20±4.24
Median	2.49±1.71	3.06±2.21	47.30±14.11	17.23±4.08
Last	2.66±1.88	3.28±2.51	28.80±11.10	12.97±3.69
	PAES-SFG			
First	1.61±0.34	2.24±1.11	104.53±27.86	25.77±4.59
Median	1.64±0.35	2.12±0.62	64.13±21.15	20.97±4.81
Last	1.79±0.62	2.31±1.11	45.10±15.76	17.33±4.19

Finally, Figures 6 and 7 show an example of complete Pareto fronts generated by the three algorithms in a single trial on the training and test sets, respectively. We have plotted only the solutions of the final archive with the MSEs lower than 5, since the other solutions in the archive are characterized by quite high MSEs and tend to visually flatten the plot. On the

ISBN: 978-989-95079-6-8

other hand, these solutions, though characterized by very low complexity, are not very interesting due to the high MSEs. We observe that the approximated Pareto fronts obtained by PAES-SFG are wide and quite dense. Further, except for some solutions in the low complexity zone, all the solutions in the Pareto front generated by PAES-SFG dominate the solutions at the same complexity of the other two fronts.

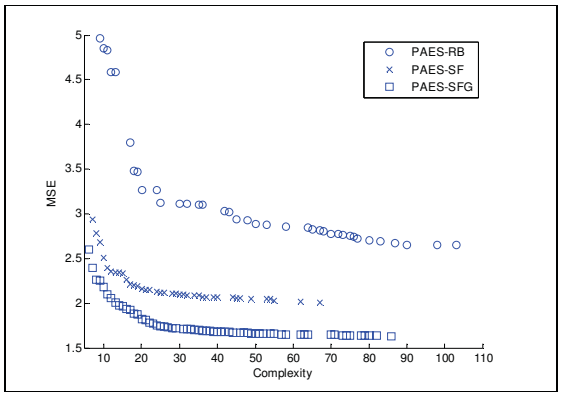


Figure 6. Examples of complete Pareto fronts on training set.

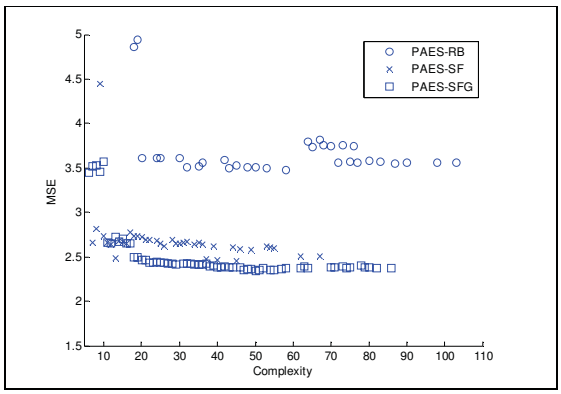


Figure 7. Examples of complete Pareto fronts on test set.

6 Conclusions

In this paper we have proposed a multi-objective evolutionary algorithm to generate a set of Mamdani rule-based fuzzy systems with optimal trade-offs between accuracy and complexity. We have exploited a modified version of the well known (2+2)PAES where the chromosome consists of three parts which codify the RB, and, for each variable, the number of fuzzy sets and the parameters of a piecewise linear transformation of the membership functions, respectively. To manage the dependence between granularity, and rule and membership parameters definition, we have adopted the following solution: the RB coded in the chromosome and the piecewise linear transformation are always defined on linguistic variables partitioned with a fixed maximum number of fuzzy sets. Only when accuracy and complexity have to be evaluated, the RB is actualized by using the real number of fuzzy sets determined by the second part of the chromosome and the DB by computing the corresponding piecewise linear transformation of the fuzzy sets. This approach has proved to be very efficient and effective, allowing both a good exploitation of the solutions and an accurate exploration of the search space.

The algorithm has been tested on a real world regression problem and has provided Pareto fronts with solutions characterized by better trade-offs between accuracy and complexity than solutions belonging to Pareto fronts generated by learning only the RB with a fixed BD, and concurrently the RB and the MF parameters with fixed granularities, respectively.

References

[1] E. H. Mamdani and S. Assilian, An experiment in linguistic synthesis with a fuzzy logic controller, International Journal on Man-Machine Studies, 7: 1-13, 1975.

[2] J. Casillas et al. (eds.), Interpretability issues in fuzzy modeling, Springer, 2003.

[3] F. Herrera, Genetic fuzzy systems: taxonomy, current research trends and prospects, Evolutionary Intelligence, 1(1): 27-46, 2008.

[4] H. Ishibuchi, Multiobjective genetic fuzzy systems: review and future research direction, in Proc. of FUZZ-IEEE 2007 Int. Conf. on Fuzzy Systems, London, 2007.

[5] M. Cococcioni et al., A Pareto-based multi-objective evolutionary approach to the identification of Mamdani fuzzy systems, Soft Computing, 11(11): 1013-1031, 2007.

[6] A. Botta et al., Context adaptation of fuzzy systems through a multi-objective evolutionary approach based on a novel interpretability index, Soft Computing, 13(5): 437-449, 2009.

[7] R. Alcalá et al., A multi-objective genetic algorithm for tuning and rule selection to obtain accurate and compact linguistic fuzzy rule-based systems, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 15(5): 539-557, 2007.

[8] M. J. Gacto, R. Alcalá, and F. Herrera, Adaptation and application of multi-objective evolutionary algorithms for rule reduction and parameter tuning of fuzzy rule-based systems, Soft Computing, 13(5): 419-436, 2009.

[9] P. Ducange et al., Knowledge Base Learning of Linguistic Fuzzy Rule-Based Systems in a Multi-objective Evolutionary Framework, Lecture Notes on Computer Science, Hybrid Intelligent Systems, pp. 747-754, 2008.

[10] P. Pulkkinen and H. Koivisto, Fuzzy classifier identification using decision tree and multiobjective evolutionary algorithms, International Journal of Approximate Reasoning, 48: 526-543, 2008.

[11] H. Ishibuchi and Y. Nojima, Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning, International Journal of Approximate Reasoning, 44(1): 4-31, 2007.

[12] M. Antonelli et al., A multi-objective genetic approach to concurrently learn partition granularity and rule bases of Mamdani fuzzy systems, in Proc. of the 8th International Conference on Hybrid Intelligent Systems (HIS-2008), pp. 278–283, Barcelona, Spain, 2008.

[13] O. Cordón, F. Herrera, and P. Villar, Analysis and guidelines to obtain a good uniform fuzzy partition granularity for fuzzy rule-based systems using simulated annealing, International Journal of Approximate Reasoning 25(3):187-215, 2000.

[14] F. Klawonn, Reducing the number of parameters of a fuzzy system using scaling functions. Soft Computing, 10(9): 749-756, 2006.

[15] W. Pedrycz and F. Gomide, Fuzzy Systems Engineering: Toward Human-Centric Computing, Wiley-IEEE Press, 2007.

[16] H. Ishibuchi et al., Selecting fuzzy if-then rules for classification problems using genetic algorithms, IEEE Transactions on Fuzzy Systems, 3(3): 260-270, 1995.

[17] H. Altay Guvenir and I. Uysal, Bilkent University Function Approximation Repository, http://funapp.cs.bilkent.edu.tr, 2000.

ISBN: 978-989-95079-6-8

Semantically-driven flexible division in fuzzy object oriented models

Nicolás Marín[1] Carlos Molina[2] Olga Pons[1] M. Amparo Vila[1]

1. Department of Computer Science and AI, University of Granada, 18071, Granada, Spain
2.Department of Informatics, University of Jaén, Campus Las Lagunillas, A3, 23071, Jaén, Spain
Email: nicm@decsai.ugr.es, carlosmo@ujaen.es, {opc,vila}@decsai.ugr.es

Abstract— *Fuzzy Databases have been a fertile research area that has produced a wide variety of remarkable solutions for the storage and manipulation of imperfect information. Proposals can be found in the three most relevant data models: relational, object-oriented, and object-relational. Query capabilities have been especially studied in the context of the relational data model, while in object models many problems continue being a matter of research. In this paper, we focus on the resolution of "division queries" (in its relational sense) in an object oriented data model. As we will see, the presence of fuzzily described objects in the database make necessary to use suitable operators that take into account the resemblance that governs the comparison in the underlying reference universe. We also analyze the role of cardinality in this kind of queries.*

Keywords— Division, Quotient, Fuzzy, Object-Oriented, Object-Relational, Database.

1 Introduction

Fuzzy Set Theory[1] has proved to be an adequate tool in handling real world data when they are affected by imperfections of very different kinds. The application of this theory in order to extend conventional databases has led to the development of fuzzy database systems.

Object-oriented and object-relational database management systems allow the representation of schema when complex relationships make the use of Codd's relational model difficult. The object-oriented data model is more powerful from a modeling point of view, because it incorporates important features such as inheritance and encapsulation. As it was the case with the relational data model in the past, many researchers have recently tried to improve object orientation with the help of fuzzy concepts. As a result, fuzzy object-oriented database models (FOODBM) have appeared [2, 3, 4].

In [5, 6] a framework is presented which allows programmers to handle imprecision in the description of objects. An object-relational implementation of this model can be found in [7].

In this paper, we focus on the study of operators that can be used to solve a special kind of queries in fuzzy object-oriented databases: those queries that, in the relational context, are known as fuzzy divisions. Fuzzy division queries in the relational data model has been deeply studied during the last decades: from the initial works of Dubois et al.[8] till the last works of Bosc et al.[9], the interested reader can find a wide survey of remarkable proposals(e.g. [10, 11, 12]).

This work analyzes how to solve this kind of queries over a fuzzy object model, where the equality is substituted by resemblance as the basis of comparisons. Due to this fact, we have to suitably soften the division operator, so that resemblance can be appropriately taken into account.

The paper is organized as follows: section 2 briefly introduces the fuzzy object-oriented data model that is the basis of our proposal; section 3 describes the type of queries we want to solve in an object-oriented context; section 4 analyzes the problem of considering the resemblance when computing the inclusion associated to division queries; section 5 analyzes the role that the cardinality of the involved sets plays in the resolution of the query; finally, some conclusions are outlined in section 6.

2 A fuzzy object-oriented model

As we have previously mentioned in the introduction, this paper is devoted to the problem of how to solve division queries in a fuzzy object model like the one introduced in [13]. We briefly present in this section a summary of its main characteristics.

As in any object model, the state of an object is equated with a set of attribute values according to its class description. The approach described in [13] considers different types of attribute values in order to give support to the representation of fuzzily described objects. Together with the *precise* values, objects, and *crisp* collections of any conventional object-oriented data model, the fuzzy model permits to use more powerful values, as imprecise labels and fuzzy collections, in order to deal with imprecision in the state of the object.

Fig. 1 shows a description of different types of attribute values that we consider in our approach. That is, the object state can be composed by:

- Precise values: This category of values involves all the classical basic classes that usually appear in an object-oriented data model (e.g. numerical classes, string classes, etc.). Values in these domains are easily represented and compared using conventional built-in data types and the classical set of relational operators.
- Imprecise values: The case of imprecise (atomic) values is a bit more complex. In many cases, linguistic labels[14, 15, 16] are associated to this kind of values, but different types of imprecise values must be considered according to their semantics.
- Objects: the attribute value may be a reference to another object (constituting what is called a complex object).
- Collections: the attribute may be conformed by a set of values or, even, by a set of objects. Imprecision in this kind of attributes may appear at two levels:

ISBN: 978-989-95079-6-8

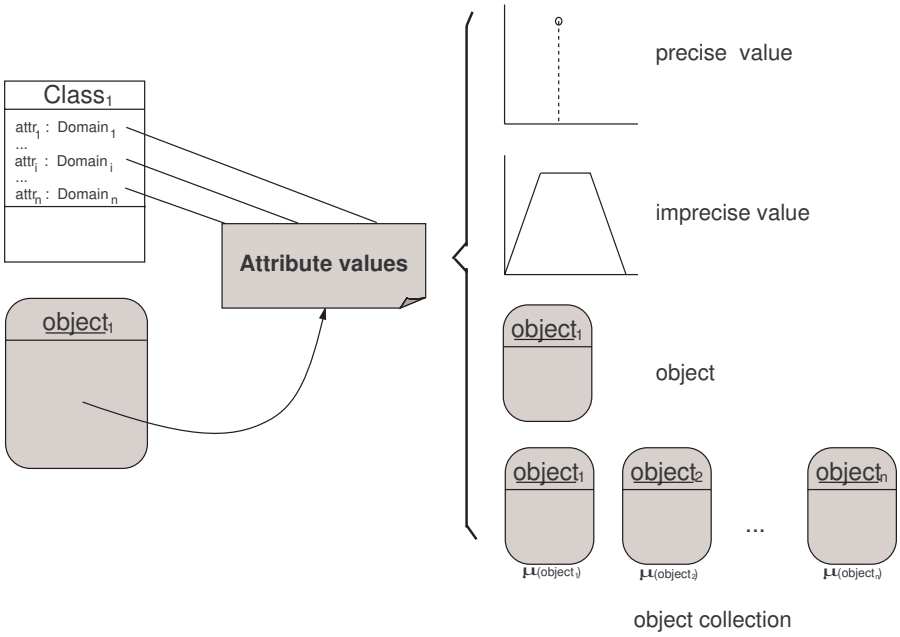


Figure 1: Different kinds of attribute values

– The set may be fuzzy. The semantics of membership degrees depends on the problem, but, in collections, the fuzzy set is considered to be conjunctive. For example, the set of languages a person can speak may be fuzzy if we take into account the degree of command for every language.
– The elements of the set may be fuzzy values or, in general, fuzzily described objects. For example, we can have a set of fuzzily described students.

An example of a fuzzy set of fuzzily described objects may be the collection of friends of a given person, if we use a degree to measure the friendship relation and the friends are described by means of fuzzy attribute values.

2.1 Generalization of equality

We do not only need a representation of fuzzily described objects in our model but also a way to manage this kind of objects. This includes a basic capability to compare the state of two objects belonging to a given class.

With fuzzily described objects like the ones presented in the previous paragraphs, this comparison has to be performed by a process which consists of two steps:

- computing degrees of resemblance for pairs of attribute values, and
- aggregating these resemblance degrees to obtain a general degree of object resemblance.

That is, if o_1 and o_2 are two objects belonging to class C which is characterized by type T_C whose structural component Str_C is an attribute set $\{x_1, x_2, \ldots, x_n\}$, then our goal is to find a resemblance degree between o_1 and o_2 by the aggregation of resemblance degrees of the pairs $(o_1.x_i, o_2.x_i)$.

$S_{x_i}(o_1, o_2)$ will stand for the resemblance degree observed between the x_i attribute values for objects o_1 and o_2, and $S(o_1, o_2)$ is the aggregated resemblance degree we want to calculate.

In this process, not all attributes need to have the same importance: each attribute x_i has an associated weight $p_{x_i} \in [0, 1]$ that represents its relative importance in the final decision.

Figure 2 summarizes the process of calculating resemblance degrees. In order to compare objects o_1 and o_2, we first compare their attribute values and obtain partial compatibility degrees $S_{x_i}(o_1, o_2)$. Then, we aggregate them to obtain a global resemblance opinion $S(o_1, o_2)$ according to the attribute importance established in the class.

As can be observed, the comparison of two objects involves a recursive procedure. For the sake of space, we omit here the operators for base cases and the analysis of the recursive aggregation procedure used in our model. Interested readers can found a detailed description in [17].

3 Division queries in fuzzy object-oriented databases

Codd defined a set of eight basic operators for his relational model. Some of them are implemented directly in SQL, but some other require particular implementation to fit into the SQL language. Relational division is one of the eight basic operations in Codd's relational algebra[18] that has these requirements.

Suppose that we have two relations t_a and t_b whose schema are, respectively, $T_A(X, Y)$ and $T_B(Y')$, where Y and Y' are compatible attributes. The division $t_a \div t_b$ obtains a new relation t_d with schema $T_D(X)$ and with the following set of tuples:

$$\{a | a \in domain(X) \wedge (\forall b, (b \in t_b) \rightarrow (< a, b > \in t_a)\} \quad (1)$$

Assume now that T_A and T_B are fuzzy relations, i.e., that their tuples are weighted by a number between 0 and 1. That

ISBN: 978-989-95079-6-8

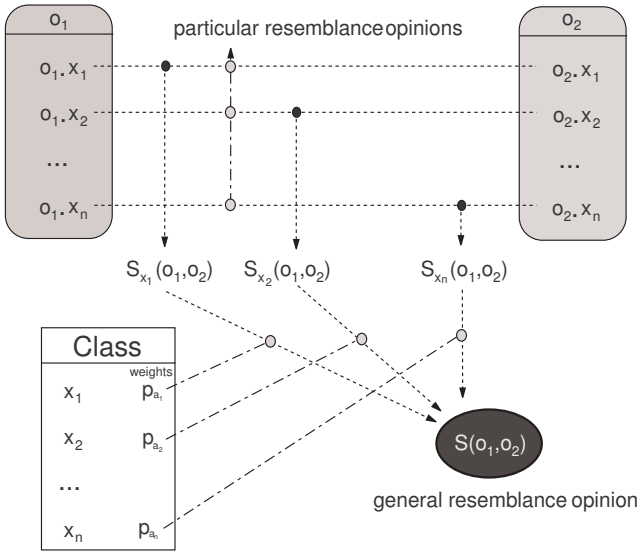


Figure 2: Obtaining the compatibility between two objects.

is, each tuple of t_a is affected by a membership degree, $\mu_{t_a}(< a, b >) \in [0, 1]$ and, similarly, each tuple of t_b has its corresponding $\mu_{t_b}(< b >) \in [0, 1]$.

The membership degree of the tuples of the division result can be obtained as follows [8]:

$$\mu_{t_d}(a) = inf_b\{\mu_{t_b}(b) \rightarrow \mu_{t_a}(a, b)\} \quad (2)$$

In the case of an object-oriented model, we have to consider (in the more general case) that attribute values are objects. Figure 3 depicts the tables with the new situation in an object-oriented context.

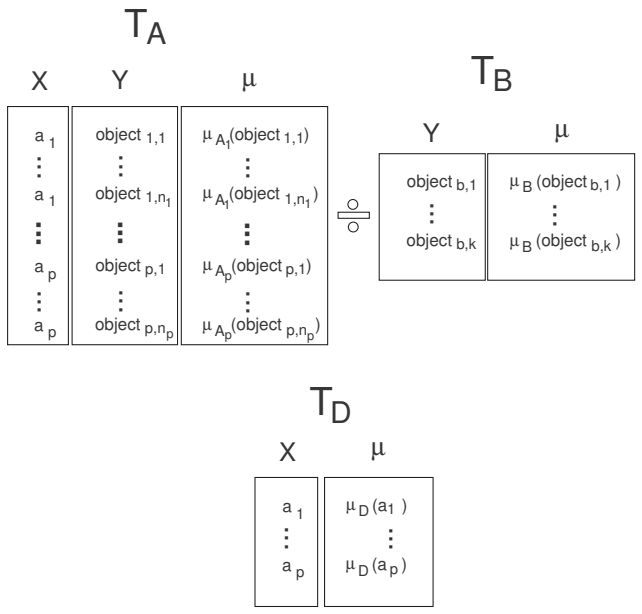


Figure 3: Tables involved in the division

Moreover, as in an object-oriented model attribute values do not necessarily have to be atomic, table T_A (and, similarly T_B) could have the schema showed in Fig. 4.

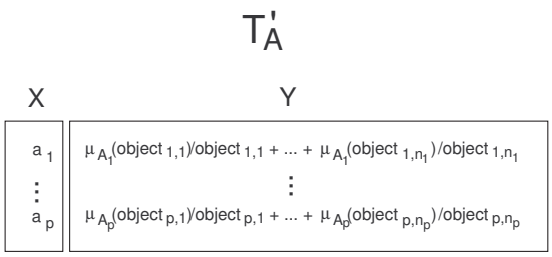


Figure 4: Object-oriented version of T_A

In any case, whatever the schema version of dividend and divisor tables are, the solution for division queries entails the resolution of the inclusion problem described in Fig. 5. Next sections are devoted to analyze this inclusion problem.

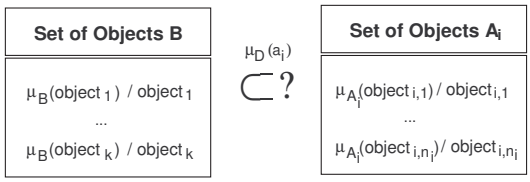


Figure 5: An inclusion problem

4 Two resemblance-based approaches

According to the previous paragraphs, in order to solve division queries in object-oriented models, we need to perform suitable computations of the above mentioned inclusion operator.

To compute the inclusion between any couple of fuzzy sets of objects (like in Fig. 5), we need to generalize the fuzzy inclusion operators, taking into account that the objects of the set may be fuzzily described.

Several proposals for the calculus of this inclusion degree can be found in the literature. In [19] the inclusion degree between two fuzzy sets A and B is calculated as follows:

$$N(B|A) = \min_{u \in U}\{I(\mu_A(u), \mu_B(u))\}, \quad (3)$$

where I stands for an implication operator, and μ_X is the membership function that describes the fuzzy set X. This degree coincides with the fuzzy division operator described in (2). The implication operator can be chosen in accordance to the properties we want the inclusion degree to fulfil [8, 10].

Nevertheless, independently of the chosen implication operator, this formulation supposes that both A and B are defined over a reference universe $\mathcal{U}$ made up of precise elements, where the classical equality is the basis of the comparisons. That is, the implication operator compares to what extent the presence of an element of the universe $\mathcal{U}$ in A forces its presence in B (i.e. we compare the membership degrees of the *same* object to both sets).

However, in the context of fuzzy databases, it frequently happens that the elements of the universe $\mathcal{U}$ are imprecise objects among which classical equality cannot be applied (as we

ISBN: 978-989-95079-6-8

commented in section 2. Instead, a similarity or a (more relaxed) resemblance relationship must be used. That is, for a given element in the set A, it is not clear which element of B has to be taken in order to compare the membership degrees.

In our example of object databases, the objects of the two set of Fig. 5 may be defined with imprecision and two *apparently* different objects (from the identity point of view) may be the *same* one (from the value equality point of view).

That is, if the reference universe $\mathcal{U}$ is formed by fuzzily described elements, a generalized version of (3) that takes into account resemblance among the involved objects must be used. We can consider two distinct ways of doing that, namely, making resemblance acts as a limit of the implication or making resemblance acts as a limit of membership to A_i.

4.1 Resemblance as a limit of the implication

If the reference universe $\mathcal{U}$ is formed by fuzzily described elements, (3) can be generalized as follows [20]:

Definition 1 (Resemblance driven inclusion degree) *Let A and B be two fuzzy sets defined over a finite reference universe $\mathcal{U}$, S be a resemblance relation defined over the elements of $\mathcal{U}$, and $\otimes$ be a t-norm. The inclusion degree of B in A driven by the resemblance relation S is calculated as follows:*

$$\Theta_S(A|B) = \min_{x\in\mathcal{U}} \max_{y\in\mathcal{U}} \theta_{B,A,S}(x,y) \tag{4}$$

where

$$\theta_{B,A,S}(x,y) = \otimes(I(\mu_B(x),\mu_A(y)),\mu_S(x,y)) \tag{5}$$

In (3), the inclusion degree is calculated taking into account the element of $\mathcal{U}$ that fulfils in a lower degree the implication condition between the membership degrees of this element to both sets. In (4), since the membership degrees of similar (but not necessarily equal from the identity point of view) elements can be compared, we restrict the implication using the resemblance degree of the two elements. In summary, for each element that belongs with a certain degree to the set A, we look for a *quite similar* object in $\mathcal{U}$ that belongs to the set B with an equal or higher degree.

Consider the example of Fig. 6.

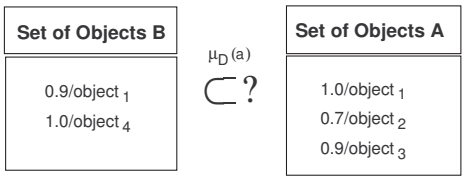


Figure 6: Example

In this example, $\{o_1, o_2, o_3, o_4\}$ is the reference universe $\mathcal{U}$ over which the resemblance relation S of table 1 is defined.

In this situation:

$\Theta_S(A|B)$= min {max {$\otimes$(I($\mu_B(o_1)$, $\mu_A(o_1)$), $\mu_S(o_1,o_1)$), ..., $\otimes$ (I($\mu_B(o_1)$, $\mu_A(o_4)$), $\mu_S(o_1,o_4)$)}, ..., max{$\otimes$(I($\mu_B(o_4)$, $\mu_A(o_1)$), $\mu_S(o_4,o_1)$), ..., $\otimes$(I($\mu_B(o_4)$, $\mu_A(o_4)$), $\mu_S(o_4,o_4)$)}}.

If we use the minimum as t-norm and (6) as implication operator, then:

$\Theta_S(A|B)$= min {max {1,0,0,0}, max {0,1,0,0}, max {0,0,1,0.7}, max {0,0,0.7,0}}= min{1,1,1,0.7}=0.7.

Table 1: Resemblance Relation

	o_1	o_2	o_3	o_4
o_1	1.0	0.0	0.0	0.0
o_2		1.0	0.0	0.0
o_3			1.0	0.7
o_4				1.0

$$I(x,y) = \begin{cases} 1, & \text{if } x \le y \\ y/x, & \text{otherwise} \end{cases} \tag{6}$$

4.2 Resemblance as a limit of relationship

As an alternative to the strategy used in the previous subsection, we can consider that the resemblance only restricts the membership degree of a certain object to the set A.

The assumption of this second version of flexible inclusion is to enlarge the set A into a superset A' obtained by composing A with the resemblance relation as suggested in [9]. The idea is to expand A in the sense that the objects of the reference class $\mathcal{U}$ *similar* to an object initially present in A, are added to A. This resemblance based tolerant inclusion can be defined as follows.

Definition 2 (Resemblance based tolerant inclusion degree) *Let A and B be two fuzzy sets defined over a finite reference universe $\mathcal{U}$, S be a resemblance relation defined over the elements of $\mathcal{U}$, and $\otimes$ be a t-norm. The tolerant inclusion degree of B in A according to the resemblance relation S is calculated as follows:*

$$\Psi_S(A|B) = \min_{x\in\mathcal{U}} \max_{y\in\mathcal{U}} \psi_{B,A,S}(x,y) \tag{7}$$

where

$$\psi_{B,A,S}(x,y) = I(\mu_B(x), \otimes(\mu_A(y),\mu_S(x,y)) \tag{8}$$

According to the example of Fig. 6 and the resemblance relation defined in Table 1:

$\Psi_S(A|B)$= min {max {I($\mu_B(o_1)$, $\otimes$($\mu_A(o_1)$,$\mu_S(o_1,o_1)$)), , ..., I($\mu_B(o_1)$, $\otimes$($\mu_A(o_4)$,$\mu_S(o_1,o_4)$))}, ..., max {I($\mu_B(o_4)$, $\otimes$($\mu_A(o_4)$,$\mu_S(o_4,o_1)$)), ...,I($\mu_B(o_4)$, $\otimes$($\mu_A(o_4)$,$\mu_S(o_4,o_4)$))}}.

If we use the minimum as t-norm and (6) as implication operator, then:

$\Psi_S(A|B)$= min {max {1,0,0,0}, max {1,1,1,1}, max {1,1,1,1}, max {0,0,0.7,0}}= min{1,1,1,0.7}=0.7.

4.3 Some notes on the two approaches

Although with our example, both flexible implementations of inclusion deliver the same value, normally the behaviour of these alternatives will differ.

An example where the second approach is more optimistic than the first one is described below.

Consider the naive example of Fig. 7.

In this example, $\{o_1, o_2\}$ is the reference universe $\mathcal{U}$ over which the resemblance relation S of table 2 is defined.

If we use the minimum as t-norm and (6) as implication operator, then:

$\Theta_S(A|B)$=0.6

$\Psi_S(A|B)$=1.0

ISBN: 978-989-95079-6-8

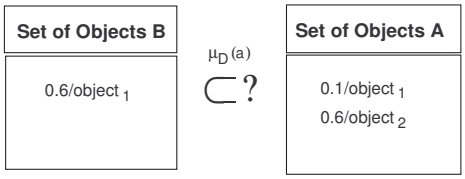


Figure 7: Example

Table 2: Resemblance Relation

	o_1	o_2
o_1	1.0	0.6
o_2		1.0

If we change the resemblance relation as in Table 3 and maintain the rest of data, then
$\Theta_S(A|B)$=0.9
$\Psi_S(A|B)$=1.0

Table 3: Resemblance Relation

	o_1	o_2
o_1	1.0	0.9
o_2		1.0

From a semantic point of view, with this configuration of the *Tolerant Inclusion*, in the computation of $\psi_{B,A,S}(x, y)$, the membership degree to A of a given object y bounds the needed resemblance between x and y. This effect can diminish if we use t-norms more restrictive than the minimum.

5 The role of cardinality

In order to finish our analysis of inclusion operators between fuzzy sets of fuzzily described objects, we now pay attention to cardinality, which may play a relevant role when computing the inclusion for two main reasons: On the first hand, if we soften the inclusion so that *similar* objects can fulfil the implication constraint, we have to be careful with the number of involved objects (the same object of A can be *matched* to more than one object of B); on the other hand, another point where the inclusion operators are able to be softened is to be tolerant with the number of objects of set B that we force to be included in A.

These two reasons lead us to new approaches to the inclusion operator which is supporting the division we want to solve.

5.1 Cardinality as a constraint of similarity

The use of resemblance relations instead of equality when computing inclusion degrees may make the obtained value be high, even if there is a great difference in terms of cardinality between the set of elements of B and the subset of elements of A matched to them during the computation of inclusion.

In some situations, cardinality may not be important. For example, imagine that A and B are two sets of *tools* and that we want to solve the question *Can a person do with tools of A all the tasks he can do with the tools of B?*. In this case, for each tool of B, we look for a tool of A with similar capabilities. In this context, it does not matter if the same tool of A is matched to more than one tool of B. That is, the number of *selected tools of A* is not relevant. However, if we need to assign this set of tools to a group of people that have to work independently, then, the number of *selected tools of A* is relevant and cardinality is important.

If we wish to distinguish between these situations, we need to weight the inclusion degree with a factor that takes into account the distance between the cardinalities of the fuzzy sets that are being compared.

Definition 3 (Cardinality Factor) *Let $\mathcal{U}$ be a reference universe. Let $|X|$ stands for the crisp cardinality of X. $CF : P(\mathcal{U}) \times P(\mathcal{U}) \to [0, 1]$ is a cardinality factor if:*

1. *If $|X| = |Y|$, $CF(X, Y) = 1$*
2. *If $|X| >= |Y|$, $CF(X, C) >= CF(Y, C)$*
3. *If $|X| >= |Y|$, $CF(C, Y) >= CF(C, X)$*

A simple example of CF is the following one:

$$CF(X,Y) = \begin{cases} 1, \text{ if } |X| >= |Y| \\ |X|/|Y|, \text{ otherwise} \end{cases} \quad (9)$$

This CF can be composed with a *relative fuzzy quantifier*[21] to adjust the behavior to the user needs.

Using (9) in the example of Fig. 8 and the resemblance relation of Table 4, we have:
$\Theta_S(A|B)$=1.0
$\Psi_S(A|B)$=1.0
but CF(matched elements of A,B)=0.5.

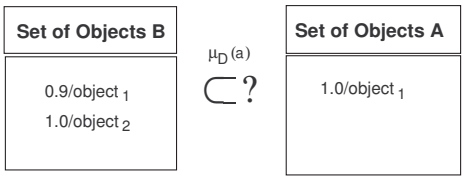


Figure 8: Example

Table 4: Resemblance Relation

	o_1	o_2
o_1	1.0	1.0
o_2		1.0

5.2 Cardinality as a relaxation of the inclusion

The other role cardinality can play in an inclusion operator is to soften the amount of elements of B that have to fulfil the inclusion relation. That is, with this kind of relaxation, the idea is to compute to what extent *almost all* elements of B are included in A. For example, in the case of tools, we want to solve the question *Can a person do with tools of A almost all the tasks he can do with the tools of B?*

There are many possible implementations of this idea in relation to what is called approximate division[8, 12]. One of the possible implementations of the corresponding approximate inclusion is to substitute the infimum in (3) by an OWA operator based on the desired *almost all* quantifier[22]. This way, we can adapt Θ and Ψ operators as follows.

ISBN: 978-989-95079-6-8

Definition 4 (Approximate Θ) *Let A and B be two fuzzy sets defined over a finite reference universe $\mathcal{U}$, S be a resemblance relation defined over the elements of $\mathcal{U}$, $\otimes$ be a t-norm, and $OWA_{\mathcal{Q}}$ an OWA operator induced by a regular non-decreasing quantifier $\mathcal{Q}$. The approximate inclusion degree of B in A driven by the resemblance relation S according to $OWA_{\mathcal{Q}}$ is calculated as follows:*

$$\Theta_{\mathcal{Q},S}(A|B) = OWA_{\mathcal{Q}}(\{\max_{y\in\mathcal{U}} \theta_{B,A,S}(x,y)\}_{x\in\mathcal{U}}) \qquad (10)$$

Definition 5 (Approximate Ψ) *Let A and B be two fuzzy sets defined over a finite reference universe $\mathcal{U}$, S be a resemblance relation defined over the elements of $\mathcal{U}$, $\otimes$ be a t-norm, and $\mathcal{Q}$ a relative fuzzy quantifier. The approximate tolerant inclusion degree of B in A based on the resemblance relation S according to $\mathcal{Q}$ is calculated as follows:*

$$\Psi_{\mathcal{Q},S}(A|B) = OWA_{\mathcal{Q}}(\{\max_{y\in\mathcal{U}} \psi_{B,A,S}(x,y)\}_{x\in\mathcal{U}}) \qquad (11)$$

6 Conclusions

In this paper, we have presented an approach to solve queries that are similar to the division queries of the relational data model but, now, in an object-oriented context. The fact that attribute values can be fuzzily described objects make us to extend the division operator by means of resemblance measures. We have considered two ways of incorporating such resemblance measures in the computation of inclusion, namely, restricting membership and restricting inclusion. Additionally, we have analyzed the role that cardinality can play in the division operator. Thus, we have considered to weight the inclusion operators with a cardinality factor that takes into account the number of elements of set A matched during the inclusion analysis. We have also propose two approximate inclusion operators based on the use of a quantifier for relaxing the division condition. With the described survey of operators, the user can semantically adapt the division operators to her/his needs in an object-oriented context.

Acknowledgment

Work partially supported by the projects *Representación y Manipulación de Objetos Imperfectos en Problemas de Integración de Datos: Una Aplicación a los Almacenes de Objetos de Aprendizaje*, Junta de Andalucía (TIC03175), and *Tratamiento de Información Difusa Temporal en un SGBD Relacional: Formulación Teórica, Implementación y Aplicaciones*, Estado Español (TIN2008-02066).

References

[1] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[2] Rita De Caluwe. *Fuzzy and Uncertain Object-Oriented Databases: Concepts and Models*, volume 13 of *Advances in Fuzzy Systems-Applications and Theory*. World Scientific, 1997.

[3] J. Lee, J-Y. Kuo, and N-L. Xue. A note on current approaches to extent fuzzy logic to object oriented modeling. *International Journal of Intelligent Systems*, 16:807–820, 2001.

[4] Zongmin Ma. *Advances In Fuzzy Object-oriented Databases: Modeling And Applications*. IGI Publishing, Hershey, PA, USA, 2004.

[5] N. Marín, O. Pons, and M. A. Vila. Fuzzy types: A new concept of type for managing vague structures. *International Journal of Intelligent Systems*, 15:1061–1085, 2000.

[6] F. Berzal, N. Marín, O. Pons, and M. A. Vila. Development of applications with fuzzy objects in modern programming platforms: Research articles. *Int. J. Intell. Syst.*, 20(11):1117–1136, 2005.

[7] L. Cuevas, N. Marín, O. Pons, and M. A. Vila. pg4db: A fuzzy object-relational system. *Fuzzy Sets Syst.*, 159(12):1500–1514, 2008.

[8] Didier Dubois and Henri Prade. Semantics of quotient operators in fuzzy relational databases. *Fuzzy Sets and Systems*, 78:89–93, 1996.

[9] P. Bosc, A. Hadjali, and O. Pivert. Exceptions and resemblance: Two keys for tolerant division operators. *Fuzzy Information Processing Society, 2008. NAFIPS 2008. Annual Meeting of the North American*, pages 1–6, May 2008.

[10] Patrick Bosc, Didier Dubois, Olivier Pivert, and Henri Prade. Flexible queries in relational databases—the example of the division operator. *Theor. Comput. Sci.*, 171(1-2):281–302, 1997.

[11] Xiaohui Tang and Guoqing Chen. A complete set of fuzzy relational algebraic operators in fuzzy relational databases. *Fuzzy Systems, 2004. Proceedings. 2004 IEEE International Conference on*, 1:565–569 vol.1, July 2004.

[12] P. Bosc, O. Pivert, and D. Rocacher. About quotient and division of crisp and fuzzy relations. *J. Intell. Inf. Syst.*, 29(2):185–210, 2007.

[13] F. Berzal, N. Marín, O. Pons, and M. A. Vila. Managing fuzziness on conventional object-oriented platforms: Research articles. *Int. J. Intell. Syst.*, 22(7):781–803, 2007.

[14] L. A. Zadeh. The concept of linguistic variable and its application to aproximate reasoning I. *Information Sciences*, 8:199–251, 1975.

[15] L. A. Zadeh. The concept of linguistic variable and its application to aproximate reasoning II. *Information Sciences*, 8:301–357, 1975.

[16] L. A. Zadeh. The concept of linguistic variable and its application to aproximate reasoning III. *Information Sciences*, 9:43–80, 1975.

[17] Fernando Berzal, Juan C. Cubero, Nicolás Marín, María Amparo Vila Miranda, Janusz Kacprzyk, and Slawomir Zadrozny. A general framework for computing with words in object-oriented programming. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 15(Supplement-1):111–131, 2007.

[18] E. F. Codd. A relational model of data for large shared data banks. *Commun. ACM*, 13(6):377–387, 1970.

[19] J-P. Rossazza, D. Dubois, and H. Prade. A hierarchical model of fuzzy classes. In *Fuzzy and Uncertain Object-Oriented Databases. Concepts and Models.*, volume 13, pages 21–61. Advances in Fuzzy Systems- Applications and Theory, 1998.

[20] N. Marín, J. M. Medina, O. Pons, D. Sánchez, and M. A. Vila. Complex object comparison in a fuzzy context. *Information and Software Technology*, 45(7):431 – 444, 2003. Software Engineering with Computational Intelligence S.I.

[21] Lotfi A. Zadeh. A computational approach to fuzzy quantifiers in natural languages. *Computers and Mathematics with Applications*, 9(1):149 – 184, 1983.

[22] R. R. Yager. Families of owa operators. *Fuzzy Sets and Systems*, (59):125–148, 1993.

ISBN: 978-989-95079-6-8

Rule Base and Inference System Cooperative Learning of Mamdani Fuzzy Systems with Multiobjective Genetic Algorithms

Antonio A. Márquez Francisco A. Márquez Antonio Peregrín

Information Technologies Department, University of Huelva
Huelva, Spain
Email: {amarquez, alfredo.marquez, peregrin}@dti.uhu.es

Abstract—*In this paper, we present an evolutionary multiobjective learning model achieving positive synergy between the Inference System and the Rule Base in order to obtain simpler, more compact and still accurate linguistic fuzzy models by learning fuzzy inference operators together with Rule Base. The Multiobjective Evolutionary Algorithm proposed generates a set of Fuzzy Rule Based Systems with different trade-offs between interpretability and accuracy in linguistic fuzzy modeling, allowing the designers select the one that involves the most adequate equilibrium for the desired application.*

Keywords— Linguistic fuzzy modeling, interpretability-accuracy trade-off, Multiobjective genetic algorithms, adaptive Inference System, adaptive defuzzification, rule learning.

1 Introduction

The main objective in system modelling is to develop reliable and understandable models. Interpretability and accuracy are usually contradictory requirements in the design of linguistic fuzzy models (FMs). Recent research into genetic fuzzy systems has focused on methods aimed at generating Fuzzy Rule-Based Systems (FRBS) with an appropriate trade-off between accuracy and interpretability [1, 2].

Two important tasks in the design of a linguistic FM for a particular application are the derivation of the linguistic Rule Base (RB) and the setup of the Inference System and defuzzification method. In the framework of the trade-off between interpretability and accuracy in fuzzy modeling, adaptive Inference Systems and defuzzification methods have acquired greater importance [3, 4].

Recently, the use of Multiobjective Evolutionary Algorithms (MOEA) has been applied to improve the aforementioned trade-off between interpretability and accuracy of linguistic fuzzy systems [5, 6, 7, 8, 9, 10, 11]. Some of them obtain the complete Pareto (the set of non-dominated solutions with different trade-offs) by selecting or learning the set of rules which best represents the example data, i.e., improving the system accuracy and decreasing the FRBS complexity. In [7, 8, 11] the authors also propose tuning the membership functions together with the rule selection to obtain simpler yet still accurate linguistic FMs.

Following these ideas on the advantage of using parametric operators and MOEAs to improve the trade-off between interpretability and accuracy, in [9] we presented a MOEA capable of learning the fuzzy inference operators (including inference and defuzzification) and of performing rule selection for Mamdani linguistic fuzzy systems. The proposed model aimed to achieve a positive synergy, that is, cooperation between the fuzzy operators and the RB to improve accuracy while at the same time simplifying the RB to improve interpretability.

Our main objective in this work is to include a new and highly important element in the learning process, the complete RB. Thus, we propose a MOEA capable of generating a set of FRBS (each with a high degree of cooperation between the RB and the Inference System) with varying optimal trade-offs between accuracy and complexity, so as to obtain compact and accurate linguistic fuzzy models by learning fuzzy operators and RB.

To do this, Section 2 describes the parametric fuzzy operators, Section 3 shows the RB learning used in this work, Section 4 is devoted to describing the MOEA learning proposal, Section 5 develops an experimental study, and finally, Section 6 presents some concluding remarks.

2 Adaptive Fuzzy Operators

In this section we describe the adaptive Inference System as well as the adaptive defuzzification method used in our learning proposal.

2.1 Adaptive Inference System

Linguistic FRBSs for system modeling use IF - THEN rules of the following form:

$$R_i: \text{If } X_{i1} \text{ is } A_{i1} \text{ and } \ldots \text{ and } X_{im} \text{ is } A_{im} \text{ then } Y \text{ is } B_i$$

with i = 1 to N, where N stands for the number of rules of the RB, X_{i1} to X_{im} and Y for the input and output variables respectively, and A_{i1} to A_{im} and B_i for the involved antecedents and consequent labels, respectively.

The expression of the Compositional Rule of Inference in fuzzy modeling with punctual fuzzification is the following: $\mu_{B'}(y) = I(C(\mu_{A1}(x_1), \ldots, \mu_{Am}(x_m)), \mu_B(y))$, where $\mu_{B'}(\cdot)$ is the membership function of the inferred consequent, $I(\cdot)$ is the implication operator, $C(\cdot)$ is the conjunction operator, $\mu_{Ai}(x_i)$ are the values of the matching degree of each input of the system with the membership functions of the rule antecedents, and $\mu_B(\cdot)$ is the consequent of the rule.

ISBN: 978-989-95079-6-8

The two components, the conjunction (C(·)) and the implication operator (I(·)) are suitable for parametrization in order for the Inference System to be adapted. Our previous studies in [3] show that models based on the adaptive conjunction is a more valuable option than those based on the adaptive implication operator. Hence, we selected the adaptive conjunction in this study in order to insert parameters in the Inference System.

Taking into account the aforementioned studies in [3], we have selected the Dubois adaptive t-norm with a separate connector for every rule, the expression for which is shown in (1).

$$T_{Dubois}(x, y, \alpha) = \frac{x \cdot y}{Max(x, y, \alpha)}, \quad (0 \leq \alpha \leq 1) \qquad (1)$$

This adaptive t-norm showed the highest accuracy in previous studies, compared with Frank and Dombi t-norms and is more efficiently computed. The use of an adaptive t-norm for the antecedent connection seeks better performance than traditional t-norms. Dubois t-norm performs between minimum ($\alpha = 0$) and algebraic product ($\alpha = 1$).

2.2 Adaptive Defuzzification Interface

There are various tendencies in the development of adaptive defuzzification methods reported in the literature. These employ one or more parameters in their expression for modifying the behaviour of the defuzzifier or, in most cases, to achieve higher accuracy.

Following the studies developed in [13], in this work we consider applying the defuzzification function to the fuzzy set inferred by each rule (getting a characteristic value) and computing them by a weighted average operator, because of its fine performance, efficiency and easier implementation. This way of working is named FITA (First Infer, Then Aggregate) [12].

We also consider the use of a product functional term of the matching degree between the input variables and the rule antecedent fuzzy sets (h_i), $f(h_i) = h_i \cdot \beta_i$ where β_i corresponds to one parameter for each rule R_i, i=1 to N, in the RB, as it is more efficiently computed and obtains similar results to other functions [13]. The adaptive defuzzification formula selected is shown in (2).

$$y_0 = \frac{\sum_{i}^{N} h_i \cdot \beta_i \cdot V_i}{\sum_{i}^{N} h_i \cdot \beta_i}, \qquad (2)$$

where V_i represents a characteristic value of the fuzzy set inferred from rule R_i, the Maximum Value or the Gravity Center (GC), the latter being the one selected in this paper.

The product functional term with a different parameter for each rule has the effect of weighted rules. This value associated with the rule indicates the importance of that rule for the inference process.

3 Rule Base Learning

The linguistic RB learning used in this work is based on the ad-hoc data driven methodology named COR [14]. This methodology manages a set of consequent label sets (one per rule). Instead of selecting the consequent with the best performance in each subspace as usual (Wang and Mendel [15]), the COR methodology considers the possibility of using another consequent, different from the best, which allows the FRBS to be more accurate thanks to having a RB with best cooperation. For this purpose, COR performs a combinatorial search among the candidate rules looking for the set of consequents which globally achieves the best accuracy.

COR consists of two stages:

1) *Construction of the search space*—This obtains a set of candidate consequents for each rule.

2) *Selection of the most cooperative fuzzy rule set*—This performs a combinatorial search among these sets seeking the combination of consequents with the best global accuracy.

In order to perform this combinatorial search, an *explicit enumeration* or an *approximate search technique* can be considered. In this work, we use a search technique because it is effective and quick.

4 Rule Base and Inference System Cooperative Learning with Multiobjective Algorithms

This Section describes the evolutionary multiobjective model proposed in this work. As was previously mentioned, our objective is to obtain a set of fuzzy systems with different trade-offs between accuracy and interpretability, using adaptive inference and defuzzification, and Rule Base learning (including rule selection). To do this, we exploit two specific MOEAs considering a threefold coding scheme (coding of rules and coding of the parameters of the Inference Systems and Defuzzification). We adopted two of the most representative second generation MOEAs, SPEA2 [16] and NSGA-II [17], as two general purpose MOEAs for performing the cooperative adaptation of the fuzzy operators and fuzzy rule learning.

4.1 SPEA2 and NSGA-II

The SPEA2 algorithm [16] (*Strength Pareto Evolutionary Algorithm for Multiobjective Optimization*) is one of the most well-known techniques for solving multiobjective problems. It is characterized by the following two aspects: a *fitness* assignment strategy, which takes into account both dominating and dominated solutions for each individual, and a density function, estimated by employing the nearest neighbourhood, which guides the search more efficiently.

NSGA-II algorithm [17] is another of the most well-known and frequently-used MOEAs for general multi-objective optimization in the literature. It is a parameterless

ISBN: 978-989-95079-6-8

approach with several interesting principles: a binary tournament selection based on fast non-dominated sorting, an elitist strategy and a crowding distance method to estimate the diversity of a solution.

4.2 Questions related to the MOEAs.

The evolutionary model uses a chromosome with threefold coding scheme (C_C+C_D+C_R) where:

- C_C encodes the α_i parameters of the conjunction connective. They are N real coded parameters (genes), one for each rule, R_i ,of the linguistic RB. Each gene can take any value in the interval [0, 1], that is, between the minimum and the algebraic product. This is represented by the C_C part of the chromosome shown in Figure 1.

- C_D encodes the β_i parameters of the defuzzification. They are N real coded parameters, one for each rule, of the linguistic RB. Each gene can take any value in the interval [0, 10]. This interval has been selected according to the study developed in [13]. It allows attenuation as well as enhancement of the matching degree. This is represented by the real part C_D shown in Figure 1.

- C_R encodes the learning Rule Base. It is an integer string of N genes, each one representing a candidate rule consequent of the initial RB. Furthermore, depending on whether a rule is selected or not, the value '-1' is assigned to the corresponding gene. This is represented by the integer part of the chromosome shown in Figure 1.

The initial population is randomly initialized in the fuzzy operators part with the exception of a single chromosome:

- C_C with the N genes is initiated to 0 in order to make Dubois t-norm equivalent to Minimum t-norm initially.

- C_D also with the N genes is initiated to 1 with the objective of beginning like the standard WCOA method.

The initial population in the fuzzy rule part, C_R, is initialized following these two exceptions:

- A single chromosome with the N rules obtained by the WM-method [15], that is, with all the genes initialized to correspondent consequent.

- Default chromosomes randomly initiated with all rules activated. In this case, in order to achieve solutions with a high accuracy we should not lose rules that could present a positive cooperation once their FM parameters have been evolved. The best way to do this is to start with solutions that select all the possible rules. This favors a progressive extraction of bad rules (those that do not improve with the tuning of parameters).

The crossover operator employed by the fuzzy operators part is BLX-0.5 [18] while the one used for the rule learning part is HUX [19].

Finally, four offspring are generated by combining the two from the C_R part with the two from the operators part (the two best replace their parents). The mutation operator changes a gene value at random in the C_R and operators part (one in each part) with probability 0.2.

In this work, to obtain an optimal set of FRBS with different trade-offs, the fitness, based on the interpretability (using the number of rules) and the accuracy (using the error measure), must be minimized.

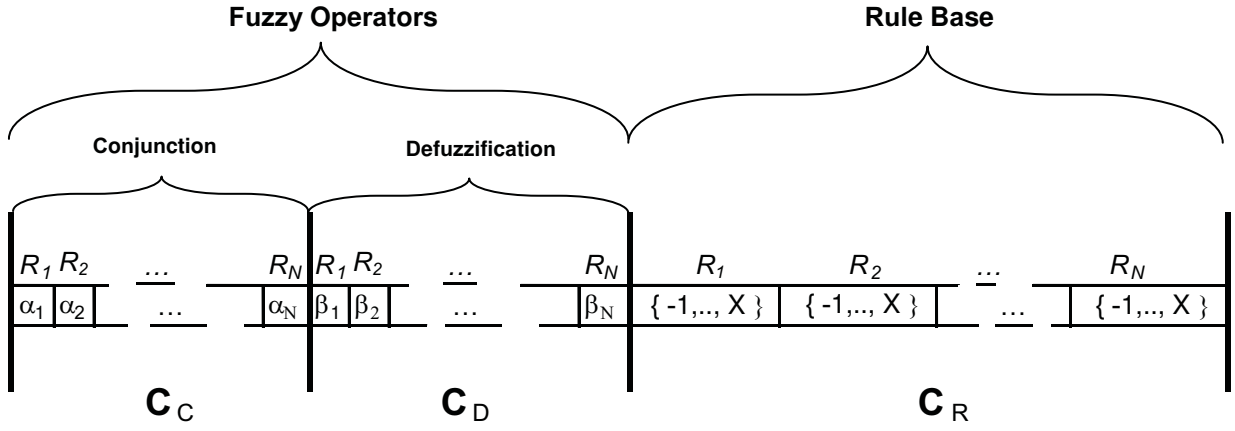


Fig. 1. Coding scheme for the MOEA with N rules

ISBN: 978-989-95079-6-8

5 Experimental Study

In order to analyse the practical behaviour of the proposed methods, we built several FMs in a real-world problem [20] of four input variables consisting in estimating the maintenance costs of medium voltage lines in a town. Methods considered for the experiments are briefly described in Table 1, where WM and COR methods are considered as reference. S denotes the methods that perform rule selection. If rule selection is performed after another method we give it a "+" denotation (eg. COR+S). However, if rule selection is performed jointly with another method we denote this in subscript (eg. COR_S). S–C–D means rule selection and fuzzy operators learning together. $SPEA2_{S\text{-}C\text{-}D}$, and $NSGA\text{-}II_{S\text{-}C\text{-}D}$ are the methods that learn the fuzzy operators and the rule selection together, while $SPEA2COR_{S\text{-}C\text{-}D}$ and $NSGA\text{-}IICOR_{S\text{-}C\text{-}D}$ are methods that also learn the RB as previously mentioned.

Table 1. Methods considered for comparisson

Ref.	Method	Description
[15]	WM	Wang & Mendel algorithm
[21]	WM + S	Wang & Mendel and then Rule Selection
[9]	WM + S–C–D	Wang & Mendel and then Rule Selection and Adaptive Fuzzy Operators
[14]	COR	COR
	COR + S	COR and then Rule Selection
	COR_S	COR with Rule Selection
	$COR_{S\text{–}C\text{–}D}$	COR with Rule Selection and Adaptive Fuzzy Operators
[9]	$SPEA2_{S\text{-}C\text{-}D}$	SPEA2 algorithm with Rule Selection and Adaptive Fuzzy Operators
-	$SPEA2COR_{S\text{-}C\text{-}D}$	SPEA2 algorithm with COR with Rule Selection and Adaptive Fuzzy Operators
[9]	$NSGA\text{-}II_{S\text{-}C\text{-}D}$	NSGA-II algorithm with Rule Selection and Adaptive Fuzzy Operators
-	$NSGA\text{-}IICOR_{S\text{-}C\text{-}D}$	NSGA-II algorithm with COR with Rule Selection and Adaptive Fuzzy Operators

5.1 Application Selected and Comparison Methodology

The application selected to test the evolutionary model is the aforementioned electrical distribution problem [20] that has a data set of 1059 cities with four input variables and a single output. The RB is composed of 65 linguistic rules achieved with the Wang and Mendel method [15]. The fuzzy partition used for inputs and output has 5 labels.

We considered a 5-fold cross-validation model, i.e., 5 random partitions of the data each with 20% (4 of them with 211 examples, and one of them with 212 examples), using the combination of 4 of them (80%) as training, and the remaining one as a test. We achieved a total of 30 trials for each evolutionary process, as the learning methods were run 6 times for each one of the data partitions. We show the average values of the medium square error (MSE) as a usual performance measure, computed considering the most accurate solution from each Pareto obtained with the multiobjective algorithm. This method of working was also employed in [9] in order to compare the single objective methods with the multiobjective ones based on considering the accuracy objective only, letting us see that the Pareto fronts are not only wide but also optimal, so similar solutions obtained with the WM + S–C–D or $COR_{S\text{–}C\text{–}D}$ must appear in the final Pareto. The MSE is computed with expression (3),

$$MSE\ (FM)_B = \frac{\frac{1}{2}\sum_{k=1}^{P}(y_k - FM(x_k))^2}{P} \qquad (3)$$

where *FM* denotes the fuzzy model the Inference System of which uses the Dubois t-norm as conjunction operator showed in expression (1), the inference operator is minimum t-norm, and the adaptive defuzzificación method is the one shown in expression (2). This measure uses a set of system evaluation data formed by P pairs of numerical data $Z_k = (x_k, y_k)$, k=1,..,P, with x_k being the values of the input variables, and y_k being the corresponding values of the associated output variables. The MOEAs population size was fixed at 200. The external population size of the $SPEA2_{S\text{-}C\text{-}D}$ and $SPEA2COR_{S\text{-}C\text{-}D}$ was 61.

5.2 Results and Analysis

To compare the results obtained we also used non-parametric tests, according to the recommendations made in [22]. The results obtained are shown in Table 2, where #R is the average number of rules, MSE_{tra} and MSE_{tst} are the average MSE for training and test respectively, and *Wilcoxon-test* is the result of applying a *Wilcoxon* signed-ranks test [23] (with 95% confidence), with the following interpretation: * represents the best average result (control algorithm); + means that the best result has better performance than that of the corresponding row, while sign (=) means it is similar to the best result. As we have mentioned, Demšar [20] recommends a set of simple, safe and robust non-parametric tests for statistical comparisons of algorithms, one of which is the Wilcoxon signed-ranks test [23]. This is analogous to the paired t-test in non-parametrical statistical procedures.

Table 2 only shows the best result for each MOEAs for accuracy. Analysing the results we can highlight the two following points:

- The learning of the RB allows remarkable improvement in accuracy: Looking at Table 2, we can observe that $SPEA2COR_{S\text{-}C\text{-}D}$ improves the accuracy of $SPEA2_{S\text{-}C\text{-}D}$ and $NSGA\text{-}IICOR_{S\text{-}C\text{-}D}$ improves $NSGA\text{-}II_{S\text{-}C\text{-}D}$. In spite of the fact that the number of rules shown in Table 2 is slightly larger, it must be taken into account that the results shown are the ones with the highest accuracy along the Pareto front. Looking at Table 3, we can observe similar values for accuracy in solutions with a lower number of rules from 40 to 32, so $NSGA\text{-}IICOR_{S\text{-}C\text{-}D}$ truly obtains better fuzzy systems (for accuracy and

ISBN: 978-989-95079-6-8

interpretability) than NSGA-II$_{S\text{-}C\text{-}D}$. Thus, the cooperation between the rules and the fuzzy operators improves the results of the evolutionary multiobjective proposal without rule learning.

- The solution with the best accuracy obtained with NSGA-II$_{COR_{S\text{-}C\text{-}D}}$ and SPEA2$_{COR_{S\text{-}C\text{-}D}}$ shows a similar accuracy to the single objective evolutionary model $COR_{S\text{–}C\text{–}D}$ (as shown in the non-parametric test) with a significant reduction in the number of rules, particularly for NSGA-II$_{COR_{S\text{-}C\text{-}D}}$. Consequently, the proposed method achieves more interpretable models with similar accuracy. We also notice that the best accuracy in Table 2 is obtained by the single objective model $COR_{S\text{–}C\text{–}D}$. The difference is small, but we can deduce that the evolutionary multiobjective methods are not achieving the most accurate solution. This fact suggested in [9] the design of more specific multiobjective algorithms in order to get even better solutions than the generic MOEAs SPEA2 and NSGA-II. Figure 2 shows the Pareto progress for each evolutionary algorithm (SPEA2$_{COR_{S\text{-}C\text{-}D}}$ and NSGA-II$_{COR_{S\text{-}C\text{-}D}}$) where we can observe the Pareto movement for each generation. Because to the adaptive fuzzy operators search space is large, we consider it may be necessary to focus the search process on the Pareto zone with highest accuracy, so that the same accuracy can be achieved as with single objective evolutionary algorithms based on accuracy.

Table 2. Results obtained

Method	#R	MSE$_{tra}$	Wilcoxon-test	MSE$_{test}$	Wilcoxon-test
WM	65	56135.75	+	56359.42	+
WM + S	40.9	41517.01	+	44064.67	+
WM + S–C–D	52.8	22640.95	+	26444.43	+
COR	65	50710.80	+	54584.76	+
COR + S	44.7	40763.48	+	43228.38	+
COR$_S$	43	39530.19	+	41060.99	+
COR$_{S\text{–}C\text{–}D}$	50	**20123.39**	*	**23323.72**	*
SPEA2$_{S\text{-}C\text{-}D}$	**38,60**	24021,41	+	29333,72	+
SPEA2$_{COR_{S\text{-}C\text{-}D}}$	41,10	21254,70	+	24079,32	=
NSGA-II$_{S\text{-}C\text{-}D}$	38,90	23364,63	+	28174,76	+
NSGA-II$_{COR_{S\text{-}C\text{-}D}}$	40,27	**20689,86**	=	**23346,34**	=

Table 3. A Pareto front example obtained from NSGA-II$_{COR_{S\text{-}C\text{-}D}}$

#R	MSE$_{tra}$	MSE$_{test}$	#R	MSE$_{tra}$	MSE$_{test}$	#R	MSE$_{tra}$	MSE$_{test}$
40	20036,37	22771,09	32	22354,40	24819,15	25	30133,70	42029,38
39	20198,77	22768,15	31	23266,76	25226,77	24	32481,97	43640,18
38	20388,66	22675,47	30	24347,59	27510,49	23	35058,65	43578,74
37	20783,97	23251,03	29	24994,05	26904,67	22	39128,61	53587,74
36	20999,33	23435,43	28	26190,94	29391,74	21	43276,59	58873,29
34	21719,71	24057,71	27	27445,85	30597,32	20	47428,26	69762,81
33	21901,55	24051,67	26	28706,22	33035,11	20	47428,26	69762,81

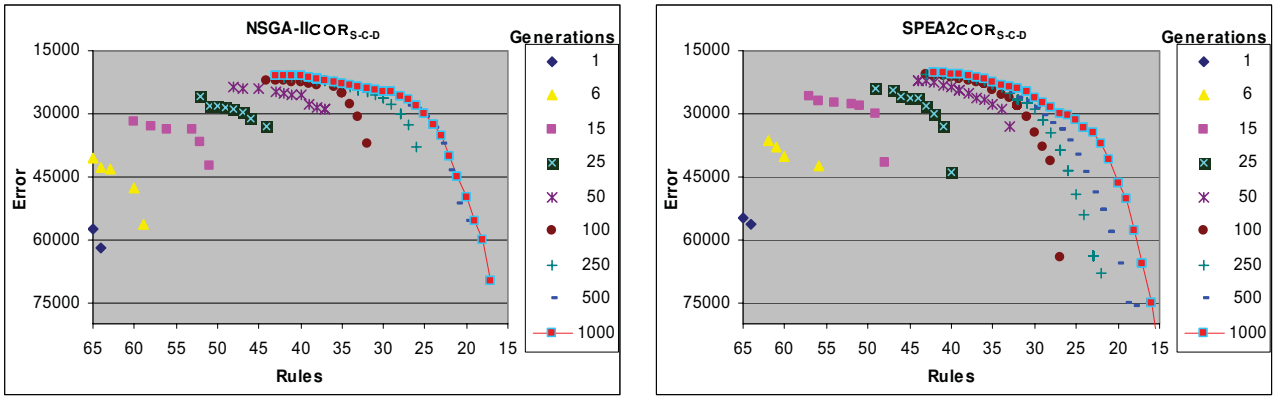


Fig. 2. Example of the Pareto front for NSGA-II$_{COR_{S\text{-}C\text{-}D}}$ and SPEA2$_{COR_{S\text{-}C\text{-}D}}$

.

ISBN: 978-989-95079-6-8

6 Conclusions

In the framework of the trade-off between accuracy and interpretability, the use of MOEAs gives a set of solutions with different levels of conciliation between both features. In this work we have proposed a multiobjective evolutionary learning model where the adaptive fuzzy operator parameters are learnt together with the RB. This fact allows both elements to cooperate, improving the accuracy as well as the interpretability.

The results obtained have shown that the use of MOEAs can represent a way to obtain a set of FRBSs in a single run with optimal trade-off between accuracy and interpretability. In terms of future work, some improvements may be developed in order to guide the search towards the desired Pareto zone with higher accuracy (right and central zone) where the FRBSs obtained are perhaps more interesting in more applications.

By focusing the search process we can reduce the effort of the search, and a better precision in the non-dominated solutions can be obtained, because the search effort is concentrated on a reduced zone of the Pareto, such that the density of the obtained solutions is higher. An improvement could be a change in the MOEAs used or a change in the non-dominated definitions in order to give more weight to the objective of accuracy.

Acknowledgment

Paper Supported by Projects TIN2008-06681-C06-06 and P07-TIC-03179.

References

[1] J. Casillas, O. Cordón, F. Herrera, et al. *Interpretability issues in fuzzy modeling*. Springer-Verlag , 2003.

[2] J. Casillas, O. Cordón, F. Herrera, et al. *Accuracy improvements in linguistic fuzzy modeling*. Springer-Verlag , 2003.

[3] J. Alcala-Fdez, F. Herrera, F.A. Márquez, et al. Increasing Fuzzy Rules Cooperation Based On Evolutionary Adaptive Inference Systems. *Int. J. of Intelligent Systems* 22(9): 1035-1064, 2007.

[4] F.A. Márquez, A. Peregrín, F. Herrera. Cooperative Evolutionary Learning of Fuzzy Rules and Parametric Aggregation Connectors for Mamdani Linguistic Fuzzy Systems. *IEEE Trans. on Fuzzy Syst*. 15(6): 1162-1178, 2007.

[5] H. Ishibuchi, T. Yamamoto. Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining. *Fuzzy Sets and Syst*. 141: 59-88, 2004.

[6] K. Narukawata, Y. Nojima, H. Ishibuchi. Modification of evolutionary Multiobjective optimization algorithms for Multiobjective design of fuzzy rule-based classification systems. *Proc. 2005 IEEE Int. Conf. on Fuzzy Syst*. 809-814. Reno , 2005

[7] R. Alcalá, J. Alcala-Fdez, M.J. Gacto, et al. A Multi-Objective Evolutionary Algorithm for Rule Selection and Tuning on Fuzzy Rule-Based Systems. *Proc. of the 16th IEEE Int. Conf. on Fuzzy Syst. FUZZ-IEEE'07*, 1367-1372, London , 2007.

[8] R. Alcalá, M.J. Gacto, F. Herrera, et al. A Multi-Objective Genetic Algorithm for tuning and Rule Selection to obtain Accurate and Compact Linguistic Fuzzy Rule-Based Systems. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Syst.,* 15(5): 539-557, 2007.

[9] A.A. Márquez, F.A. Márquez, A. Peregrín. Cooperation between the Inference System and the Rule Base by using Multiobjective Genetic Algorithms. *Proceedings of Third Internacional of Hybrid Artificial Intelligence System,(HAIS'08)* 739-746 Burgos, Spain, 2008.

[10] M. Cococcioni, P. Ducange, B. Lazzerini, et al. A Pareto-based Multi-objective Evolutionary approach to the identification of Mamdani Fuzzy Systems. *Soft Computing* 11(11): 1013-1031, 2007

[11] M.J. Gacto, R. Alcalá, F. Herrera. Adaptation and Application of Multi-Objective Evolutionary Algorithms for Rule Reduction and Parameter Tuning of Fuzzy Rule-Based Systems. *Soft Computing* 13(5): 419-436, 2009.

[12] J.J. Buckley, Y. Hayashi. Can approximate reasoning be consistent?. *Fuzzy Sets and Syst.,* 65(1): 13-18, 1994.

[13] O. Cordón, F. Herrera, F.A. Márquez, et al. A Study on the Evolutionary Adaptive Defuzzification Methods in Fuzzy Modelling. *Int. J.of Hybrid Intelligent Syst*, 1(1): 36-48, 2004.

[14] J. Casillas, O. Cordón, F. Herrera. COR: A methodology to improve ad hoc data-driven linguistic rule learning methods by inducing cooperation among rules. *IEEE Trans. Syst., Man, Cybern.—Part B: Cybern*. 32(4): 526-537, 2002.

[15] L.X. Wang, J.M. Mendel. Generating fuzzy rules by learning from examples. *IEEE Trans. on Syst., Man, and Cybernetics* 22(6): 1414-1427, 1992.

[16] E. Zitzler, M. Laumanns, L. Thiele. SPEA2: Improving the strength pareto evolutionary algorithm for multiobjective optimization. *Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems (EUROGEN'01*), 95-100, 2001.

[17] K. Deb, A. Pratap, S. Agarwal, et al. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. on Evolutionary Computation* 6(2): 182-197 , 2002.

[18] L.J. Eshelman, J.D. Schaffer. Real-coded genetic algorithms and interval-schemata. *Found. of Genetic Algorithms* 2: 187-202, 1993.

[19] L.J. Eshelman. The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination. *Found. of Genetic Algorithms* 1: 265-283 , 1991.

[20] O. Cordón, F. Herrera, L. Sánchez. Solving electrical distribution problems using hybrid evolutionary data analysis techniques. *Appl. Intell.,* 10, 5-24,1999.

[21] J. Casillas, O. Cordon, M.J. del Jesus, et al. Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction. *IEEE Trans. on Fuzzy Syst*. 13(1):13–29, 2005.

[22] J. Demšar, Statistical comparisons of classifiers over multiple data sets, *J. Mach. Learn. Res*. 7: 1-30, 2006.

[23] D.J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, CRC Press, Boca Raton, FL, 2003.

ISBN: 978-989-95079-6-8

On the information provided by uncertainty measures in the classification of remote sensing images

Luísa M S Gonçalves [1,2] *Cidália C Fonte* [2,3] *Eduardo N B S Júlio* [4] *Mario Caetano* [5,6]

1 Polytechnic Institute of Leiria, Civil Engineering Department, Portugal
2 Institute for Systems and Computers Engineering at Coimbra, Portugal
3 Department of Mathematics, University of Coimbra, Portugal
4 ISISE, Civil Engineering Department, University of Coimbra, Portugal
5 Portuguese Geographic Institute (IGP), Remote Sensing Unit (RSU), Lisboa, Portugal
6 CEGI, Instituto Superior de Estatística e Gestão de Informação, ISEGI, Universidade Nova de Lisboa, Portugal
Email: luisag@estg.ipleiria.pt, cfonte@mat.uc.pt, ejulio@dec.uc.pt, mario.caetano@igeo.pt

Abstract—*This paper investigates the potential information provided to the user by the uncertainty measures applied to the possibility distributions associated with the spatial units of an IKONOS satellite image, generated by two fuzzy classifiers, based, respectively, on the Nearest Neighbour Classifier and the Minimum Distance to Means Classifier. The deviation of the geographic unit characteristics from the prototype of the class to which the geographic unit is assigned is evaluated with the Un non-specificity uncertainty measures proposed by [1] and the exaggeration uncertainty measure proposed by [2]. The classifications were evaluated using accuracy and uncertainty indexes to determine their compatibility. Both classifications generated medium to high levels of uncertainty for almost all classes, and the global accuracy indexes computed were 70% for the Nearest Neighbour Classifier and 53% for the Minimum Distance to Means Classifier. The results show that similar conclusions can be obtained with accuracy and uncertainty indexes and the latter, along with the analysis of the possibility distributions, may be used as indicators of the classification performance and may therefore be very useful tools. Since the uncertainty indexes may be computed to all spatial units, the spatial distribution of the uncertainty was also analysed. It's visualization shows that regions where less reliability is expected present a great amount of detail that may be potentially useful to the user.*

Keywords—Accuracy assessment, Minimum Distance to Mean Classifier, Nearest Neighbour Classifier, Non-specificity measures, Remote Sensing Images, Uncertainty.

1 Introduction

Some classifiers used in Remote Sensing allow the assignment of each spatial unit (pixel or object) to several classes through the computation of degrees of possibility, probability or membership associated with each class, as opposed to the traditional classifiers where each spatial unit is only assigned to one class [3, 4, 5, 6]. This additional information may be interpreted as degrees of membership of the spatial unit to the classes, and in this case are usually referred to as soft classifiers [7]. Even though the use of soft classifiers is increasing, most applications still require a hard classification into disjoint classes of interest. For these cases, the spatial units may be assigned to the class presenting the larger degree of possibility or probability. The additional information provided by these degrees of possibility of probability may be used as indicators of the classifier difficulty to assign only one class to the spatial unit and, together with the application of uncertainty measures, may provide valuable information to the user [2, 8, 9, 10].

In [1] it is shown that, for a fuzzy classifier which assigns to each spatial unit degrees of membership or possibility associated with the classes, the non-specificity measures may be used to estimate the classification accuracy. They can therefore be a useful tool, since, on one hand, they can give an estimation of the classification accuracy prior to the final accuracy evaluation, enabling the classification improvement before the final classification accuracy is assessed (which is a time consuming and expensive process), and on the other hand enable the spatialization of the uncertainty [11]. In this paper a very high resolution image is classified with two fuzzy classifiers and the uncertainty measures (two non-specificity measures and an exaggeration measure) are applied to both classifications, to determine the usefulness of the information provided by them. A comparison between the information provided by the non-specificity measures and the producer's and user's accuracy indexes is made to determine their compatibility.

To evaluate the usefulness of the information provided by the spatialization of uncertainty, an approach similar to the one proposed by [2] is considered, where it is proposed that, when the spatial units are assigned to the best degree of similarity, errors of omission and commission are committed. The former occurs because, when assigning the spatial unit only to the best class, its similarity to the other classes is omitted. The latter occurs because, when the spatial unit is assigned to the best class and the degree of similarity to that class is smaller that one, a commission error is introduced, since a higher similarity is considered. To estimate omission errors [2] used an entropy measure. Since the Shannon entropy should only be applied to probability measures, the U_n uncertainty measure was used instead. To estimate the commission errors the measure of exaggeration proposed in [2] was used. The results are shown and conclusions drawn.

2 Data and methods

2.1 Data

The study was conducted in a rural area with a smooth topographic relief, situated in a transition zone between the

ISBN: 978-989-95079-6-8

centre and south of Portugal featuring diverse landscapes representing Mediterranean environments. The area is occupied mainly by agriculture, pastures, forest and agro-forestry areas where the dominant forest species in the region are eucalyptus, coniferous and cork trees. An image obtained by the IKONOS sensor was used, with a spatial resolution of respectively 1m in the panchromatic mode and 4m in the multi-spectral mode (XS) and a dimension of 11 884 m by 14 432m. The geometric correction of the multi-spectral image consisted of its orthorectification. The average quadratic error obtained for the geometric correction was 1.39 m, inferior to half the pixel size, which guarantees an accurate geo-referencing.

2.2 Classification

Two fuzzy classifiers were used in this application to obtain the elementary entities that are the basic units of landscape, like crown trees and parts of buildings, called Surface Elements (SE), to produce a Surface Elements Map (SEM). Both classifiers compute degrees of similarity [12] between the values observed at each spatial unit and a set of ideal values, which are, in this case, the spectral response observed in each spatial unit and what is considered to be the ideal characteristics of the class in terms of spectral values, respectively. Even though both classifiers use the same theoretical tools to derive the degrees of similarity, the definition of the ideal characteristics of each class is different for each of them.

The first classification method used is an object–oriented supervised Fuzzy classifier based on the Nearest Neighbor Classifier (FNNC) available on the software eCognition. In the supervised classification methods, a training set has to be chosen for each class. The spectral responses of the pixels inside this set are the signature files used to characterize each class. This classification method assigns to each pixel a degree of membership to each class, depending on the distance between its spectral response and the closest spectral response of the pixels used in the training set for each class (see Fig. 1).

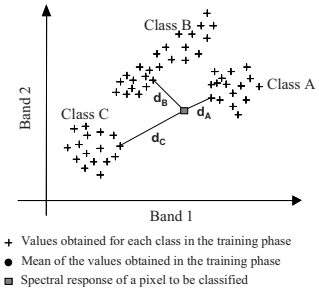


Figure 1: Nearest Neighbor Classifier.

The degrees of membership are computed using a membership function, where the ideal value, corresponding to a degree of membership equal to one, is obtained when the distance between the spectral response of the pixel to classify to the nearest reflectance value of the training set of the class is equal to zero (see Fig. 2).

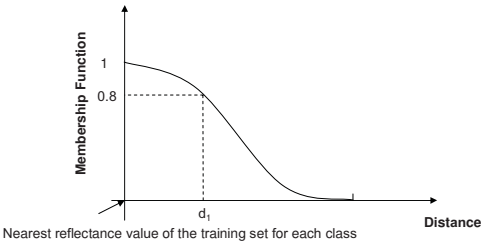


Figure 2: Membership function used in the fuzzy classifier based on the Nearest Neighbor Classifier.

The second classification method used is a pixel-based supervised Fuzzy classifier based on the underlying logic of Minimum-Distance-to-Means Classifier (FMDMC), available in the commercial software IDRISI. With this classification method, the image is classified based on the information contained in the signature files and a standard deviation unit (Z-score distance) introduced by the user. The fuzzy set membership is calculated based on a standardized Euclidean distance from each pixel reflectance, on each band, to the mean reflectance for each class signature (see Fig. 3), using a sigmoidal membership function (see Fig. 4).

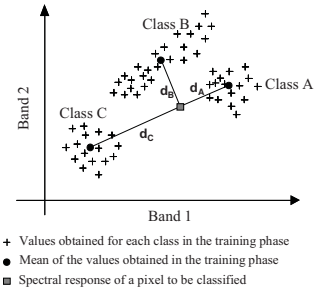


Figure 3: Minimum-Distance-to-Means Classifier.

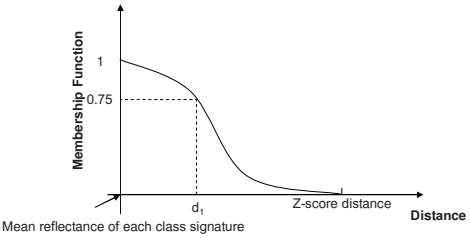


Figure 4: Membership function used in the fuzzy classifier based on the Minimum-Distance-to-Means Classifier

The underlying logic is that the mean of a signature represents the ideal point for the class, where fuzzy set membership is one. When distance increases, fuzzy set membership decreases, until it reaches the user-defined Z-score distance where fuzzy set membership decreases to zero. To determine the value to use for the standard deviation unit, the information of the training data set was used to study the spectral separability of the classes and determine their average separability measure.

The most representative surface elements of the study area are Eucalyptus Trees (ET), Coniferous Trees (CFT), Cork Trees (CKT), Shadows (S), Shallow Water (SW), Deep Water (DW), Herbaceous Vegetation (HV), Sparse Herbaceous Vegetation (SHV) and Non-Vegetated Area (NVA), and therefore these were the considered classes.

ISBN: 978-989-95079-6-8

2.3 Uncertainty measures

To evaluate the uncertainty of the possibilistic classifier the measures of nonspecificity were used. The uncertainty measure *NSp* proposed by [13] and given by

$$NSp(\Pi) = 1 - \sum_{i=1}^{n} \left[\Pi(x_i) - \Pi(x_{i+1}) \right] \frac{1}{i}$$

and a normalized version proposed by [1] of the U-uncertainty measure developed by [14], given by

$$U_n(\Pi) = \frac{[1 - \Pi(x_1)] \log_2 n + \sum_{i=2}^{n} [\Pi(x_i) - \Pi(x_{i+1})] \log_2 i}{\log_2 n}$$

In both equations Π is an ordered possibility distribution defined over a universal set X, that is, a possibility distribution such that $\Pi(x_1) \geq \Pi(x_2) \geq \ldots \geq \Pi(x_n)$, n is the number of elements of the universal set and $\Pi(x_{n+1})$ takes the value zero. A normalized version of the U-uncertainty measure was used to normalize the range of the measure, so that both non-specificity measures vary within the interval [0,1]. The non-specificity measures are appropriate to evaluate the uncertainty resulting from possibilistic classifications, since they quantify the ambiguity in specifying an exact solution [15, 16].

An additional uncertainty measure of exaggeration E, proposed by [2], was also used to quantify the commission errors at each spatial unit. This measure is given by

$$E = 1 - \Pi(x_1)$$

where $\Pi(x_1)$ is the largest degree of possibility.

2.4 Classification evaluation with the accuracy indexes

To evaluate the classifications accuracy a testing set was used. A random sampling of 900 pixels was selected considering the entire image scene. The sample unit was the pixel. The accuracy assessments was made with error matrixes, where the (i,j) entry is the number of pixels that is class i in the map and class j in the reference. The reference data were obtained from aerial images with larger resolution. The Global Accuracy was computed for both classifications. The User Accuracy (UA) and the Producer Accuracy (PA) indexes were also computed for all classes.

2.5 Comparison between uncertainty and accuracy indexes

To determine if the information given by the uncertainty measures may be used as indicators of the classifier's performance and if they are correlated with the results obtained with the classification quality assessment, two accuracy indexes were considered, namely the normalized user's and producer's accuracy; and one index for each uncertainty measure based on the mean, corresponding to the complement of the mean uncertainty per class ($I_{1-\overline{NSp}}$ and $I_{1-\overline{U_n}}$). Since high uncertainty values are expected to correspond to low accuracy values, and vice-versa, the complement of the uncertainty measures is considered to allow an easy comparison with the accuracy indexes. The correlation coefficient between the several accuracy and uncertainty indexes was also computed.

Even though uncertainty measures can be computed to the whole image, since the objective was to compare the results given by the uncertainty measures with the ones given by the error matrix, for this comparison only the sample pixels were used, so that the results were not influenced by the sample representativeness.

2.6 Spatial variation of uncertainty

Since there is a possibility distribution associated with all pixels or objects of the image, the uncertainty indexes may be computed for the whole image. This enables the visualization of the spatial distribution of uncertainty, and its spatial relation with the classes assigned to each spatial unit. As an estimator of the spatial units omission errors the U_n uncertainty measure was used, since it is more sensitive to dispersion than NSp (see [1]) and to estimate the commission errors the exaggeration E measure was used.

3 Results and discussion

3.1 Classification evaluation with accuracy indexes

The error matrix computed for the classification with the FNNC is presented in Fig. 5. The global accuracy obtained with this classification method was 70%.

	Error matrix of the classification with the fuzzy KNN									User's Accuracy (%)
	DW	SW	NVA	ET	S	HV	CKT	CFT	SHV	
DW	102									100.0
SW	3	92								96.8
NVA			76				2	1	8	87.4
ET				42	4	6	7	18	11	47.7
S	2				77		5			91.7
HV			2	1		90			7	90.0
CKT		1	11	8	18		50	6	26	41.7
CFT				3		43	10	42	9	39.3
SHV			23	3		22	16	3	73	52.1
Producer's Accuracy (%)	95.3	98.9	67.9	73.7	77.8	55.9	55.6	60.0	54.5	**69.77%**

Figure 5: Error matrix of the image classification with the method based on the Nearest Neighbour Classifier.

The error matrix shows that water classes (DW and SW) were well identified. Forestry species were often confused between each other and with other classes, such as Sparse Herbaceous Vegetation (SHV) and Herbaceous Vegetation (HV). Significant confusion was observed between Herbaceous Vegetation (HV) and Coniferous Trees (CFT). The class with the smaller value of PA is SHV (54.5%), which means it is the class with more omission error. And the class with smaller UA is CFT (39.3%) and therefore the class with more commission errors.

The global accuracy obtained with the classification with the FMDMC was 53% and the error matrix computed is presented in Fig. 6. The error matrix shows that Deep Water (DW) was the class best identified. Significant confusion was observed between Cork Trees (CKT), Eucalyptus Trees (ET) and Sparse Herbaceous Vegetation (SHV) and between ET and Coniferous Trees (CFT). The class with the smaller value of PA is SHV (14.8%), which means this is the class with more omission error. The pixels which are erroneously not included in this class are included mainly in the class Non Vegetated Areas (NVA), which is the class with smaller UA (27%) and therefore the class with more commission

ISBN: 978-989-95079-6-8

error, which receives pixels from all classes. These results show that this classifier presents great difficulty in classifying the class NVA.

	Error matrix of the classification with the fuzzy MDM									User's Accuracy (%)
	DW	SW	NVA	ET	S	HV	CKT	CFT	SHV	
DW	89				1					98.9
SW	3	54	1							93.1
NVA	16	40	104	8	19	79	15	15	89	27.0
ET				37		5	8	17	4	52.1
S	1				35		7			81.4
HV						65			1	98.5
CKT			7	13	3		50	11	18	49.0
CFT				5		16	6	27	3	47.4
SHV			3	1		2	6		20	62.5
Producer's Accuracy (%)	81.7	57.4	90.4	57.8	60.3	38.9	54.3	38.6	14.8	**53.21%**

Figure 6: Error matrix of the image classification with the method based on the Minimum Distance to Mean Classifier.

3.2 Comparison between uncertainty and accuracy indexes

For the classification made with the FNNC, the uncertainty indexes $I_{1-\overline{NSp}}$ and $I_{1-\overline{U_n}}$ along with the user's and producer's accuracy, ordered with increasing values of $I_{1-\overline{NSp}}$, are shown in Fig. 7.

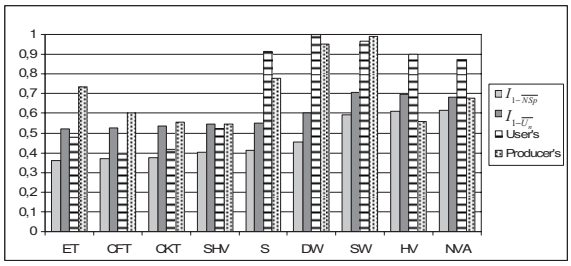


Figure 7: $I_{1-\overline{NSp}}$ and $I_{1-\overline{U_n}}$ uncertainty indexes along with the user's and producer's accuracy of the classification based on the Nearest Neighbour Classifier.

It can be observed that for all classes the uncertainty indexes present low to medium values. To determine the reason for these relatively low values of certainty, the observation of the possibility distribution was made, and it was observed that the largest degree of possibility is in general very high, but also high levels of possibility are obtained for the second and sometimes even for the third classes, which explains the high levels of uncertainty (see Table 1). The analysis of the possibility distributions and the classes associated with the second degrees of possibility for the several points enable the identification of classes that might be easily confused, which are mainly the forest classes and SHV. Similar conclusions can be taken from the confusion matrix, with the exception that considerable confusion was also expected between CFT and HV, and this is not expected from the analysis of the degrees of possibility.

From this analysis, worst results are expected for the classes ET, CFT, CKT and SHV, which are exactly the classes with low levels of UA, $I_{1-\overline{NSp}}$ and $I_{1-\overline{U_n}}$. This is an expected result, since in [1] it is shown that high correlation values are expected to occur between the uncertainty indexes and the UA. Similarly, the classes with higher values of $I_{1-\overline{NSp}}$ and $I_{1-\overline{U_n}}$ are in general the ones presenting higher UA, except for the classes S and DW, which have slightly lower value of $I_{1-\overline{NSp}}$ and $I_{1-\overline{U_n}}$.

Table 1: Mean values of the highest and second highest degrees of possibility of the possibility distributions associated with the sample points used to evaluate the accuracy of the classification with the FNNC, along with the most frequent second class.

Best class	$\overline{\Pi(x_1)}$	$\overline{\Pi(x_2)}$	Second class	% of pixels with the second class
SW	0.96	0.66	NVA	100
DW	0.99	0.87	NVA	51
NVA	0.84	0.37	SHV	52
ET	0.99	0.96	CKT	48
HV	0.93	0.55	SHV	50
SHV	0.94	0.83	CKT	52
CFT	0.98	0.94	ET	56
CKT	0.99	0.94	SHV	46
S	0.98	0.89	CKT	64

For the classification made with the FMDMC, the uncertainty indexes $I_{1-\overline{NSp}}$ and $I_{1-\overline{U_n}}$ along with the user's and producer's accuracy, ordered with increasing values of $I_{1-\overline{NSp}}$, are shown in Fig. 8. It can be seen that the values of certainty given by both NSp and U_n are low to medium for all classes, presenting $I_{1-\overline{U_n}}$ slightly larger values than $I_{1-\overline{NSp}}$.

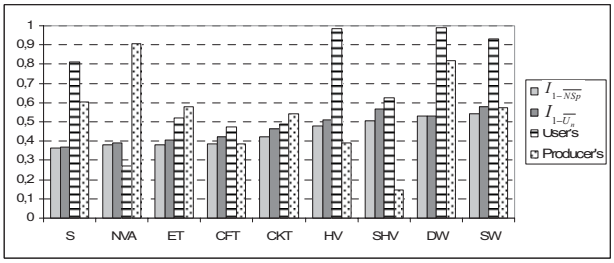


Figure 8: $I_{1-\overline{NSp}}$ and $I_{1-\overline{U_n}}$ uncertainty indexes along with the user's and producer's accuracy for the classification with the FMDMC.

To determine why such relatively high levels of uncertainty were obtained, an analysis of the possibility distributions was also made. This analysis showed that, for all classes, a high percentage of the pixels presented relatively low values for the highest degree of possibility, which explains the relatively low levels of certainty, and no point was classified with the highest degree of possibility equal to one (see Table 2). Further more, for all classes, the second highest degree of possibility for the great majority of testing sites was to the class Non Vegetated Areas (NVA) (except, of course, for the class NVA), in some cases with the degrees of possibility close to the higher degree, which means that there was great confusion between this class and all other classes.

ISBN: 978-989-95079-6-8

Table 2: Mean values of the highest and second highest degrees of possibility of the possibility distributions associated with the sample points used to evaluate the accuracy of the classification with the FMDMC, along with the most frequent second class.

Best class	$\overline{\Pi(x_1)}$	$\overline{\Pi(x_2)}$	Second class	% of pixels with the second class
SW	0.62	0.19	NVA	100
DW	0.53	0.003	NVA	26
NVA	0.41	0.07	HV	16
ET	0.48	0.14	NVA	88
HV	0.57	0.16	NVA	95
SHV	0.63	0.38	NVA	100
CFT	0.52	0.22	NVA	82
CKT	0.54	0.22	NVA	90
S	0.38	0.03	NVA	100

This indicates that one major problem of this classification was mainly due to the difficulty in discriminating NVA from the other classes. This same conclusion can be obtained from the accuracy analysis. The class NVA presents very different values for the user and producer accuracy, respectively 27% and 90%. This means that this class presents a very high percentage of commission errors, and therefore, a great amount of sites that should have been assigned to other classes were assigned to NVA, and were therefore absent from the other classes, such as SHV and HV, increasing their omission errors.

The information provided by the uncertainty measures is therefore in accordance with the one provided by the accuracy indexes, even though the analysis of the possibility distributions is necessary to reach the obtained conclusions.

The most evident discrepancies between the uncertainty and accuracy indexes obtained with the classification with the FMDMC are mainly for the class DW and S. These classes present higher values of accuracy, meaning that they were well identified according to the accuracy indexes, but considerably low levels of classification certainty, corresponding to lower values of the uncertainty indexes. This can be explained by the relatively low level of the first degree of possibility and the confusion between classes, which reinforces the assumption that uncertainty indexes capture the classifier difficulty in determining the correct class. Apparently some discrepancies occur between uncertainty and accuracy indexes for the SW and HV classes due to the fact that they present higher values of UA (93,1 and 98,5 respectively). However, the PA results are lower (57,4% and 38,9 % respectively) which reveal that these classes present higher omission errors and are in accordance with the lower values of $I_{1-\overline{U_n}}$ index. Two main aspects seem to be responsible for the results obtained with this classifier: 1) the larger degrees of possibility for each pixel are in general relatively low, which results in high uncertainty levels; 2) there is considerable confusion between all classes and the class NVA.

To explain the low degrees of possibility obtained with this classifier, a closer look to its classification approach is required. The computed degrees of possibility reflect the closeness of the spectral response of the testing set to an ideal value, which is the mean of the spectral responses obtained for the training set. The computed degrees of possibility are then obtained considering the distance between the spectral response at each testing site and that mean value, and therefore, if the obtained value is relatively distant from the mean, even if its spectral response is very close to some values obtained for the training set, a low degree of possibility will be assigned to it. This classifier and the obtained degrees of possibility, have therefore a limited capability to translate the information provided by the training set. For example, if a pixel has a spectral response equal to the spectral response of one of the pixels used in the training set, but which is relatively far from the mean, it will not have a degree of possibility of belonging to that class equal to one, as should be expected.

The fuzzy classifier based on the FNNC has a better behavior on this aspect, since it translates the information contained in the training set in a more reliable way. For example, all points with a spectral response equal to the spectral response of a point included in the training set will have a degree of possibility equal to one. Furthermore, if the spectral response of pixels is located inside the regions populated by the spectral responses of a particular class, high degrees of similarity are expected to occur to that class.

3.3 Spatial variation of uncertainty

Fig. 10 shows on a), b) and c) respectively an extract of the classification with the FNNC and the values of the U_n and E uncertainty measures. Fig. 10 d), e) and f) show the corresponding images obtained with the FMDMC.

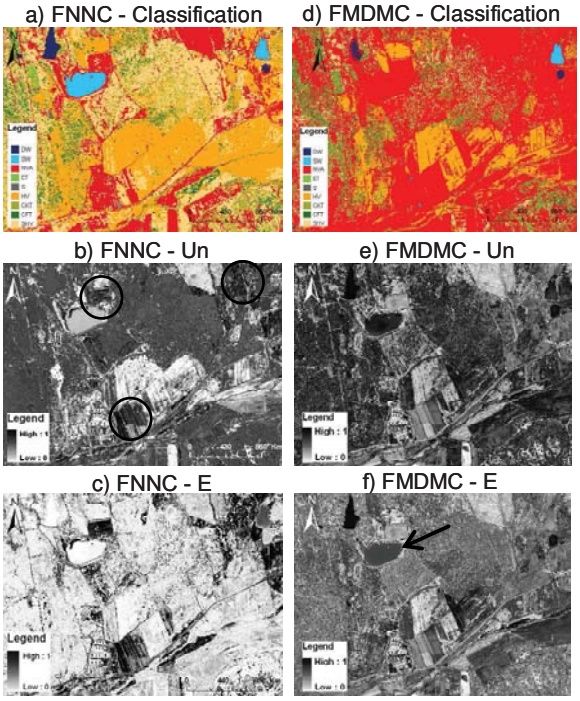


Figure 10: a), b) and c) show respectively the classification results, U_n and E uncertainty values obtained with the FNNC. Images d), e) and f) show the results corresponding to a), b) and c) obtained with the FMDMC.

It can be easily observed that the classification results are very different. Even though the accuracy evaluation and the

ISBN: 978-989-95079-6-8

uncertainty indexes already indicated that the classification with the FMDMC was worst, the visualization of the results clearly shows that it is bad. It can be seen that the FMDMC classified most of the image as NVA, while the FNNC manage to differentiate much more structures in the image.

Images shown in Fig. 10 b) and e) correspond to the spatial distribution of the U_n uncertainty measure of both classifications and may be used to illustrate the omission error committed when the pixel is assigned to the class corresponding to the largest degree of possibility. The regions with larger uncertainty are the ones where the degrees of possibility were more dispersed over other classes. Fig. 10b) shows that relatively large values of U_n were associated with regions that correspond to landscape units composed by different surface elements such as forest trees and Sparse Herbaceous Vegetation. Some region with higher U_n uncertainty can also be seen in the zones identified by the black circumferences. Fig. 10c) and f) show the commission errors, that is, the exaggeration committed when the best class is chosen. In general in Fig. 10c) low values are obtained for most of the image, higher values are mainly visible in some landscape boundaries and in the regions identified by the circumferences in Fig. 10b), which indicate, once again, that those regions may be problematic. Fig. 10f) shows that larger commission errors are obtained with the FMDMC than with the FNNC. One aspect clearly visible from Fig. 10e) and f) is that the spatial variation of uncertainty shows many structures invisible in the classification, such as the SW region shown in the region identified by the arrow.

4 Conclusions

The results obtained with the presented case study highlight that the classifiers FMDMC and FNNC presented possibility distributions with particular characteristics, which are a consequence of the method used to compute the degrees of possibility. The classification with the FNNC presented large possibility values for the best class, but also presented large values for the second and sometimes third class, and the classification with the FMDMC presented low values of the largest possibility. In both cases the presented characteristics increase uncertainty and are a signal of possible difficulties in the classification. In fact, neither of these classifications presented very good results, since Global Accuracies of only 70% and 53% were obtained.

It is also shown that when the classifications present relatively large and similar uncertainty values for all classes it is more difficult to take conclusions regarding the problematic classes only with these measures. Even though, an analysis of the possibility distributions may give valuable information. It is also very important to know how the degrees of possibility associated to the pixels are obtained, to understand what they represent.

In this study it was considered that the confusion matrix and the accuracy indexes were reliable, but in reality a human interpreted may have large difficulties in differentiating for example NVA from SHV, and this type of confusion between classes may influence greatly the accuracy indexes. In future analysis the uncertainty information present in the reference data, used to build the confusion matrixes, should be taken in consideration when comparing the results provided by the uncertainty and accuracy information.

The presented results also show that the spatialization of uncertainty may provide valuable information to the user. This will enable the identification of problematic regions, where the classification is probably less reliable. On the other hand, the spatial variation of uncertainty provides a great amount of information, which, if introduced in the classification itself, might be very useful.

References

[1] L. M. S. Gonçalves, *et al.*, Evaluation of soft possibilistic classifications with non-specificity uncertainty measures. *International Journal of Remote Sensing*. Accepted, 2009.

[2] A. Zhu, Measuring uncertainty in class assignment for natural resource maps under fuzzy logic. *Photogrammetric Engineering and Remote Sensing*, 63: 1195-1202, 1997.

[3] F. Wang, Improving remote sensing image analysis though fuzzy information representation. Photogrammetric Engineering and Remote Sensing, 56 (8): 1163-1169,1990.

[4] L. Bastin, Comparison of fuzzy c-means classification, linear mixture modeling and MLC probabilities as tools for unmixing coarse pixels. International Journal of Remote Sensing, 18 (17): 3629-3648, 1997.

[5] J. Zhang, G. M. Foody, Fully-fuzzy supervised classification of sub-urban land cover from remotely sensed imagery: statistical and artificial neural network approaches. *International Journal of Remote Sensing*, 22: 615-628, 2001.

[6] M. A. Ibrahim, *et al.*, Estimating and accommodating uncertainty through the soft classification of remote sensing data. *International Journal of Remote Sensing*, Vol.26: 2995-3007, 2005.

[7] G. M. Foody, Approaches for the production and evaluation of fuzzy land cover classifications from remotely-sensed data. *International Journal of Remote Sensing*, 17: 1317-1340, 1996.

[8] C. Ricotta, On possible measures for evaluating the degree of uncertainty of fuzzy thematic maps. *International Journal of Remote Sensing*, 26: 5573-5583, 2005.

[9] F. Maselli, *et al.*, Use of probability entropy for the estimation and graphical representation of the accuracy of maximum likelihood classifications. ISPRS Journal of Photogrammetry and Remote Sensing, 49 (2): 13-20,1994.

[10] W. G. Liu, *et al.*, Uncertainty and confidence in land cover classification using a hybrid classifier approach. Photogrammetric Engineering and Remote Sensing, 70 (8): 963-971, 2004.

[11] L. Bastin, P. Ficher, J. Wood, Visualizing uncertainty in multi-spectral remotely sensed imagery. Computers & Geosciences, 28:337-350, 2002.

[12] C. C. Fonte and W. Lodwick, Modelling the fuzzy spatial extent of geographical entities. In *Fuzzy Modeling with Spatial Information for Geographic Problems* (ed: Cobb, M., Petry, F., Robinson, V.). Springer-Verlag. 121-142. 2005.

[13] R. Yager, Measuring tranquility and anxiety in decision making: an application of fuzzy sets. *International Journal of General Systems*, 8: 139-146, 1982.

[14] M. Higashi, G. Klir, On measure of fuzziness and fuzzy complements. *International Journal of General Systems*, 8: 169-180, 1983.

[15] N. Pal, J. Bezdek, Quantifying different facets of fuzzy uncertainty. In *Fundamentals of Fuzzy Sets,* Dubois, D., Prade, H. (eds.), The Handbook of Fuzzy Sets Series, Kluwer Acad. Publ: 459-480., 2000.

[16] R. Yager, On the specificity of a possibility distribution. *Fuzzy Sets and Systems*, 50: 279-292, 1992.

ISBN: 978-989-95079-6-8

On interval and fuzzy calculations of economic uncertainty

Hans Schjær-Jacobsen

Copenhagen University College of Engineering
15 Lautrupvang, DK-2850 Ballerup, Denmark
Email: hsj@ihk.dk

Abstract—*This paper emphasizes that numerically correct calculation of economic uncertainty with intervals and fuzzy numbers requires implementation of global optimization techniques in contrast to straightforward application of interval arithmetic. In general, the latter approach produces incorrect results in the sense that the degree of uncertainty is severely overestimated with the risk of improper decision making. This is demonstrated by a simple case from managerial economics as well as a cost estimation case from a real life railway reconstruction project.*

Keywords—Economic uncertainty, fuzzy numbers, intervals, possibility, probability.

1 Introduction

Criticism has been raised towards probability theory as being a too normative framework to take all the aspects of uncertain judgement into account, Dubois and Prade [1]. In this paper we will focus on alternative methods of modelling of economic uncertainty like the interval representation and the fuzzy number representation.

The interval representation, Moore [2], [3] is particularly well suited to a situation where the knowledge of an uncertain parameter is limited to knowing its minimum and maximum value whereas nothing else is known. Based upon a mathematical theory of interval analysis this approach has shown to be useful in keeping track of worst and best cases in economic analyses and thus contribute to improved decision processes, see e.g. Schjær-Jacobsen [4], [5]. The introduction of fuzzy sets supports the concept of possibility rather than probability and translates natural language expressions into the mathematical formalism of possibility measures. It is generally recognized that possibility is distinct from probability. Probabilities can be interpreted as relative frequencies or, more generally, uncertain knowledge or belief of a statistical nature. In contrast, possibility relates to the degree of feasibility and ease of attainment or imprecise knowledge. Particularly, this paper will deal with some computational aspects of uncertainty representation by intervals and fuzzy numbers. It will be demonstrated that correct results are obtained in the general case by applying global optimization. It is the author's experience that these aspects are widely neglected, probably due to lack of communication and interaction between professional communities each one advocating either concept.

2 A simple case from managerial economics

In order to demonstrate the principles of numerical modelling of uncertainty by intervals and fuzzy numbers presented in this paper we introduce a simple yet instructive referential case from managerial economics. Consider a company selling one product at price p and quantity q into a market. For the sake of simplicity we may assume fixed cost to be zero. The turnover TR is

$$TR = TR(p,q) = p \cdot q, \tag{1}$$

and the variable cost VC is given by

$$VC = VC(q) = 20 \cdot q. \tag{2}$$

Then the profit π is

$$\pi = \pi(p,q) = TR(p,q) - VC(q) = p \cdot q - 20 \cdot q. \tag{3}$$

We want to find the price p* and quantity q* that gives the maximum profit π^*. It is seen from (3) that the profit is a monotone function of q and p which means that no maximum of π exists unless variables are constrained. If for example $p \leq 65$ and $q \leq 450$, then $p^* = 65$, $q^* = 450$, and $\pi^* = 20.250$.

Next we suppose that price and quantity are interdependent variables due to market conditions. Assuming that they are connected by the demand function

$$p = p(q) = 100 - 0{,}1 \cdot q \tag{4}$$

we get for the turnover

$$TR = TR(p) = -10 \cdot p^2 + 1.000 \cdot p, \tag{5}$$

the variable cost

$$VC = VC(p) = -200 \cdot p + 20.000, \tag{6}$$

and the profit margin

$$\pi = \pi(p) = -10 \cdot p^2 + 1.200 \cdot p - 20.000. \tag{7}$$

The profit of (7) is maximized by the optimality criterion that marginal cost MC be equal to marginal turnover MR which gives the results

ISBN: 978-989-95079-6-8

$$p^* = 60,\ q^* = 400, \text{ and } \pi^* = 16.000. \quad (8)$$

In the following, we shall refer to the above examples for explanations of computational details and special features.

3 Uncertainty modelling using intervals

2.1 Basics of interval analysis

Following Moore [3] and Caprani, Madsen, and Nielsen [6] we define a real interval number as an ordered pair [a; b] of real numbers with $a \le b$. It may also be defined as an ordinary set of real numbers x such that $a \le x \le b$, or

$$[a; b] = \{x \mid a \le x \le b\}. \quad (9)$$

If the basic arithmetic operations addition, subtraction, multiplication, and division are denoted by the symbol #, we can define operations on two intervals $\mathbf{I}_1 = [a_1; b_1]$ and $\mathbf{I}_2 = [a_2; b_2]$ based on the set-theoretic formulation:

$$\mathbf{I}_1 \# \mathbf{I}_2 = \{x \# y \mid a_1 \le x \le b_1,\ a_2 \le y \le b_2\}. \quad (10)$$

For basic operations on the intervals $\mathbf{I}_1$ and $\mathbf{I}_2$ we get the resulting interval $\mathbf{I} = [a; b]$ by the formulas

$$\begin{aligned}
&\mathbf{I} = \mathbf{I}_1 + \mathbf{I}_2 = [a_1+a_2;\ b_1+b_2],\\
&\mathbf{I} = \mathbf{I}_1 - \mathbf{I}_2 = [a_1-b_2;\ b_1-a_2],\\
&\mathbf{I} = \mathbf{I}_1 \cdot \mathbf{I}_2 = [\min(a_1 \cdot a_2,\ a_1 \cdot b_2,\ b_1 \cdot a_2,\ b_1 \cdot b_2);\\
&\qquad \max(a_1 \cdot a_2,\ a_1 \cdot b_2,\ b_1 \cdot a_2,\ b_1 \cdot b_2)], \quad (11)\\
&\mathbf{I} = \mathbf{I}_1 / \mathbf{I}_2 = [\min(a_1/a_2,\ a_1/b_2,\ b_1/a_2,\ b_1/b_2);\\
&\qquad \max(a_1/a_2,\ a_1/b_2,\ b_1/a_2,\ b_1/b_2)],\ 0 \notin [a_2; b_2].
\end{aligned}$$

It can be shown that the four basic interval operations are inclusion monotonic, commutative, and associative. However, the distributive rule is not valid in general. Instead, the so-called sub-distributivity holds, but only for addition and multiplication [6]

$$\mathbf{I}_1 \cdot (\mathbf{I}_2 + \mathbf{I}_3) \subseteq \mathbf{I}_1 \cdot \mathbf{I}_2 + \mathbf{I}_1 \cdot \mathbf{I}_3. \quad (12)$$

From a rational real valued function F of n real valued variables

$$F = F(x_1, x_2, \ldots, x_n) \quad (13)$$

we can create the interval extension function as an interval function $\mathbf{F}^I$ of n intervals

$$\mathbf{F}^I = \mathbf{F}^I(\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_n) \quad (14)$$

simply by replacing the real operators by interval operators and the real variables by intervals.

A rational function can be formulated in many ways whereas the same reformulations cannot be done for interval expressions due to the invalidity of the distributive rule. This implies that different formulations of a rational function will lead to different interval extension functions and thus to different interval results [6]. In the case of F being a monotonic function within the entire range of the input variables the minimum and maximum of $\mathbf{F}^I$ as an interval can simply be found among the function values F at the extreme points of the variables. In the general case of F being non-monotonic or variables appearing more then once, the calculation of $\mathbf{F}^I$ as an interval is non-trivial, which is demonstrated in the following example.

Example: Based on the real valued function $F = x \cdot (1 - x)$ the interval function $\mathbf{F}^I = \mathbf{I} \cdot (1 - \mathbf{I})$, $\mathbf{I} = [0; 1]$ is calculated. Straightforward application of formulas from (11) gives the result $\mathbf{F}^I = [0; 1]$ whereas the correct result is [0; 0,25].

In this paper the term "correct" is used to indicate the narrowest possible interval that can be calculated for an uncertain variable. Generally, to obtain this, iterative global optimization methods have to be used, see e.g. Hansen [7] and Kjøller *et al.* [8]. In order to obtain correct results (as in the above example) to an accuracy specified by the user, interval calculations in this paper are carried out using the Interval Solver 2000 program, Hyvönen and De Pascale [9], [10], as an add-in module to MS-Excel 2000. An overall absolute and relative precision of 10^{-6} has been applied.

Correct calculation of interval functions allows for strong statements about the uncertainties involved. *Firstly*, you can say that provided all uncertain input variables stay within their minimum and maximum values, the uncertain output function will stay within its minimum and maximum values. *Secondly*, the uncertain output function will not attain any value that is not a function value of some combination of the uncertain input values (within their minimum and maximum values).

2.2 Independent and interdependent variables

Independent variables using interval arithmetic

To calculate the uncertain profit we first use the formulas of interval arithmetic (11) by three different ways of calculation. As an example look at the independent and uncertain quantity and price

$$\mathbf{p} = [55; 65],\ \mathbf{q} = [350; 450], \quad (15)$$

Firstly, we use the turnover and variable cost as intermediate variables. By (1) we get for the uncertain turnover **TR** and by (2) for the uncertain variable cost **VC**

$$\begin{aligned}
&\mathbf{TR} = [55; 65] \cdot [350; 450] = [19.250; 29.250],\\
&\mathbf{VC} = 20 \cdot [350; 450] = [7.000; 9.000]. \quad (16)
\end{aligned}$$

Then by (3) we get for the uncertain profit

$$\boldsymbol{\pi} = \mathbf{TR} - \mathbf{VC} = [10.250; 22.250]. \quad (17)$$

However, the above calculation produces a too wide interval (17) for the profit. The reason for this is that in the expression (3) the variable q appears twice thus allowing the quantity q used to calculate **TR** to be different from the quantity used to calculate **VC**.

ISBN: 978-989-95079-6-8

Secondly, the rightmost form of (3) is used, giving

$$\boldsymbol{\pi} = [350; 450]\cdot[55; 65] - 20\cdot[350; 450] = [10.250; 22.250], \quad (18)$$

which is identical to (17) (and thus incorrect) because the arithmetic operations are identical.

Thirdly, (3) is rearranged before the interval calculations are carried out:

$$\pi = \pi(p,q) = q\cdot(p - 20). \quad (19)$$

We then get for the uncertain profit

$$\boldsymbol{\pi} = [350; 450]\cdot([55; 65] - 20) = [350; 450]\cdot[35; 45] = [12.250; 20.250], \quad (20)$$

which is a somewhat narrower interval than (17) and (18) because each variable is appearing only once in (19). Actually, (20) is the correct result. This can easily be verified simply by closer numerical inspection of the profit function (3) for various combinations of the variables p and q. The results are summarised in Table 1.

Independent variables using global optimization

Next, we calculate the uncertain profit by global optimization using Interval Solver 2000. With the same input variables (15) and intermediate variables **TR** and **VC** we get from (3)

$$\mathbf{TR} = [19.250; 29.250],\ \mathbf{VC} = [7.000; 9.000],\ \boldsymbol{\pi} = \mathbf{TR} - \mathbf{VC} = [10.250; 22.250], \quad (21)$$

which is identical to (17) and (18) for the same reason as mentioned above. With the same input variables and by way of formulas (18) and (19) we get the result

$$\mathbf{p} = [55; 65],\ \mathbf{q} = [350; 450],\ \boldsymbol{\pi} = [12.250; 20.250], \quad (22)$$

which is seen to be the correct result identical to (20). The results are summarised in Table 1.

Interdependent variables using interval arithmetic

To perform an uncertainty analysis around a given price p = 60 we set

$$\mathbf{p} = [55; 65]. \quad (23)$$

Obviously, since the quantity variable q can be eliminated by using the demand function (4), the number of variables is reduced from two to one, namely the price. Even in this simple case, as it turns out, the results obtained by interval arithmetic can be dramatically incorrect due to the fact that the profit function is now non-monotonic.

The profit is calculated in two different ways using interval arithmetic (11). *Firstly*, the profit is calculated by

Table 1. Summary of uncertainty analyses, independent price **p** = [55; 65] and quantity **q** = [350; 450]. Correct results shown in *italics*.

Calculation Method	$\boldsymbol{\pi} = \mathbf{TR} - \mathbf{VC}$	$\boldsymbol{\pi} = \mathbf{p}\cdot\mathbf{q} - 20\cdot\mathbf{q}$	$\boldsymbol{\pi} = \mathbf{q}\cdot(\mathbf{p} - 20)$
Interval Arithmetic	[10.250; 22.250]	[10.250; 22.250]	*[12.250; 20.250]*
Global Optimization	[10.250; 22.250]	*[12.250; 20.250]*	*[12.250; 20.250]*

intermediate variables **TR** and **VC** according to (5) and (6), respectively

$$\mathbf{TR} = -10\cdot[55; 65]^2 + 1.000\cdot[55; 65] = [12.750; 34.750] \quad (24)$$

and

$$\mathbf{VC} = -200\cdot[55; 65] + 20.000 = [7.000; 9.000], \quad (25)$$

which gives

$$\boldsymbol{\pi} = \mathbf{TR} - \mathbf{VC} = [3.750; 27.750]. \quad (26)$$

Secondly, the profit is calculated by the above formula

$$\boldsymbol{\pi} = -10\cdot\mathbf{p}^2 + 1.200\cdot\mathbf{p} - 20.000 \quad (27)$$

yielding

$$\boldsymbol{\pi} = -10\cdot[55; 65]^2 + 1.200\cdot[55; 65] - 20.000 = [3.750; 27.750]. \quad (28)$$

The results found above are identical but way out of order, since the correct result of the uncertain profit is [15.750; 16.000]. This is easily verified by closer numerical inspection of (7) with varying p.

Interdependent variables using global optimization

Global optimization is used to calculate the uncertain profit by means of (26) yielding [13.750; 17.750] and (27) yielding [15.750; 16.000]. It may be observed that correct results are produced in the latter case but not in the former, due to usage of intermediate variables **TR** and **VC**, which allows for a too wide resulting interval. The results are shown in Table 2.

Table 2. Summary of uncertainty analyses, interdependent price **p** = [55; 65] and quantity $\mathbf{q} = 1.000 - 10\cdot\mathbf{p}$. Correct result shown in *italics*.

Calculation Method	$\boldsymbol{\pi} = \mathbf{TR} - \mathbf{VC}$	$\boldsymbol{\pi} = -10\cdot\mathbf{p}^2 + 1.200\cdot\mathbf{p} - 20.000$
Interval Arithmetic	[3.750; 27.750]	[3.750; 27.750]
Global Optimization	[13.750; 17.750]	*[15.750; 16.000]*

ISBN: 978-989-95079-6-8

The differences in Tables 1 and 2 between correct intervals and the more wide intervals are considerable. When an uncertain variable is appearing more than once, interval arithmetic will usually produce too wide results and global optimization must to be used. With non-monotonic functions, global optimization is necessary. In all cases, intermediate variables should be avoided, since both interval arithmetic and global optimization will produce incorrect results.

4 Uncertainty modelling using fuzzy numbers

4.1 Fuzzy numbers and intervals

A fuzzy set A in X where X is a space of points (objects) with a generic element of X denoted by x, i.e. X = {x}, is characterized by a membership function $f_A(x)$ which associates with each point in X a real number in the interval [0; 1]. The value of the membership function $f_A(x)$ at x represents the "grade of membership" of x in A. Thus the closer the value of $f_A(x)$ to unity, the higher the grade of membership of x in A. Note that when A is an ordinary set, i.e. non-fuzzy, the membership function can take only two values 0 and 1. In other words, a fuzzy set is a set of ordered pairs (x, $f_A(x)$)

$$A = \{(x, f_A(x)) \mid x \in X\}. \tag{34}$$

It is also useful to define the ordinary (non-fuzzy) set A_α as the α-cut of A:

$$A_\alpha = \{x \in X \mid f_A(x) \geq \alpha, 0 \leq \alpha \leq 1\}. \tag{35}$$

In this paper we are mainly interested in the concept of fuzzy numbers as a means of representing uncertain or fuzzy information, Dubois and Prade [11], [12]. In addition to the simplest fuzzy number, namely the interval, we also make use of the triangular fuzzy number, Chiu and Park [13], [a; c; b] where a ≤ c ≤ b, that can be defined by its membership function:

$$\begin{aligned} f(x) &= (x-a)/(c-a), && a \leq x \leq c, \\ &= (b-x)/(b-c), && c \leq x \leq b, \\ &= 0, && \text{otherwise.} \end{aligned} \tag{36}$$

Mathematical operations on triangular fuzzy numbers can be facilitated by introducing the left L(α) and right R(α) representation of a fuzzy triangular number $\mathbf{F}^T$, refer to the α-cut (35):

$$\begin{aligned} &\mathbf{F}^T = [L(\alpha); R(\alpha)], \text{ where} \\ &L(\alpha) = a + (c-a)\alpha \text{ and } R(\alpha) = b + (c-b)\alpha, \\ &\alpha \in [0, 1]. \end{aligned} \tag{37}$$

Observe that in this notation a fuzzy number is written as an interval with upper and lower bounds depending on α. This means that addition, subtraction, multiplication, and division can be carried out by using interval methods for all values of α. Likewise, for any triangular function, the resulting triangular functional values can be calculated and represented by L and R functions using interval methods for all values of α.

Example: Based on the real valued function F = x·(1 – x) calculate the corresponding fuzzy function with triangular argument [0; 0,5; 1]. Correct results have been calculated with global optimization and are shown in Table 3. From the results in Table 3 it can be seen that the function F has a maximum of 0,250 at x = 0,5 corresponding to α = 1. Also note that the R(α) function has been correctly calculated to 0,250 for all values of α.

To obtain simpler representations and reduce the number of calculations, triple and quadruple representations of fuzzy variables corresponding to α-cuts 0 and 1 in (35) may be used. In the above example the result then is [0; 0,25; 0,25], where the extreme function values is obtained by global optimization on the interval [0; 1] and the interior point is obtained by conventional calculation at x = 0,5.

4.2 Simple case with triangular fuzzy numbers

Next we calculate the simple case from Section 2 using profit function (3) with independent triangular fuzzy input parameters **p** = [55; 60; 65], **q** = [350; 400; 450] defined by (36). The resulting membership function is shown in Table 4 for different values of α. It is easily seen that the value α = 0 corresponds to the correct profit interval previously found and α = 1 corresponds to the single point calculation of the profit. This is one of the important features of the triangular fuzzy number representation of uncertainty: It is easily communicated and understood that the ordinary single point calculation is extended to an interval around the single point representing the uncertain value of the input variable. The resulting triangular fuzzy profit is interpreted as follows: With the given uncertain input variables, the most possible value of the profit is 16.000 and profits outside the interval [12.250; 20.250] are impossible.

Calculations have also been carried out with interdependent triangular fuzzy input price parameter corresponding to **p** = [55; 60; 65] and the profit function (7), the results are shown in Table 5. For all values of α the correct maximum profit of 16.000 has been found. No profit values outside the interval [15.750; 16.000] are possible given the uncertain price input variable **p**.

Table 3. Fuzzy extension of x·(1-x) calculated with triangular argument [0; 0,5; 1].

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
L(α)	0,000	0,047	0,090	0,127	0,160	0,188	0,210	0,227	0,240	0,248	0,250
R(α)	0,250	0,250	0,250	0,250	0,250	0,250	0,250	0,250	0,250	0,250	0,250

ISBN: 978-989-95079-6-8

Table 4. Uncertain profit (3) by global optimization, triangular fuzzy input **p** = [55; 60; 65], **q** = [350; 400; 450].

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
L(α)	12.250	12.603	12.960	13.323	13.690	14.063	14.440	14.823	15.210	15.603	16.000
R(α)	20.250	19.803	19.360	18.923	18.490	18.063	17.640	17.223	16.810	16.403	16.000

Table 5. Uncertain profit (7) by global optimization, triangular fuzzy input **p** = [55; 60; 65].

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
L(α)	15.750	15.797	15.840	15.877	15.910	15.937	15.960	15.977	15.990	15.977	16.000
R(α)	16.000	16.000	16.000	16.000	16.000	16.000	16.000	16.000	16.000	16.000	16.000

Table 6. Cost estimation for railway reconstruction case by triangular fuzzy numbers.

Var	Code	Item	[a; c; b]
	0,00	Management and specs.	[1.732; 1.780; 1.884]
$\mathbf{X_1}$	0,10	Project management	[524; 540; 575]
$\mathbf{X_2}$	0,20	Construction management etc.	[975; 1.000; 1.050]
$\mathbf{X_3}$	0,30	Design specifications etc.	[233; 240; 259]
$\mathbf{X_4}$	10,00	Environmental and soil eng.	[864; 888; 950]
$\mathbf{X_5}$	20,00	Traffic tasks	[48; 50; 53]
	30,00	Renewal of tracks	[8.905; 9.190; 9.664]
$\mathbf{X_6}$	30,10	New outbound main track	[975; 1.000; 1.050]
$\mathbf{X_7}$	30,20	Track renewal at platform 3/5	[5.432; 5.600; 5.880]
$\mathbf{X_8}$	30,30	New platform edge	[1.533; 1.580; 1.643]
$\mathbf{X_9}$	30,40	Track renewal depot, West	[285; 300; 321]
$\mathbf{X_{10}}$	30,50	Track layout design	[682; 710; 770]
$\mathbf{X_{11}}$	40,00	Platform and station	[538; 560; 602]
$\mathbf{X_{12}}$	50,00	Safety and signal installations	[5.035; 5.245; 5.586]
	60,00	Informatics incl. power supply	[2.374; 2.417; 2.626]
$\mathbf{X_{13}}$	60,10	Phase 2-4	[78; 80; 86]
$\mathbf{X_{14}}$	60,20	Sub project management	[249; 259; 275]
$\mathbf{X_{15}}$	60,30	Passenger information	[1.009; 1.030; 1.123]
$\mathbf{X_{16}}$	60,40	Electrical power supply	[1.038; 1.048; 1.142]
	70,00	Overhead line incl. pylons	[3.507; 3.624; 3.787]
$\mathbf{X_{17}}$	70,10	Overhead cables	[3.021; 3.122; 3.262]
$\mathbf{X_{18}}$	70,20	Layout and planning	[486; 502; 525]
$\mathbf{Y_1}$		Total cost before corrections	[23.003; 23.754; 25.152]
$\mathbf{X_{19}}$	A	Internal decision process	[1,006; 1,032; 1,098]
$\mathbf{X_{20}}$	B	Design specifications etc.	[1,009; 1,040; 1,100]
$\mathbf{X_{21}}$	C	Working process	[1,021; 1,042; 1,084]
Y		Total cost after corrections	[23.842; 26.565; 32.930]

Table 7. Total cost after corrections for railway reconstruction case by triangular fuzzy numbers.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
L(α)	23.842	24.108	24.353	24.624	24.875	25.148	25.480	25.734	26.020	26.276	26.565
R(α)	32.930	32.244	31.592	30.895	30.269	29.626	28.983	28.387	27.743	27.161	26.565

ISBN: 978-989-95079-6-8

5 A railway reconstruction project

Consider the case of estimating the total cost incurred by a railway reconstruction project described by independent fuzzy input variables, namely 18 cost items $\mathbf{X}_1$,..., $\mathbf{X}_{18}$ and 3 correction factors $\mathbf{X}_{19}$,..., $\mathbf{X}_{21}$. This case has been treated previously by using the concept of imprecise stochastic variables, Schjær-Jacobsen [14]. The correction factors are introduced in order to account for overall influences not accounted for by the individual cost items. The total cost before corrections is the sum

$$\mathbf{Y}_1 = \mathbf{X}_1 + \mathbf{X}_2 + ... + \mathbf{X}_{18}. \quad (38)$$

The total cost after corrections $\mathbf{Y} = \mathbf{Y}(\mathbf{X})$ is a function of all 21 uncertain variables

$$\mathbf{Y} = (\mathbf{X}_1 + \mathbf{X}_2 + ... + \mathbf{X}_{18}) \cdot \mathbf{X}_{19} \cdot \mathbf{X}_{20} \cdot \mathbf{X}_{21}. \quad (39)$$

The uncertain input variables should be estimated by railway experts with relevant project experience. Subsequently, the total cost $\mathbf{Y}$ is calculated by means of global optimization. The total cost estimation results are shown in Table 6 in the case of triangular fuzzy input variables. The term "code" refers to the cost structure hierarchy. For example the item carrying code 0,00 is the sum of cost items at lower levels carrying the codes 0,10, 0,20, and 0,30. Thus the latter are input variables whereas the former is an output variable. The cost item carrying the code 10,00 is an input variable itself because it does not have cost items at lower levels involved. In Table 7 the membership function of total cost $\mathbf{Y}$ after corrections is shown.

For $\alpha = 1$ the single point calculation of 26.565 is shown corresponding to an ordinary calculation without uncertainty. This value is the most possible one and is attained when all input variables are attaining their most possible values. For $\alpha = 0$ the total cost after corrections is within the interval [23.842; 32.930] and any value outside this interval is impossible considering the uncertain input variables.

It should be mentioned here that comparative studies of approaches involving probabilistic methods (including Monte Carlo simulation), trapezoidal fuzzy numbers, and imprecise stochastic variables have been made in [14] and [15] and is further investigated in a research project concerning economic uncertainty in mega projects.

6 Conclusion

In order to calculate correct results with intervals and fuzzy numbers global optimization methods must be implemented in contrast to straightforward application of interval arithmetic. However, the experimental calculations reported in the paper also show that further care should be taken not to introduce intermediate variables, which may also result in excess width intervals. When calculating fuzzy extensions of non-monotonic functions global optimization must be used in order to produce correct results. Otherwise, additional and unnecessary uncertainty will arise eventually leading to wrong decisions.

The results of the paper indicate that correct calculations of intervals (and fuzzy number membership functions) allow for rather strong statements pertaining to economic uncertainties. For example, it can be said that provided all uncertain input variables stay within their limits the uncertain output variables will stay within their limits. Obviously, it is an advantage of the interval and triangular (and trapezoidal) fuzzy representation of uncertainty that the meaning is easily communicated and understood.

References

[1] D. Dubois and H. Prade. *Possibilty theory – An approach to computerized processing of uncertainty.* New York: Plenum Press, 1988.

[2] R.E. Moore. *Interval arithmetic and automatic error analysis in digital computing.* Ph.D. dissertation, Stanford University, USA, 1962.

[3] R.E. Moore. *Interval analysis.* New Jersey: Prentice-Hall, 1966.

[4] H. Schjær-Jacobsen. A new method for evaluating worst- and best-case economic consequences of technological development. *International Journal of Production Economics,* 46-47: 241-250, 1996.

[5] H. Schjær-Jacobsen. Handling economic risks and uncertainties of production technology investments. *Proceedings of the 14th International Conference on Production Research, Osaka, Japan,* 1: 346-349, 1997.

[6] O. Caprani, K. Madsen, H.B. Nielsen. *Introduction to interval analysis*. Lecture notes, Department of Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark, November 20, 2002.

[7] E. Hansen. *Global optimization using interval analysis*. Marcel Dekker, New York, USA, 1992.

[8] S. Kjøller, P. Kozine, K. Madsen, O. Stauning. *Non-linear global optimization using interval arithmetic and constraint propagation*. In: A. Törn and J. Zilinskas (Eds.), Models and algorithms for global optimization, Springer Verlag, 2007.

[9] E. Hyvönen and S. de Pascale. A new basis for spreadsheet computing: Interval Solver for Microsoft Excel. *Proceedings of the 11th Innovative Applications of Artificial Intelligence (IAAI-99)*, AAAI Press, Menlo Park, California, 1999.

[10] E. Hyvönen and S. de Pascale. *Interval Solver 2000 for Microsoft Excel*. User's Guide, Version 4.0. Program Release 4.0.0.2, Delisoft Ltd., Helsinki, Finland, 2000.

[11] D. Dubois and H. Prade. Operations on fuzzy numbers. *International Journal of System Science*, 9: 613-626, 1978.

[12] D. Dubois and H. Prade. Fuzzy real algebra: Some results. *Fuzzy Sets and Systems*, 2: 327-348, 1979.

[13] C.U. Chiu and C.S. Park. Fuzzy cash flow analysis using present worth criterion. *The Engineering Economist*, 39(2): 113-138, 1994.

[14] H. Schjær-Jacobsen. A unified approach to variability and imprecision in modelling of economic uncertainty. *Proceedings of the 15th International Working Seminar on Production Economics*, 1: 437-447, Innsbruck, Austria, March 3-7, 2008.

[15] H. Schjær-Jacobsen. Modeling of economic uncertainty, *Fuzzy Economic Review*, IX(2): 49-73, 2004.

ISBN: 978-989-95079-6-8

A method for locating the iliac crests based on the fuzzy Hough transform

Sergio Jaime-Castillo Juan Miguel Medina Antonio Garrido

Department of Computer Science and Artificial Intelligence, University of Granada
C/ Periodista Daniel Saucedo Aranda s/n, 18071 Granada, Spain
Email: {sjaime,medina,a.garrido}@decsai.ugr.es

Abstract— *In this paper we present a method based on the fuzzy Hough transform to locate the iliac crests in posteroanterior radiographs. Firstly, a Canny edge detector is applied to the X-ray images in order to obtain an edge map. Then, the fuzzy Hough transform is used in combination with several constraints extracted from our domain knowledge. At the bottom of the radiograph we search for two circular arcs, as we can represent the shape of the iliac crests by means of these curves. Locating the iliac crests can be useful for a later processing of the X-rays in order to obtain a fully automated segmentation of the whole spine.*

Keywords— Computer-aided diagnosis, fuzzy Hough transform, medical imaging, X-rays.

1 Introduction

Computer-aided diagnosis is increasing its importance in healthcare. This kind of systems can help physicians to perform routine tasks that require a long training and a great amount of time, and can provide more accurate and consistent measures that may improve the diagnostic process.

Modern medical imaging techniques such as computed tomography (CT) and magnetic resonance imaging (MRI) provide images of high quality which can help physicians enormously in diagnostic process. However, these techniques are usually expensive and give patients high doses of radiation. Therefore, they are not suitable for all diagnoses and, although X-rays do not provide as much quality as the other ones, they are widely used in diagnosis of many diseases.

Much research has been done to segment anatomical structures in CT and MRI. Nevertheless, X-rays have not been as widely studied as the other ones due to its lower quality. In our work we use X-ray images because they are extensively used in diagnosing scoliosis, which is the problem we are focused on.

Scoliosis is a three-dimensional deformation of the spine that produces vertebral rotation and crushing, and lateral curvature. It is typically classified as congenital (caused by vertebral anomalies present at birth), idiopathic (sub-classified as infantile, juvenile, adolescent or adult according to the time when the onset occurred) or as a secondary symptom of another condition, such as cerebral palsy, spinal muscular atrophy or due to a physical trauma.

Diagnosis of scoliosis is performed by calculating Cobb angle ([1]) manually on X-rays. Automatic measure of Cobb angle could improve diagnosis since it can provide more reliable and consistent measures. In order to assess Cobb angle automatically we need to segment the spine accurately and, as we will see afterwards, the proposed method can be useful to do so.

The iliac crest is the long and curved upper border of the wing of the ilium. According to a study performed by Render in [2], the line connecting the iliac crests intersects the spine at the level of the fourth lumbar vertebra or the interspace between the fourth and the fifth lumbar vertebra in 78.5% of cases, while the point of intersection was lower than the fourth lumbar vertebra in 17.8% and at the level of the interspace between the third and the fourth lumbar vertebra in 3.7%. We are interested in locating these anatomical structures as we can use them as a reference to initialise a segmentation algorithm that is intended to segment the whole spine. In the same way, this reference point will be useful to label the vertebrae found by the segmentation algorithm, since we can use the results of Render.

In this work, we have used full-length standing posteroanterior X-rays. The pictures were acquired from a PACS (Picture Archiving and Communication System) as DICOM (Digital Imaging and COmmunication in Medicine) files with 5700×2400 pixels of resolution. The problem with this kind of images is the presence of noise and distortions due to its acquisition process, that makes difficult to detect and identify the anatomical structures present in the X-rays accurately. Other important problems are poor contrast and low visibility of some details in the images. In order to deal with these problems, first of all we have to preprocess the images to obtain new ones with less quantity of noise and, as far as possible, better visibility than the initial ones.

The rest of the paper is organised as follows: in section 2 we describe the foundations and the operation of the fuzzy Hough transform. In section 3 we present our method for locating the iliac crests in posteroanterior radiographs and, in section 4, we discuss the results achieved by it. Finally, in section 5, we comment some aspects of our future work related to this paper.

2 The fuzzy Hough transform

The Hough transform (HT) was originally proposed as a technique to detect straight lines by Hough in 1962 ([3]). It was extended to detect general curves with known analytical expressions by Duda and Hart in [4] and, later, Ballard proposed a new extension in [5] to detect any curve although its expression was not known. In this work we use the fuzzy Hough transform (FHT), another extension of the original algorithm developed by Han *et al.* in [6].

A straight line can be represented as $\rho = x\cos\theta + y\sin\theta$, where θ is the angle of the vector from the origin to the closest point of the line, ρ is the lowest distance between the line and the origin and (x, y) are the coordinates of a point in the image space. In that way, a line is defined by these two parameters ρ

ISBN: 978-989-95079-6-8

and θ. Similarly, a circle and other curves can be represented by its parameters. The main goal of the algorithm is to detect shapes by checking which points in the edge map belong to the curve defined by a certain point in the parameter space. This is done by mapping the image space to the parameter space, where each point defines a different curve. By doing so we simplify the problem to locate a peak in an accumulator array.

In the original HT, the value at a point of the parameter space is the number of points in the image space that occur strictly on the shape defined by this point of the parameter space. There is no difference between a point being far from the ideal line or being just a little off it. Because of this, with the HT is difficult to detect approximate shapes that may appear in noisy environments due to the noise itself, or due to the not accurate edge detection performed by the feature detector. Fuzzy theory can help to deal with the detection of those approximate shapes.

We have assumed that we can represent the shape of the iliac crest by two circular arcs. However, the iliac crest is not exactly a circular arc, although it is close to. Because of this, the results obtained by applying the conventional HT to the X-rays are not accurate, so we propose the use of the FHT in order to achieve better results. In section 4 we make a comparison between both techniques based on the results achieved by them.

The main idea in the FHT is that each edge point contributes more or less to one point in the parameter space depending on its distance from the ideal curve defined by this point. To do this, once the accumulator is computed, Han *et al.* proposed to convolve it with the membership function

$$g(d) = \begin{cases} ke^{-d^2/\sigma^2} & if\ d < R \\ 0 & otherwise \end{cases} . \qquad (1)$$

As they proved, when the membership function of fuzzy points is defined as a circle the transforms of those points lying inside it is bounded in the parameter space by r, which is the radius of that membership function. That is true for straight lines and circles and, as we are considering circular arcs in such a way that they are particular instances of circles with some restrictions, we shall use this smoothing function.

In (1) k, σ and R are constants that are chosen empirically and d is the distance between the two points of the image considered. In our case, they are set as $k = 1$, $R = 3$ and $\sigma = 1.3$ in the experiments.

3 The proposed method to locate the iliac crest

In this section we present our method for automated detection of the iliac crest, based upon the fuzzy Hough transform. The method starts with the acquisition of the X-rays and finishes with the location of the iliac crests. Fig. 1 shows the operation of the method. It involves basically the search for circular arcs at the bottom of the radiograph, where the iliac crests are supposed to be, by means of the FHT.

As we have mentioned beforehand, a preprocessing stage is necessary to deal with noise and distortions present in the images. In this work we have used a combination of unsharp masking (UM), adaptive histogram equalisation and median filter to achieve images more suitable to be analysed by a computer vision system.

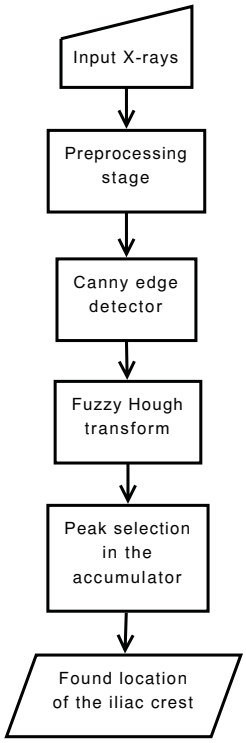


Figure 1: Flowchart of the proposed method to detect the iliac crest

Unsharp masking intends to improve the visibility of images by emphasising smoothed edges while the noise components are suppressed. The effect of the algorithm onto the images is edge sharpening, but usually it tends to reduce its contrast. Because of this, after the unsharp masking procedure we perform adaptive histogram equalisation to enhance it.

In unsharp masking, the high-frequency components of the image are extracted and then are added to the original image. First of all, a slightly blurred version I_{blur} of the image I_{orig} is obtained by convolving it with a Gaussian kernel and then is subtracted from I_{orig} to obtain I_{edge}. After that, I_{edge} is multiplied by a sharpness factor and the result, which contains the enhanced high-frequency components of the image, is then added to I_{orig}.

In our case, the unsharp masking is done by using

$$I_{sharp}(x,y) = I_{orig}(x,y) + 6 \cdot (I_{orig}(x,y) - I_{blur}(x,y)). \quad (2)$$

In (2) $I_{orig}(x,y)$ represents the original image and $I_{blur}(x,y)$ is a blurred version of the original image obtained by convolving $I_{orig}(x,y)$ with a Gaussian kernel of size 5×5 pixels and a variance of 5.

After the unsharp masking, we apply adaptive histogram equalisation to the image in order to enhance the contrast of the resultant image. Adaptive histogram equalisation divides the image into several regions and then distributes pixels values in each region of the image uniformly throughout the grey scale. It is performed in order to achieve a higher contrast for the images by means of the *adaptivehisteq* MATLAB function.

Finally, a median filter of size 11×11 pixels is applied to the image with the *medfilt2* MATLAB function to reduce the presence of impulse noise, naturally present in the X-rays.

ISBN: 978-989-95079-6-8

Once the image has been preprocessed, Canny edge detector is applied to it in order to obtain an edge map that will be used as a starting point for the FHT in the locating procedure. The Canny edge detector is one of the most widely used algorithms among all the features detectors. It was proposed by Canny in [7] in 1986. This method consists of several stages: first of all noise reduction is performed by means of a Gaussian filter; then it computes the gradient in every point of the image and discard those ones whose value is under a certain threshold; next, the pixels whose gradient magnitude is not maximum in the gradient direction are suppressed; finally a hysteresis thresholding is performed by using two thresholds to suppress spurious noisy pixels.

We have used the implementation available in MATLAB by means of the $edge$ function, and we have used the default parameters of the function.

In order to apply the FHT we need to identify the parameters that define the considered curve and the equation that relates them. A circular arc is defined by five parameters: the two coordinates of its centre, its radius, its orientation and its angular size. We could use the generalised Hough transform proposed in [5], but this would imply that we would have to use a five-dimensional accumulator. By doing so, the memory and computational requirements of the algorithm would rise enormously. Nevertheless, if we can fix one or more parameters, the requirements will remain reasonable.

In this case, after studying the characteristics of the iliac crests showed in the X-rays, we can assume that the radius of the circular arcs we use to represent them has to be between fifteen and forty pixels. We can also assume that an appropriate orientation for the circular arc could be 0 rad and a suitable angular size for it could be π rad. These assumptions we have made permit us to simplify the original problem reducing its computational requirements, which is one of the major drawbacks of the Hough transform. Instead of using a five-dimensional accumulator we can use a less computational expensive three-dimensional array.

In this work we have used a parametric representation of the circular arc, which allow us to use the parametric equation of the circle by restricting the permitted values for the parametric coordinate between 0 rad and π rad, instead of between 0 rad and 2π rad that we would use if we were searching for a whole circle. Moreover, the parametric representation facilitates the computation of the positions of the accumulator that have to be incremented. Thus, the equation used to describe the circular arc is

$$\begin{aligned} x(t) &= R \cdot \cos(t) + a \\ y(t) &= R \cdot \sin(t) + b \\ t &\in [0, \pi] \end{aligned} \tag{3}$$

where t is the parametric coordinate, R is the radius of the circular arc and (a, b) are the coordinates of the centre of the arc.

Next, we introduce the formal description of our implementation of the FHT:

1. Set the initial values of the accumulator A to 0.
2. For each edge point (x, y), do:
   (a) For each radius R between 15 and 40 pixels, do:
      i. For every $t \in [0, \pi]$, do:
         A. $a = x - R \cdot \cos(t)$
         B. $b = y - R \cdot \sin(t)$
         C. $A[a, b, R] = A[a, b, R] + 1$
3. Convolve the accumulator A with the smoothing function (1).

At this point we can apply the FHT, as we have just described it, to the edge map obtained in the previous step. By doing so, we get an accumulator array where the maximum values represent the most suitable arcs for the given image. We shall use the one hundred biggest ones of those peaks.

After the use of the FHT, several constraints coming from the domain knowledge are applied to the peaks found in the accumulator array to improve the performance of the algorithm.

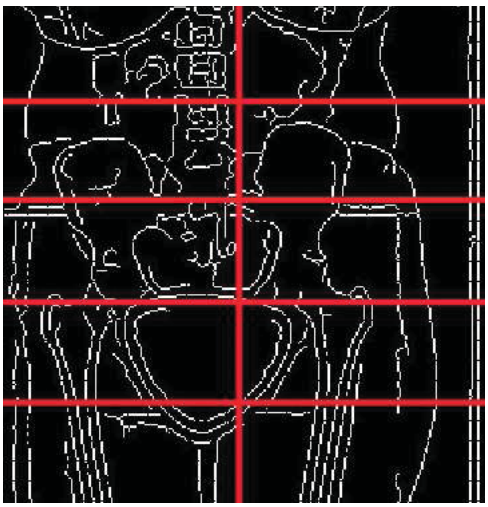


Figure 3: Canny edge map divided into regions to seek local maxima

We divide the lower part of the radiograph into two parts, and we seek the left iliac crest in the left half and the right iliac crest in the right half. Let us consider five regions of equal size within each considered division of the radiograph as we can see in Fig. 3. The search is performed separately in the left and in the right parts of the X-ray. In each region we search for the maximum local peak and record the parameters of the correspondent circular arc. Thus, we have ten local maxima, five for the left part and five for the right part, which represent ten circular arcs that are suitable locations for the iliac crests. We can make twenty five possible couples with those peaks by choosing one from the left and another one from the right.

We can assume that the locations of the left and the right iliac crests are, approximately, at the same height in the X-ray. Let us consider a tolerance factor of 15 pixels. In that case, we can discard those couples of peaks which do not satisfy this criterion. Finally, for the couples of peaks which satisfy the criterion, we add its values and we choose the one with the largest value. Thus, at the end we have a couple of peaks which corresponds to the location of the iliac crests obtained by the method. This process is called "peak selection in the accumulator" in Fig. 1.

We performed several experiments in order to choose the most appropriate values for the tolerance factor and the number of regions in terms of its performance.

ISBN: 978-989-95079-6-8

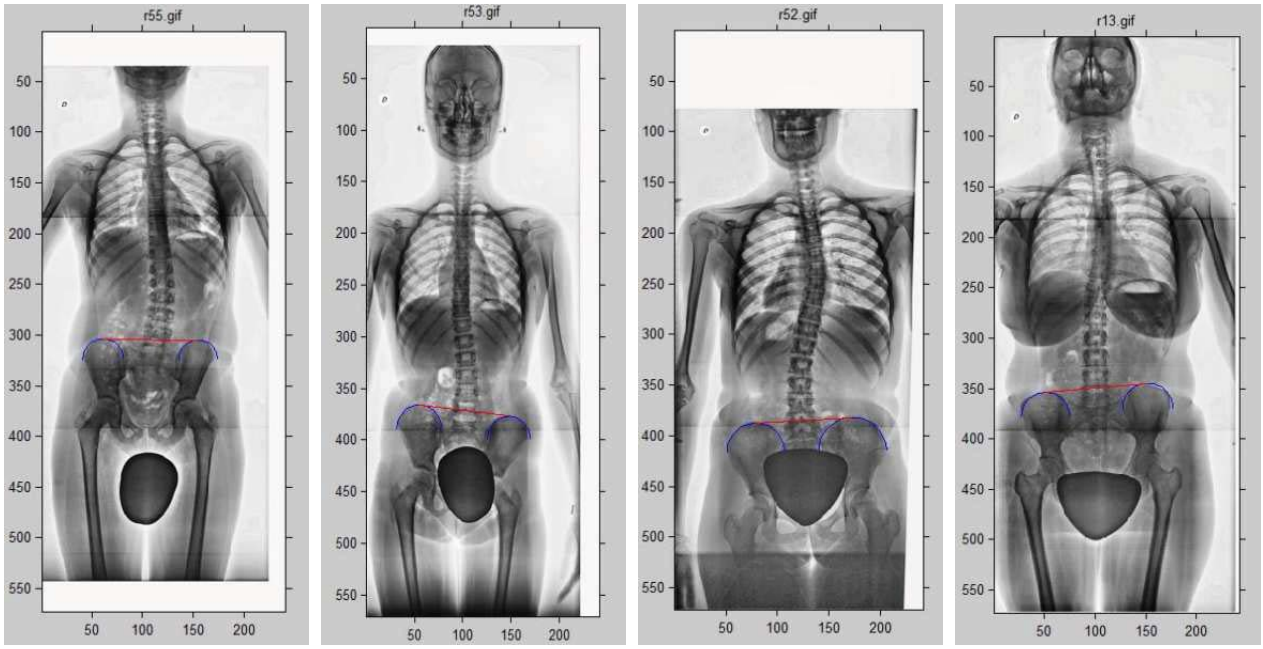
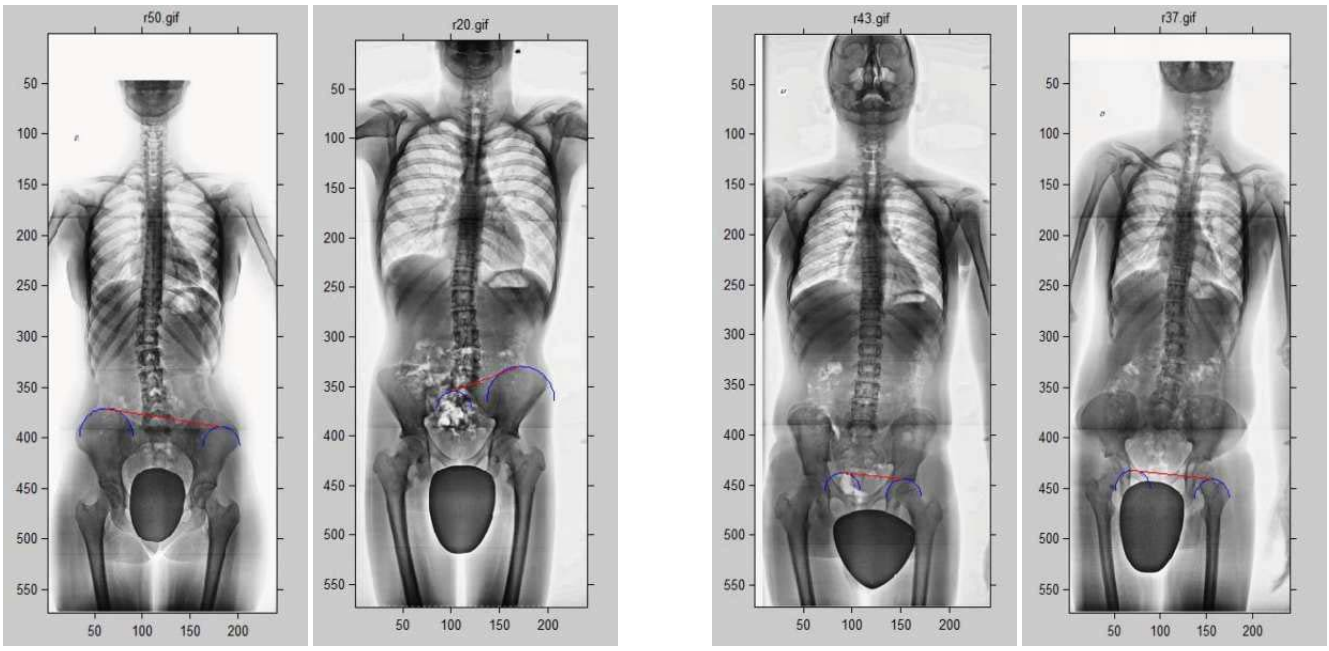


Figure 2: Some examples of the found locations of the iliac crests



(a) Only one iliac crest accurately located

(b) No iliac crest accurately located

Figure 4: Some examples which do not locate accurately the iliac crests

4 Results

In this section we discuss the results achieved by our method. We used a database with 45 X-ray images obtained from a PACS system in a local hospital. The images were in DICOM format with 5700×2400 pixels of resolution. Almost all the images were from patients suffering from scoliosis because we are focused on the study of this pathology by means of digital images. The resolution of the images was reduced to a tenth of its original one to reduce the time of processing before starting the location procedure. Experiments performed with several images suggest us that reduction in image size does not have a significant influence on the accuracy of our method.

It is important to remark that we are not interested in extract the exact shape of the iliac crest, but in the line drawn from the highest point of the left arc to the highest one of the right arc that, as we said beforehand, in an large number of cases will pass through the fourth lumbar vertebra or the interspace between the fourth and the fifth lumbar vertebrae. Therefore, accurate shape delimitation for the iliac crests is not essential.

The experiments we performed showed that in 28 out of 45 cases iliac crests were properly located, while in 3 images

ISBN: 978-989-95079-6-8

the method found the correct location for one iliac crest but failed to locate the another one. In Fig. 2 we can see some of the satisfactorily solved cases after performing the experiments. Moreover, Fig. 4 shows some cases that are not properly solved by our method.

The most important reasons that make our method fails in those cases are two: on the one hand, the presence of distortions in the area of the radiograph were the iliac crest is located, due to the position of the intestines which partially occlude it; on the other hand, the shape of some iliac crests is not completely round and this causes that our method fails because other rounder objects in the image, such as the femur heads, are located. In those cases other constraints should be included in our algorithm.

We also ran the same experiments using the method described here but using the standard HT instead of the FHT. Results obtained were satisfactory in 17 out of 45 images, whereas they were satisfactory for only one iliac crest in 14 out of 45. It is clear then that FHT can achieve better results in those images with a great presence of noise or in which the shapes sought are not clear.

5 Concluding remarks and future work

In this paper we have proposed a method for locating the iliac crests in X-ray images based on the fuzzy Hough transform. We can use the found location of the iliac crests to identify the fourth lumbar vertebra by drawing a line between the two crests and, that will be useful to initialise another algorithm designed to segment the spine. We are still working on improving accuracy and success rate of our method, as well as its efficiency, by incorporating additional constraints into the method.

It is important to remark that results achieved by means of FHT are clearly better than those obtained by standard HT. As shape of iliac crests is largely variable, it is difficult to find them by using a rigid template as HT does. For that reason FHT, which relaxes shape constraints by means of fuzzy logic, is more suitable in this problem.

Acknowledgement

This work has been partially supported by the Spanish "Ministerio de Ciencia y Tecnologa" (MCYT) under grants TIN2006-07262, and the "Consejería de Innovación, Ciencia y Empresa de Andalucía" (Spain) under research project P06-TIC-01570.

References

[1] J.R. Cobb. Outline for the study of scoliosis. *American Academy of Orthopaedic Surgeons, Instructional Course Lectures*, 5:261–275, 1948.

[2] C.A. Render. The reproducibility of the iliac crest as a marker of lumbar spine level. *Anaesthesia*, 51(11):1070–1071, 1996.

[3] P.V. Hough. Methods and means for recognizing complex pattern, 1962. US Patent 3,069,654.

[4] R.O. Duda and P.E. Hart. Use of the Hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15(1):11–15, 1972.

[5] D.H. Ballard. Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2):111–122, 1981.

[6] J.H. Han, L.T. Kóczy, and T. Poston. Fuzzy hough transform. *Pattern Recognition Letters*, 15(7):649–658, 1994.

[7] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

ISBN: 978-989-95079-6-8

Membership dependent stability analysis of TS fuzzy controlled systems using coupling attenuation

Jean-Yves Dieulot[1] Nesrine Elfelly[2]
[1]LAGIS UMR CNRS 8146, University of Lille
[2]LAGIS UMR CNRS 8146, Ecole Centrale de Lille
Cité Scientifique, 59650 Villeneuve d'Ascq, France
Email: dieulot@polytech-lille.fr, nesrine.elfelly@ed.univ-lille1.fr

Abstract— *The use of Linear Matrix Inequalities and Common Quadratic Lyapunov Functions is a powerful and commonplace tool for Takagi-Sugeno fuzzy controlled system analysis and synthesis. However, in practice, few practical and performing results are available when the subsystems exhibit different input matrices, because of the strong coupling between the subsystems/subcontrollers. In this paper, this coupling is demonstrated and a method is proposed which allows to synthesize, for a number of subsystems higher than 2, the local gains of a Parallel Distributed Controller. It is shown that the controller gains depend on the values of the input matrices and of the membership functions, and are thus able to relax classical stability conditions by embedding information on the fuzzy premises.*

Keywords— Fuzzy Control, Takagi-Sugeno Fuzzy Systems, Stability, Parallel Distributed Control.

1 Introduction

T-S fuzzy model is a fuzzy dynamic model which embeds a set of fuzzy rules in order to describe a global nonlinear system as a set of local linear models, these being smoothly connected by fuzzy membership functions [1]. As T-S fuzzy models are equivalent to polytopic linear models, systematic approaches to stability analysis and controller design can be developed using powerful conventional (linear and nonlinear) control theory with the help of optimization techniques such as Linear Matrix Inequalities solvers [2].

Control of T-S fuzzy system is most of the time achieved by the so-called Parallel Distributed Compensation method (PDC) for which the fuzzy controller shares the same fuzzy sets with the fuzzy model in the premise parts, and, thus, each control rule is distributively designed for the corresponding rule of a T-S fuzzy model. Since the consequent parts of T-S fuzzy models are described by linear state equations, linear control theory and corresponding stability tools can be used to design the fuzzy control local gains [3].

Among the mathematical tools used for assessing the stability of a T-S fuzzy system (with or without PDC control), the use of a Common Quadratic Lyapunov Function (CQLF) for all subsystems is the most popular, because it allows to derive easily explicit conditions [2] [4] [5]. It has been noted that Common Quadratic Lyapunov functions tend to be conservative, and even worse, might not exist for many complex highly nonlinear systems. Several works have allowed to relax stability conditions [4] [6] [7], and as an alternative, such tools as piecewise quadratic Lyapunov functions or fuzzy Lyapunov functions have been proposed (e.g. [8]). The drawbacks of these methods lie in the difficulty of synthesizing control gains - and not only analyze closed-loop stability with fixed controllers -, and that the results are quite independent of the membership functions, leading sometimes to poor overall performances. However, extended results have been obtained when considering additional bounds or shapes on membership functions considering linear consequents and PDC [9] [10] or nonlinear / non PDC controllers [11] [12] [13].

In short, there still remains a need for designing further tools to synthesize PDC control for T-S fuzzy systems, which appears to be very difficult when the input matrices of the systems of the fuzzy consequents are not identical or not proportional. In this case, the closed-loop system is not a parallel distribution of the individual closed-loop subsystems, because additional coupling terms appear, as the control gain designed for one subsystem is distributed over the remaining subsystems. Whereas most papers embed the general case, the examples are nearly always proposed for proportional or identical input matrices, or for a very limited number of input matrices.

When input matrices are different, it is not only necessary that all subsystems be stable, but also that any subsystem should be at least stable (or performing) under every other local controller, and, moreover, that all these closed-loop controlled subsystems should share a CQLF. A solution to this problem consists of solving a Riccati equation which has a positive definite solution that corresponds to a Lyapunov function, using some results from linear control theory [14]. In this case, the set of coupling terms is represented by a product of matrices involving a single uncertain matrix with a norm smaller than one, leading to a global Riccati equation [2].
Whereas this method is powerful and depends on the amplitude and variations of the membership functions, finding a global bounding matrix for the coupling terms is often not easy to work out, because these terms depend on the control gains themselves; in the case of control synthesis, when the gains are a priori unknown, this task can be impossible. The cancelation of such coupling terms has been tackled only for large-scale systems [15]. To sum up, practical and membership function-dependent of PDC control synthesis has not yet be performed when subsystems exhibit different input matrices as the explicit cancelation of coupling terms has not been proposed.

In this paper, the closed-loop T-S system under PDC control is rewritten in such a way that the cross-coupled terms are clearly

ISBN: 978-989-95079-6-8

seen as a weighted sum of the pairwise products involving the difference of two input matrices and their corresponding control gains. This allows, first, to determine easily a Riccati equation for each subsystem, which does not depend on the other subsystems control gains, and that takes the variation of fuzzy membership functions into account. Moreover, adding a little more conservativeness, allowing the control gains to be bounded, turns the stability conditions into a set of very simple Lyapunov equations, which might be a good help for control synthesis. Finally, it is shown that, when the number of subsystems is sufficient, the cross-coupled terms can be canceled by proposing nonlinear control gains.

2 Takagi-Sugeno fuzzy model and controller

2.1 Closed-loop model decomposition

The fuzzy model proposed by Takagi and Sugeno consists of a set of r fuzzy IF...THEN rules for which the consequents are linear models:

Plant Rule R_i: IF z_1 IS M_{i1} AND $\cdots$ AND z_g IS M_{ig} THEN $\dot{x} = A_i x + B_i u$;
where $x(t)$, $u(t)$ are state and input vectors, $z_i(t)$, M_{ij} are the premise variables and corresponding fuzzy rules, the final output of the fuzzy system being inferred as follows:

$$\dot{x} = \sum_{i=1}^{r} \mu_i \big(A_i x + B_i u\big), \tag{1}$$

where $\mu_i = \frac{\omega_i}{\sum_{i=1}^{r} \omega_i}$, ω_i is the grade of membership function of rule R_i.

For every subsystem S_i, a local controller can be defined as $u = K_i x$, where K_i is a control gain (possibly nonlinear). It is intuitive to build a fuzzy controller using a set of rules which share the same premises as the fuzzy model, hence distributing the local controllers within the global controllers according to their systems' weights.

Controller C_i: IF z_1 IS M_{i1} AND $\cdots$ AND z_g IS M_{ig} THEN $u = K_i x$, yielding:

$$u = \sum_{i=1}^{r} \mu_i K_i x. \tag{2}$$

Lemma 2.1. *Let the system $\dot{x} = \sum_{i=1}^{r} \mu_i (A_i x + B_i u)$ with PDC control $u = \sum_{i=1}^{r} \mu_i K_i x$ such that $A_i + B_i K_i = G_i$ and $\sum_{i=1}^{r} \mu_i \leq 1$, $\mu_i \geq 0$. The closed-loop system is:*

$$\dot{x} = \left(\sum_{i=1}^{r}\sum_{j=1}^{r} \mu_i \mu_j G_i + \sum_{i=1}^{r} A_i \Big(1 - \sum_{j=1}^{r} \mu_j\Big) + \sum_{i,j=1,j\neq i}^{r} \mu_i \mu_j (B_i - B_j)(K_j - K_i) \right) x. \tag{3}$$

Proof. One has

$$\dot{x} = \sum_{i=1}^{r} \mu_i \left(A_i x + B_i \sum_{j=1}^{r} \mu_j K_j x \right)$$
$$= \sum_{i=1}^{r} \mu_i \left(A_i + \mu_i B_i K_i + B_i \sum_{j=1,j\neq i}^{r} \mu_j K_j \right) x.$$
$$\dot{x} = \sum_{i=1}^{r} \left(\mu_i^2 G_i + \mu_i A_i (1 - \mu_i) + \mu_i B_i \sum_{j=1,j\neq i}^{r} \mu_j K_j \right) x.$$

Moreover,

$$\sum_{i=1}^{r} \mu_i B_i \sum_{j=1,j\neq i}^{r} \mu_j K_j = \sum_{i=1}^{r} \sum_{j=1,j\neq i}^{r} \mu_i \mu_j (G_i - A_i) + \sum_{i=1}^{r} \mu_i B_i \sum_{j=1,j\neq i}^{r} \mu_j K_j - \sum_{i=1}^{r} \sum_{j=1,j\neq i}^{r} \mu_i \mu_j B_i K_i.$$

In this equation, one can rearrange the two last sums into a sum of pairwise terms:

$$\sum_{i,j=1,j\neq i}^{r} \mu_j B_j \mu_i K_i + \mu_i B_i \mu_j K_j - \mu_i \mu_j B_i K_i - \mu_j \mu_i B_j K_j = \sum_{i,j=1,j\neq i}^{r} \mu_i \mu_j (B_i - B_j)(K_j - K_i).$$

Hence,

$$\sum_{i=1}^{r} \mu_i B_i \sum_{j=1,j\neq i}^{r} \mu_j K_j = \sum_{i=1}^{r} \sum_{j=1,j\neq i}^{r} \mu_i \mu_j (G_i - A_i) + \sum_{i,j=1,j\neq i}^{r} \mu_i \mu_j (B_i - B_j)(K_j - K_i).$$

One has now:

$$\dot{x} = \left(\sum_{i=1}^{r} \Big(\mu_i^2 G_i + \mu_i A_i (1 - \mu_i) \Big) + \sum_{i=1}^{r} \sum_{j=1,j\neq i}^{r} \mu_i \mu_j (G_i - A_i) + \sum_{i,j=1,j\neq i}^{r} \mu_i \mu_j (B_i - B_j)(K_j - K_i) \right) x.$$

As $\sum_{i=1}^{r} \mu_i^2 G_i + \sum_{i=1}^{r} \sum_{j=1,j\neq i}^{r} \mu_i \mu_j G_i = \sum_{i=1}^{r} \sum_{j=1}^{r} \mu_i \mu_j G_i$, and

$$\sum_{i=1}^{r} \mu_i A_i (1 - \mu_i) - \sum_{i=1}^{r} \sum_{j=1,j\neq i}^{r} \mu_i \mu_j A_i = \sum_{i=1}^{r} \mu_i A_i \Big(1 - \sum_{j=1}^{r} \mu_j\Big).$$

We demonstrate the final result:

$$\dot{x} = \left(\sum_{i=1}^{r}\sum_{j=1}^{r} \mu_i \mu_j G_i + \sum_{i=1}^{r} \mu_i A_i \Big(1 - \sum_{j=1}^{r} \mu_j\Big) + \sum_{i,j=1,j\neq i}^{r} \mu_i \mu_j (B_i - B_j)(K_j - K_i) \right) x.$$

□

ISBN: 978-989-95079-6-8

Remark: When $\sum_{i=1}^{r} \mu_i = 1$, formula (3) is reduced to:

$$\dot{x} = \left(\sum_{i=1}^{r} \mu_i G_i + \sum_{i,j=1, j\neq i}^{r} \mu_i \mu_j (B_i - B_j)(K_j - K_i) \right) x.$$

2.2 Global Stability Verification

It should be recalled that a Linear Matrix Inequality (LMI) is a set of equations which can be put under the form $F_O + F_1 x_1 + \cdots F_n x_n \prec 0$, where $F_i^T = F_i$. A matrix F is said negative definite, which is noted $F \prec 0$ if $\forall x \neq 0, x^T F x < 0$.
A Bilinear Matrix Inequality involves bilinear terms and cannot be solved in a straightformard way as LMIs are.

Theorem 2.2. *The system $\dot{x} = \sum_{i=1}^{r} \mu_i (A_i x + B_i u)$, under PDC control $u = \sum_{i=1}^{r} \mu_i K_i x$, such that $A_i + B_i K_i = G_i$ and $A_i + B_i K_j = G_{ij}$, is stable if there exists a common positive definite matrix P such that:*

$$\begin{array}{l} \forall\, i = 1 \cdots r, P G_i + G_i^T P \prec 0, \\ \forall\, i < j, P\left(G_{ij} + G_{ji}\right) + \left(G_{ij} + G_{ji}\right)^T P \prec 0. \end{array} \quad (4)$$

Remark: Theorem (2.2) allows the determination of both the Lyapunov matrix and the controller gain, using a change of variable $N_i = K_i P^{-1}$, when being replaced in the stability conditions, leads to a set of LMIs in N_i and in P, the PDC controller being provided by $K_i = N_i P$. The conservativeness of the result comes from several reasons: of course, the global system can be stable without sharing a Common Quadratic Lyapunov function. The conditions are independent of the membership functions, and those roughly mean that system *i* is considered to behave "well" with the corresponding local controller $u = K_i x$, but also with any other local controller $u = K_j x, j \neq i$. Of course, it cannot be expected that a system with a controller designed for another plant has a "good" behavior, and hence, the PDC controller is designed according to the "worst" case among the pairs {Plant *i*, Controller *j*}.

3 Coupling terms attenuation

Theorem 3.1. *[14] First, we consider the linear uncertain system for which $\dot{x} = A + \sum_{i=1}^{r} D_i \delta_i E_i, \|\delta_i\| \leq 1$, and the elements of δ are Lebesgue measurable. Then the positive-definite matrix P is a common Lyapunov matrix for this system if there exists r positive scalars η_i such that:*

$$PA + A^T P + \sum_{i=1}^{r} \eta_i P D_i D_i^T P + \eta_i^- 1 E_i^T E_i \prec 0,$$

or simply, if $\forall i, \eta_i = 1$

$$PA + A^T P + \sum_{i=1}^{r} P D_i D_i^T P + E_i^T E_i \prec 0. \quad (5)$$

Remark: This Theorem was applied first by Tanaka [2] and then by numerous authors to the whole coupling term. Whereas this method provides for a rather non-conservative solution, it is clear that finding individual uncertain matrices might be a tedious task, because the rate of variation and thus the bounds on the uncertain matrix depend on the control gains themselves. It can thus be applied to analyze an existing solution (the gains are fixed) but not for gain synthesis considering models/controllers coupling. The following Theorem proposes a different application of this method to every individual component of the coupling terms.

Theorem 3.2. *Consider the system $\dot{x} = \sum_{i=1}^{r} \mu_i (A_i x + B_i u)$, under PDC controller:*

$$\dot{x} = \left(\sum_{i=1}^{r} \mu_i G_i + \sum_{i,j=1, j\neq i}^{r} \mu_i \mu_j (B_i - B_j)(K_j - K_i) \right) x.$$

Let us suppose that: $\forall\, i$, there exists b_i such that:

$$\sum_{j/B_i \neq B_j} \mu_j (B_j - B_i) = b_i \delta_i, \text{ where } \|\delta_i\| \leq 1.$$

The matrices δ_i thus depend on membership functions and other input matrices μ_j and B_j; as μ_j may vary with time, δ_i is a matrix which may vary with time or with the state space x. The corresponding norm is the Euclidean norm.
The closed-loop system is quadratically stable if:

$$\forall\, i = 1 \cdots r, P G_i + G_i^T P + P b_i b_i^T P + K_i^T K_i \prec 0. \quad (6)$$

This can be turned into:

$$\forall\, i = 1 \cdots r, \begin{pmatrix} P G_i + G_i^T P & P b_i & K_i^T \\ b_i^T P & -I & 0 \\ K_i & 0 & -I \end{pmatrix} \prec 0. \quad (7)$$

Proof.

$$\begin{aligned} \dot{x} &= \sum_{i=1}^{r} \left(\mu_i G_i + \sum_{i,j=1, j\neq i}^{r} \mu_i \mu_j (B_i - B_j)(K_j - K_i) \right) x \\ &= \sum_{i=1}^{r} \mu_i \sum_{j=1, j\neq i}^{r} G_i + \mu_j (B_i - B_j) K_i. \end{aligned}$$

One has now: $G_i + \mu_j (B_i - B_j) K_i = G_i + b_i \delta_i K_i$, and one can apply the Theorem (3.1). □

Remark: Uncertain matrices δ_i do not depend anymore on the control gains but only on input matrices and membership functions which are known a priori. Their determination is thus quite easy and the membership functions are indeed embedded in the control synthesis. Note also that the corresponding *i* Riccati equations in (6) are decoupled, i.e. the *i*th equation only depends on the *i*th control gain, the influence of the other subsystems being lumped into the matrix $b_i \delta_i$. However, one can realize that equation (7) is a BMI in P and K_i, and thus less tractable for control gain synthesis, which motivates the following Corollary.

Corollary 3.3. *Let us suppose that:*

$$\forall\, i = 1 \cdots r,\ \exists\, Q_i \succ 0,\ K_i^T K_i - Q_i \prec 0.$$

Then, condition (7) can be expressed as:

$$\exists\, P \prec 0, \forall\, i = 1 \cdots r, P G_i + G_i^T P + Q_i' \prec 0, \quad (8)$$

ISBN: 978-989-95079-6-8

where $Q'_i = Pb_ib_i^TP + Q_i$, *with* $K_i^TK_i - Q_i \prec 0$,
which can be turned into:

$$\forall\, i = 1 \cdots r, \begin{cases} \begin{pmatrix} PG_i + G_i^TP + Q_i & Pb_i \\ b_i^TP & -I \end{pmatrix} \prec 0, \\ K_i^TK_i - Q_i \prec 0. \end{cases}$$

The Corollary simply reduces the search for a common Lyapunov matrix to a series of r Lyapunov equations and thus r LMIs. This is really a drastic improvement to other methods because, now, control gains can nearly be selected independently without the need to taking care of coupling terms, at the expense of a priori gain limitation.

4 Example

Let us take the 3 following systems:

$$A_1 = \begin{pmatrix} -1 & 2 \\ 0 & -2 \end{pmatrix}, B_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix};$$
$$A_2 = \begin{pmatrix} -1 & -1 \\ 1 & 0 \end{pmatrix}, B_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix};$$
$$A_3 = \begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix}, B_3 = \begin{pmatrix} 1 \\ 0 \end{pmatrix};$$

along with local gains:

$K_1 = (\ 2 \quad 1\ ), K_2 = (\ -2 \quad 1\ ), K_3 = (\ 2 \quad 0\ ).$

The premisses corresponding to systems 1, 2 and 3 are:
$\mu_1 = z$, $\mu_2 = 1 - z$ and $\mu_3 = z$ where $z \in [-1 \cdots 1]$.

For every subsystem i, it is quite easy to compute the matrices b_i such that $\sum_{B_i \neq B_j} \mu_i(B_j - B_i) = b_i\delta_i$, since the upper bound depends on the fuzzy variable z.
One finds:
${b_1}^T = (\ 1 \quad 0.25\ )$, ${b_2}^T = (\ 0.75 \quad 0.25\ )$,
${b_3}^T = (\ 1 \quad 1\ )$.

The application of Theorem (3.2) allows to find a common positive definite matrix $P = \begin{pmatrix} 1.28 & -0.37 \\ -0.37 & 0.87 \end{pmatrix}$ whereas it is impossible to find one by the classical method; it is easy to check that the gain K_2 is unable to stabilize matrix A_1 and the converse for K_1 and A_2. It is quite interesting to note that the result is quite tied to the value of the matrices b_i.
When all other variables keep the same values, but $b_2^T = (\ 1 \quad 1\ )$, then the Theorem (3.2) is no more applicable because a positive definite CQLF cannot be found. Thus, Theorem (3.2) is able to relax stability conditions, depending strongly on the membership functions and input matrices values.

5 Conclusion

In this paper, the stability of a Takagi Sugeno fuzzy system under the Parallel Distributed Compensation controller has been studied. This control strategy allocates the same weight to a local controller that the one in the fuzzy combination of the local submodels.
The influence of the coupling between any local subsystem and any local controller (different from the corresponding local controller designed from the local subsystem considered) in the closed-loop response has been highlighted, and it has been shown to be effective when the input matrices of the subsystems are different. It has been subsequently shown that a controller synthesis based on an analysis of each local subsystem controlled by any local compensator, would lead to conservative results.

A new approach has been proposed which, for every local subsystem, takes the coupling term coming from other subsystems into account, and proposes to choose the gain in order to cope with the effect of the coupling terms. This strategy allows to minimize the number of Linear Matrix Inequalities to be solved for controller synthesis and to take into account the shape of the membership functions.

References

[1] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modelling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, 15:116–132, 1985.

[2] K. Tanaka, T. Ikeda, and H. O. Wang. Robust stabilization of a class of uncertain nonlinear systems via fuzzy control: Quadratic stabilizability, H∞ control theory, and linear matrix inequalities. *IEEE Transactions on Fuzzy Systems*, 4:1–13, 1996.

[3] P. Gahinet, A. Nemirovski, A. Laub, and M. Chilali. *The LMI Control Toolbox*. The Mathworks, Inc., Natick, MA, 1995.

[4] K. Tanaka, T. Ikeda, and H. O. Wang. Fuzzy regulators and fuzzy observers: Relaxed stability conditions and LMI-based designs. *IEEE Transactions on Fuzzy Systems*, 6:250–265, 1998.

[5] J. Yoneyama. Robust H∞ control analysis and synthesis for Takagi-Sugeno general uncertain fuzzy systems. *Fuzzy Sets and Systems*, 157:2205–2223, 2006.

[6] C.H. Fang, Y.S. Liu, S.W. Kau, L. Hong, and C.H. Lee. A new LMI-based approach to relaxed quadratic stabilization of T-S fuzzy control systems. *IEEE Transactions on Fuzzy Systems*, 14:286–297, 2006.

[7] E. Kim and H. Lee. New approaches to relaxed quadratic stability condition of fuzzy control systems. *IEEE Transactions on Fuzzy Systems*, 8:523–534, 2000.

[8] G. Feng, C. L. Chen, D. Sun, and X. P. Guan. H∞ controller synthesis of fuzzy dynamic systems based on piecewise Lyapunov functions and bilinear matrix inequalities. *IEEE Transactions on Fuzzy Systems*, 13:94–103, 2005.

[9] A. Sala and C. Ariño. Relaxed stability and performance LMI conditions for Takagi-Sugeno fuzzy systems with polynomial constraints on membership function shapes. *IEEE Transactions on Fuzzy Systems*, 16:1328–1336, 2008.

[10] A. Sala and C. Ariño. Relaxed stability and performance conditions for Takagi-Sugeno fuzzy systems with knowledge on membership function overlap. *IEEE Transactions On Systems, Man, And Cybernetics-Part B*, 37:727–732, 2007.

[11] M. Abbaszadeh and H. J. Marquez. LMI optimization approach to robust H∞ observer design and static output feedback stabilization for discrete-time nonlinear uncertain systems. *IEEE Int. J. Robust Nonlinear Control*, 19:313–340, 2009.

[12] H.K. Lam and F.H.F. Leung. Stability analysis of fuzzy control systems subject to uncertain grades of membership. *IEEE Transactions on Systems, Man and Cybernetics - Part B*, 35:1322–1325, 2005.

ISBN: 978-989-95079-6-8

[13] H.K. Lam and F.H.F. Leung. LMI-based stability and performance conditions for continuous-time nonlinear systems in Takagi-Sugeno's form. *IEEE Transactions on Systems, Man and Cybernetics - Part B*, 37:1396–1406, 2007.

[14] P.P. Khargonekar, I.R. Petersen, and K. Zhou. Robust stabilization of uncertain linear systems: Relations between quadratic stabilizability and H1 control theory. *IEEE Transactions on Automatic Control*, 37:356–61, 1990.

[15] W.J. Wang and W.W. Lin. PDC Synthesis for T-S fuzzy large-scale systems. *IEEE Transactions on Fuzzy Systems*, 12:309–315, 2004.

ISBN: 978-989-95079-6-8

Characterization of L-fuzzy semi-filters and semi-ideals

Jorge Jiménez, Susana Montes[1], Branimir Šešelja, Andreja Tepavčević[2]
1. University of Oviedo, Spain
2. University of Novi Sad, Serbia
Email: {meana,montes}@uniovi.es, {seselja,andreja}@im.ns.ac.yu

Abstract— *Lattice-valued up-sets and down-sets of a poset are investigated under the cutworthy approach. It is proved that the collections of all lattice valued up-sets and down-sets of a given poset are complete lattices under fuzzy inclusion. Properties of these are investigated. For a given collection of crisp up-sets of a poset, necessary and sufficient conditions are given under which the family represents a collection of cuts of a lattice valued up-set. The corresponding conditions are also obtained for the analogous case of down-sets.*

Keywords— L-valued sets (L-fuzzy sets), L-valued up-sets, L-valued down-sets.

1 Introduction

Lattice valued sets and structures have been widely studied from Goguen's first paper [1] on this topic, and since Negoita and Ralescu published their book [2] back in 1975. They appear when the membership grades can be represented by elements of an arbitrary partially ordered set L, instead of just by numbers in the unit interval $[0, 1]$. Although a lot of studies have been done for fuzzy order structures (see, e.g., [3, 4, 5, 6] among many others), the relaxation of the requirements allow us to use lattice-valued sets in a wider context, since they can be more appropriate to model natural problems. It turns out that recently this approach to fuzziness attracts more and more interest, in particular when a structure has a residuated or similar lattice as a co-domain. Thus, in the last decade, lattice-valued mathematics have undergone a significant development. This have created bridges between them and algebraic theories, quantales and order-theoretic structures, various subdisciplines of topology and theoretical computer science (see, e.g., [7, 8, 9, 10, 11, 12, 13, 14]).

Due to the importance of up-sets (semi-filters) and down-sets (semi-ideals) in the classical representation theory, the main purpose of this paper is to study the generalization of these concepts for L-valued sets. Fuzzy semi-ideals (down-sets) and semi-filters (up-sets) were investigated in [15], refereing to a particular T-norm, therefore not in the framework of our cutworthy approach. Let us remember that a property generalized to a fuzzy structure is said to be cutworthy if the corresponding crisp property is preserved by the cut-structures. It is known that many properties of relations or algebraic structures are cutworthy if the co-domain of the investigated fuzzy structure is a complete lattice (e.g., properties of equivalence or ordering relations, then subalgebras of algebraic structures etc.). On the other hand, if this co-domain is a lattice with some additional operations like residuation, product etc., then this cutwortheness is not fulfilled. Since our aim is to connect a lattice valued structure with its cut-structures, the co-domain of all fuzzy structures here is a complete lattice, without additional operations.

In the present investigation, we start with a finite poset X and a complete lattice L. Then we introduce L-valued up-sets (semi-filters) and L-valued down-sets (semi-ideals) as isotone and anti-isotone mappings from X to L, respectively. It turns out that the cut-sets of these are precisely crisp up-sets and down-sets on X. Next we investigate collections of all L-valued up-sets and all L-valued down-sets on X. We prove that under the order induced by the one in L, these collections are complete lattices, fulfilling all identities satisfied by the lattice L. Moreover, we analyze the problem of existence of a representation of collections of crisp up-set on X by collections of cut-set of an L-valued up-set on X. We give necessary and sufficient conditions (fulfilled by the lattice L) under which such representation exists.

Our main motivation in this work was to present a mathematical characterization of these algebraic objets, namely semi-filters and semi-ideals, in the fuzzy environment. The justification to consider lattice valued fuzzy sets has been widely explained in the literature, since lattices are more richer structure and we can obtain non-comparable values of fuzzy sets. They can be applied, for instance, in image processing.

The work is structured in 4 sections. In Section 2 we recall the most relevant concepts concerning classical order sets, up-sets and down-sets, as well as lattice-valued sets. In Section 3 we introduce L-valued up-sets and down-sets and we relate these concepts. The first part of this section is devoted to a wide study of the properties and characterizations for L-valued up-sets. In the second part of the section, analogue problems are investigated for L-valued down-sets. In Section 4 we briefly address some conclusions and future work.

2 Preliminaries

In this section, some well-known definitions and preliminary results are recalled. They will be necessary in order to understand the new concepts introduced in this paper and their associated studies.

2.1 Order, up-sets and down-sets

Some necessary notions from the classical order theory are listed in the sequel, together with relevant properties. For more comprehensive presentation, see e.g., books [16, 17].

A **poset** is a nonempty set X equipped with an ordering relation $\leq$. A poset is usually denoted as an ordered pair $(X, \leq)$, or simply by the underlying set X. A sub-poset of $(X, \leq)$ is a poset on a subset Y of X in which the order is the one restricted from X, and usually denoted in the same way ($\leq$). An **up-set** (**semi-filter**) on a poset X is any subposet U, satisfying the following: for $x \in U, y \in X$, $x \leq y$

ISBN: 978-989-95079-6-8

implies $y \in U$. Dually, an **down-set** (**semi-ideal**) on X is any sub-poset D, satisfying: for $x \in D, y \in X$, $y \leq x$ implies $y \in D$. A **lattice** is a poset L in which for each pair of elements x, y there is a greatest lower bound (glb, infimum, meet) and a least upper bound (lub, supremum, join), denoted respectively by $x \wedge y$ and $x \vee y$. These are binary operations on L. A non-empty poset L is said to be a **complete** lattice if infimum and supremum exist for each subset of L. Complete lattice possesses the **top** (1) and the **bottom** element (0). On the other hand, a lattice L is **distributive**, if each operation is distributive with respect to the other. An example of distributive lattices is presented in the following lemma.

Lemma 1 *The collection of all up-sets (down-sets) of a poset X is a distributive lattice under inclusion.*

Given a lattice L, an element a in L is **completely meet-irreducible** if $a \neq 1$ and for every family $\{x_i \mid i \in I\}$ of elements from L, from $a = \bigwedge_{i \in I} x_i$ it follows that $a = x_i$ for some $i \in I$. When L is a lattice of finite length, a is just said to be **meet-irreducible**.

A **closure operator** on a lattice L is a function $C : L \to L$ such that, for all $p, q \in L$, it fulfills the three following requirements:

- $p \leq C(p)$,
- $p \leq q \to C(p) \leq C(q)$,
- $C(C(p)) = C(p)$.

If $p = C(p)$, then p is a **closed element** under the corresponding closure operator.

Lemma 2 *Let C be a closure operator on a lattice L. Then,*

1. *The subset of all closed elements of L is closed under meets in L.*
2. *The top element of L is a closed element under C.*

Lemma 3 *Let L be a complete lattice. Then,*

1. *For any closure operator C on L, the subset of all closed elements of L is a complete lattice under the order inherited from L.*
2. *If F is a subset of L closed under arbitrary meets, then the map such that to any element p in L associates the value $\bigwedge\{q \in F \mid p \leq q\}$ is a closure operator on L.*

By the definition of the supremum and the infimum in a complete lattice, we also have that $\bigwedge \emptyset = 1$ and $\bigvee \emptyset = 0$.

Lemma 4 *Let C be a closure operator on a complete lattice L. If we consider the relation $\sim$ on L defined by $x \sim y$ iff $C(x) = C(y)$, then*

1. *$\sim$ is an equivalence relation on L.*
2. *Each such equivalence class has the top element and this element is closed.*
3. *The set $L/\sim$ (denoted also by L/C) can be ordered: $[x]_\sim \leq [y]_\sim$ iff $C(x) \leq C(y)$ in L.*
4. *The poset $(L/\sim, \leq)$ is a lattice isomorphic with the poset of closed elements of L, under C.*

2.2 Lattice-valued sets

We present some notions from the theory of lattice-valued structures. More details about the relevant properties can be found in [18, 19].

Lattice-valued, **L-valued sets** or **L-fuzzy sets** are here considered to be mappings from a non-empty set X (domain) into a complete lattice L (co-domain) with the top and bottom elements 1 and 0, respectively. This concept can be seen as a generalization of the concept of fuzzy or valued set and it was introduced by Goguen (see [1]).

If $\alpha : X \to L$ is an L-fuzzy set on X then, for $p \in L$, the set

$$\alpha_p := \{x \in X \mid \alpha(x) \geq p\}$$

is called the **p-cut**, a **cut set** or simply a **cut** of α.

Example 1 *Let us consider the poset $(X, \leq)$ and the complete lattice $(L, \leq)$ represented in Figure 1.*

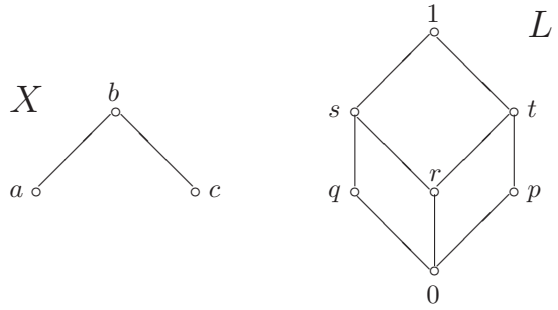


Figure 1: Hasse diagrams of X and L.

Some examples of L-valued sets are

X	a	b	c
α	r	t	p

X	a	b	c
ν	q	s	r

X	a	b	c
μ	r	0	p

X	a	b	c
λ	q	s	t

where we consider a tabular representation for the L-valued sets on X.

In the case of α, for instance, its p-cut is $\alpha_p = \{b, c\}$ and its t-cut is $\alpha_t = \{b\}$.

We can notice that the ordering in X does not have a role in this example. This ordering is used later in the paper.

Some characterizations of the collection of cuts of a lattice-valued set are presented in the next three results.

Proposition 1 *[18] Let $\mathcal{F}$ be a family of subsets of a nonempty set X which is closed under intersections and contains X. Let $\alpha : X \to \mathcal{F}$ be defined by*

$$\alpha(x) = \bigcap(p \in \mathcal{F} \mid x \in p).$$

Then, α is a $\mathcal{F}$-valued set on X with the codomain lattice $(\mathcal{F}, \supseteq)$ such that its family of p-cuts is $\mathcal{F}$ and for every $p \in \mathcal{F}$ it holds that $p = \alpha_p$.

Theorem 1 *[20] Let L be a fixed complete lattice. Necessary and sufficient conditions under which $\mathcal{F} \subseteq \mathcal{P}(X)$ is the collection of cut sets of an L-valued set with domain X are:*

ISBN: 978-989-95079-6-8

1. *$\mathcal{F}$ is closed under arbitrary intersections and contains X,*
2. *The dual poset of $\mathcal{F}$ under inclusion can be embedded into L, such that all infima and the top element are preserved under the embedding.*

Proposition 2 *[21] Let L be a lattice of finite length and let $\alpha : X \to L$ be an L-valued set. Then, all p-cuts of α are distinct if and only if all meet-irreducible elements of L belong to $\alpha(X)$.*

3 L-valued up-sets and down-sets

In classical order theory, the families of up-sets and down-sets, associated to any ordered set, play a central role in the representation theory. Once the L-valued sets are considered, a step forward would be to define and characterize up-sets and down-sets in the L-valued logic. That study is made in this section.

Classical up-sets and down-sets can be seen as sets which are "closed under going up" and "closed under going down", respectively. In fact, a subset is an up-set if and only if its characteristic function is order-preserving. Moreover, a subset is a down-set if and only if its characteristic functions is order-preserving, when the dual of the usual order is considered in $\{0, 1\}$. Thus, a natural generalization of these concepts for L-valued sets is presented in Definition 1.

Definition 1 *Let $(X, \leq)$ be a poset and let $(L, \leq)$[1] be a complete lattice in which 0 and 1 are the bottom and the top element, respectively.*

A function $\alpha : X \to L$ such that for all $x, y \in X$

$$x \leq y \implies \alpha(x) \leq \alpha(y)$$

*is an **L-valued up-set** or an **L-valued semi-filter** on X.*

Dually, a function $\mu : X \to L$ such that for all $x, y \in X$

$$x \leq y \implies \mu(y) \leq \mu(x)$$

*is an **L-valued down-set** or an **L-valued semi-ideal** on X.*

Example 2 *If we consider the L-valued sets introduced in Example 1, α and ν are L-valued up-sets, μ is an L-valued down-set and λ is neither L-valued up-set nor L-valued down-set.*

If $L = \{0, 1\}$, that is, if we are in a two-valued logic, it is immediate that any L-valued up-set is a classical up-set and the same happens for down-sets. Thus, these concepts generalize the classical ones.

The duality between these two concepts is showed in the following proposition, by means of complement operators [22].

Proposition 3 *Let $(X, \leq)$ be a poset and let $(L, \leq)$ be a complete lattice in which 0 and 1 are the bottom and the top element, respectively. If there exists a complement operator c on L, that is, a map $c : L \to L$ fulfilling*

1. *Boundary values: $c(0) = 1$ and $c(1) = 0$;*
2. *Non-increasing: $p \leq q \Rightarrow c(q) \leq c(p), \forall p, q \in L$;*
3. *Involution: $c(c(p)) = p, \forall p \in L$;*

then an L-valued set $\alpha : X \to L$ is an L-valued up-set if and only if, its c-complement α^c is an L-valued down-set, where α^c denotes the L-valued set on X defined by $\alpha^c(x) = c(\alpha(x))$.

3.1 Characterization of the family of L-valued up-sets

In this section we present a wide study of the concept of L-valued up-sets. This study is repeated later, in next section, for L-valued down-sets, due to the duality between these two classes of L-valued sets. Let us start with a characterization of the concept of L-valued up-set by means of α-cuts.

Theorem 2 *Let $(X, \leq)$ be a poset, let $(L, \leq)$ be a complete lattice and let α an L-valued set on X. The following two statements are equivalent:*

- *α is an L-valued up-set on X.*
- *The p-cut α_p of α is crisp up-set (semi-filter) on X, for any $p \in L$.*

The proof of this theorem is a immediate consequence of the definitions of p-cut and L-valued up-set.

Example 3 *Let us consider again the L-valued up-set α introduced in Examples 1 and 2. The family of its cuts is formed by the elements*

$$\alpha_0 = X, \alpha_p = \{b, c\}, \alpha_r = \{a, b\}, \alpha_t = \{b\}$$

$$\text{and } \alpha_q = \alpha_s = \alpha_1 = \emptyset.$$

As a consequence of the previous theorem, we know that all these crisp sets are up-sets. Conversely, since all the cuts are crisp up sets on X the previous theorem says that α is an L-valued up-set which is easy to check.

From now on, we denote by $\mathcal{U}_L(X)$ the subset of the L-valued power set of X formed by all L-valued up-sets on the poset X, that is,

$$\mathcal{U}_L(X) = \{\alpha : X \to L \mid \alpha \text{ is an } L\text{-valued up-set}\}.$$

This poset can be ordered naturally using the order induced by the one from the lattice L. Thus, for any $\alpha, \beta \in \mathcal{U}_L(X)$, we say that $\alpha \leq \beta$ if and only if for each $x \in X$ $\alpha(x) \leq \beta(x)$.

Moreover, meet and join exist for any subset of $\mathcal{U}_L(X)$, as it is stated in the following theorem.

Theorem 3 *The poset $(\mathcal{U}_L(X), \leq)$ is a complete lattice.*

In fact, the poset $(\mathcal{U}_L(X), \leq)$ is a complete sublattice of the lattice $(L^X, \leq)$. More specifically, it is a $\{0, 1\}$-complete sublattice with $0(x) = 0$ and $1(x) = 1$, for all $x \in X$.

Due to the definition of the order in $\mathcal{U}_L(X)$, we have the following obvious consequence.

Proposition 4 *If a lattice identity holds in L, then the same identity is satisfied in the lattice $(\mathcal{U}_L(X), \leq)$.*

[1] For simplicity, we consider the same notation for the order in X and in L. Of course these orders do not have to coincide. However, possible ambiguity is removed by the context, so that in any moment we are able to identify order represented by $\leq$.

ISBN: 978-989-95079-6-8

From Proposition 4, if L is e.g., a distributive lattice, then we also have that the lattice $(\mathcal{U}_L(X), \leq)$ is distributive.

Now, we prove a result which allows us to know when, starting from a given collection of crisp up-sets, it is possible to built an L-valued up-set.

Theorem 4 *Let $(X, \leq)$ be a poset, let $\mathcal{F} \subseteq \mathcal{P}(X)$ be a family of some up-sets of a poset X, and let $(L, \leq)$ be a complete lattice. Then, there is an L-valued up-set $\alpha : X \to L$ such that its family of cuts is equal to $\mathcal{F}$ if and only if the following two conditions hold:*

1. *$\mathcal{F}$ is closed under intersections and contains X;*
2. *there is a closure operator C on L, such that the poset $(L/C, \leq)$ is order isomorphic to $(\mathcal{F}, \supseteq)$.*

In the proof of this theorem we have to prove the equivalence among these conditions and the conditions imposed in Theorem 1.

Example 4 *Let us consider again the poset X, the complete lattice L and the L-valued sets α, ν introduced in Example 1. Let $\mathcal{F}$ be the family of up-sets of X defined by*

$$\mathcal{F} = \{X, \{a, b\}, \{b, c\}, \{b\}, \emptyset\}.$$

This family fulfills Conditions 1 and 2 in Theorem 4 and therefore, there exists an L-valued up-set whose family of cuts is equal to $\mathcal{F}$. This L-valued up-set is precisely the L-valued set α. Of course, it does not need to be unique. For instance, also ν is an L-valued up-set with $\mathcal{F}$ as its family of cuts.

In the previous theorem, we have characterized families of up-sets which are the cut-sets of an L-valued up-set. Now, we are going to characterize the L-valued up-set such that this family is formed by all the (crisp) up-sets.

Proposition 5 *Let $(X, \leq)$ be a poset, let $(L, \leq)$ be a complete lattice and let $\alpha : X \to L$ be an L-valued up-set. The following two statements are equivalent:*

- *The family of cut sets of α is formed by all the crisp up-sets of poset $(X, \leq)$.*
- *For every family $\{x_i \mid i \in I\}$ of elements from X and every $x \in X$, it holds that*

$$\alpha(x) \geq \bigwedge_{i \in I} \alpha(x_i) \Rightarrow \exists i \in I \mid x \geq x_i.$$

A main step in the proof of the previous proposition is to consider that the $\bigcup_{i \in I} {\uparrow} x_i$ is a crisp up-set of α, where $\{x_i \mid i \in I\}$ is the family of elements from X such that $\alpha(x) \geq \bigwedge_{i \in I} \alpha(x_i)$ for $x \in X$ and $x \not\geq x_i$ for all $i \in I$.

As a consequence of the previous proposition, we obtain two necessary conditions for a representation by means of all the crisp up-sets. They are presented in the two following corollaries.

Corollary 1 *Let $(X, \leq)$ be a poset, let $(L, \leq)$ be a complete lattice and let $\alpha : X \to L$ be an L-valued up-set. If the family of cut sets of α is formed by all the crisp up-sets of poset $(X, \leq)$, then α is an order-embedding.*

In the following result it is necessary to consider the subposet L^α of L associated to any L-valued set α on X and generated by taking all infima of all the subsets of $\alpha(X)$ (including infimum of the empty set - the top element of L). Thus, given an L-valued set α on X,

$$L^\alpha = \{p \in L \mid p = \wedge B \text{ with } B \subseteq \alpha(X)\}.$$

Let us note that it is immediate that L^α is a complete lattice.

Corollary 2 *Let $(X, \leq)$ be a poset, let $(L, \leq)$ be a complete lattice and let $\alpha : X \to L$ be an L-valued up-set. If the family of cut sets of α is formed by all the crisp up-sets of poset $(X, \leq)$, then $\alpha(x)$ is a completely meet-irreducible element in the lattice L^α for every $x \in X$ with $\alpha(x) \neq 1$.*

In order to illustrate Proposition 5 and Corollaries 1 and 2, we show an example of an L-valued set such that its family of cuts is formed by all the crisp up-sets of its domain.

Example 5 *Let us consider again the L-valued up-set α introduced in Example 1. In Example 4 we have obtained its family of cut sets. It is immediate that this is the family of all up-sets of X. Thus, by Corollaries 1 and 2, we know that α is an order-embedding and $\alpha(x)$ is a completely meet-irreducible element in the lattice $L^\alpha = \{0, p, r, t, 1\}$ for every $x \in X$.*

Now, we suppose that L is a lattice of finite length. Then, combining Proposition 2 and Proposition 5, we obtain the following result.

Theorem 5 *Let $(X, \leq)$ be a poset, let $(L, \leq)$ be a complete lattice of finite length and let $\alpha : X \to L$ be an L-valued up-set such that*

1. *α is an bijection to set $M(L)$ of all meet-irreducible elements of lattice L.*
2. *For every family $\{x_i \mid i \in I\}$ of elements from X, if $\alpha(x) \geq \bigwedge_{i \in I} \alpha(x_i)$ for $x \in X$, then $x \geq x_i$, for some $i \in I$.*

Then, the lattice L is isomorphic with the family of all up-sets of the poset X and therefore L is distributive.

Example 6 *Let us consider again the L-valued up-set α introduced in Example 1. Since $\alpha_s = \alpha_q = \emptyset$, it is not true that all its p-cuts are distinct (see Proposition 2) and therefore Condition (1) in the previous theorem does not hold. We can also notice that s is a meet irreducible element in lattice L and it is not a value of the function α. Therefore, we cannot use this example in order to illustrate Theorem 5.*

Let us consider a different poset $(X, \leq)$, a different complete lattice $(L, \leq)$ and a different L-valued up-set α, which are represented in Figure 2.

Given the L-valued set $\alpha : X \to L$, the p-cuts of α are all the up-sets of X:

$$\alpha_0 = \{0, a, b, c, 1\},\ \alpha_n = \{a, b, c, 1\},\ \alpha_p = \{a, b, 1\},$$

$$\alpha_q = \{a, c, 1\},\ \alpha_r = \{b, c, 1\},\ \alpha_s = \{a, 1\}\ \alpha_t = \{b, 1\},$$

$$\alpha_u = \{c, 1\},\ \alpha_v = \{1\},\ \alpha_1 = \emptyset,$$

which are all distinct.

Observe that using the mapping $E(l) = \alpha_l$ for all $l \in L$, that is, if we change l by α_l, the result is the set of all the upsets ordered by $\supseteq$.

ISBN: 978-989-95079-6-8

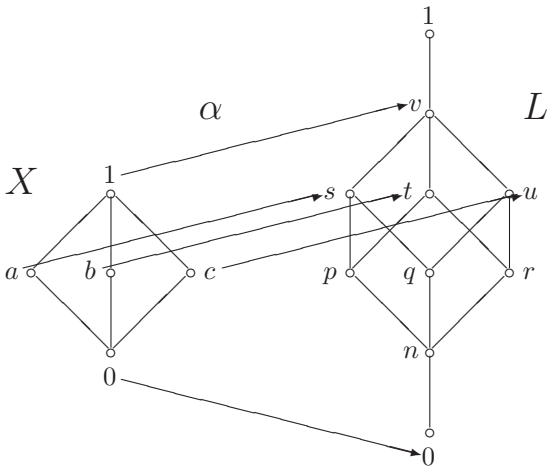


Figure 2: Hasse diagrams of X and L and L-valued set α.

Let us conclude this section with a version of the well-known Birkhoff Representation Theorem: on the one hand, by this theorem, we know that any finite distributive lattice L can be represented by the set of all up-sets on the set of all meet-irreducible elements of L. Now by Theorem 5, we can define an L-valued up-set whose cut sets are order isomorphic to L. Therefore, we can obtain a representation of any distributive lattice as a family of cut-sets of an L-valued up-set. This L-valued up-set is the embedding of $M(L)$ in L, that is, $\alpha : M(L) \to L$ with $\alpha(p) = p, \forall p \in M(L)$.

3.2 Characterization of the family of L-valued down-sets

All the previous studies made for L-valued up-sets could be repeated for L-valued down-sets. By the duality between these two concepts (see, for instance, Proposition 3), the obtained results are totally analogous.

Theorem 6 *Let $(X, \leq)$ be a poset and let $(L, \leq)$ be a complete lattice. A function $\mu : X \to L$ is an L-valued down-set on X if and only if any p-cut μ_p of μ is a crisp down-set (semi-ideal) on X, for every $p \in L$.*

Theorem 7 *Let $(X, \leq)$ be a poset and let $(L, \leq)$ be a complete lattice. If we consider the set $\mathcal{D}_L(X)$ formed by the collection of all L-valued down-sets on X and the induced order on $\mathcal{D}_L(X)$ is defined for any $\mu, \nu \in \mathcal{D}_L(X)$ as follows*

$$\mu \leq \nu \text{ if and only if } \mu(x) \leq \nu(x), \forall x \in X,$$

then the poset $(\mathcal{D}_L(X), \leq)$ is a complete lattice.

Proposition 6 *Let $(X, \leq)$ be a poset and let $(L, \leq)$ be a complete lattice. If a lattice identity holds in L, then the same identity is satisfied in the lattice $(\mathcal{D}_L(X), \leq)$. Thus, for a distributive lattice L, the lattice $(\mathcal{D}_L(X), \leq)$ is also distributive.*

Theorem 8 *Let $(X, \leq)$ be a poset, let $\mathcal{F} \subseteq \mathcal{P}(X)$ be a family of some down-sets of a poset X, and let $(L, \leq)$ be a complete lattice. Then, there is an L-valued down-set $\mu : X \to L$ such that its family of cuts is equal to $\mathcal{F}$ if and only if the following two conditions hold:*

1. *$\mathcal{F}$ is closed under intersections and contains X;*
2. *there is a closure operator C on L, such that the poset $(L/C, \leq)$ is order isomorphic to $(\mathcal{F}, \supseteq)$.*

Proposition 7 *Let $(X, \leq)$ be a poset, let $(L, \leq)$ be a complete lattice and let $\mu : X \to L$ be an L-valued down-set. The family of cut sets of μ is formed by all the crisp down-sets of poset $(X, \leq)$ if and only if for every family $\{x_i \mid i \in I\}$ of elements from X the condition $\mu(x) \geq \bigwedge_{i \in I} \mu(x_i)$ for $x \in X$ implies that $x \leq x_i$ for some $i \in I$.*

Corollary 3 *Let $(X, \leq)$ be a poset, let $(L, \leq)$ be a complete lattice and let $\mu : X \to L$ be an L-valued down-set. If the family of cut sets of μ is formed by all the crisp down-sets of poset $(X, \leq)$, then the two following statements hold:*

- *μ is an order-embedding.*
- *For any $x \in X$, $\mu(x)$ is a completely meet-irreducible element in the lattice L^{μ}.*

Theorem 9 *Let $(X, \leq)$ be a poset, let $(L, \leq)$ be a complete lattice of finite length and let $\mu : X \to L$ be an L-valued down-set such that*

1. *μ is an bijection to set $M(L)$ of all meet-irreducible elements of lattice L.*
2. *For every family $\{x_i \mid i \in I\}$ of elements from X, if $\mu(x) \geq \bigwedge_{i \in I} \mu(x_i)$ for $x \in X$, then $x \leq x_i$, for some $i \in I$.*

Then, the lattice L is distributive and L is isomorphic with the family of all down-sets of the poset X.

4 Conclusion

In this work, we have carried out an in-depth study of some particular classes of L-valued sets: up-sets and down-sets. We have started with the L-valued up-sets and we have proven that this is a cutworthy property. Moreover, we have established necessary and sufficient conditions under which for a given family of crisp up-sets, there exists an L-valued set such that its collection of cuts coincides with the family of crisp up-sets. In particular, we have done a deeper study for the L-valued up-sets such that its family of cuts is formed by all the crisp up-sets. In the case of lattices of finite length, we have arrived to the Birkhoff Representation Theorem. All these studies have been repeated for the case of L-valued down-sets and analogous results have been obtained.

In future works we would like to obtain some necessary and sufficient conditions not only for the existence of an L-valued up-set (resp. down-set) for a given family of up-sets (resp. down-sets), but for the uniqueness of such L-valued up-set (resp. down-set). The properties under which uniqueness is guaranteed were already studied for general fuzzy sets in [23]. Now, we try to adapt these results in two senses: L-valued sets instead of fuzzy sets (valued in the interval $[0, 1]$) and up-sets (resp. down-sets) instead of any sets.

Acknowledgements

The research reported on in this paper has been partially supported by Projects FEDER-MEC-MTM2004-01269 and FICYT-PC07-11 and also by Serbian Ministry of Science and

ISBN: 978-989-95079-6-8

Environment, Grant No. 144011 and by the Provincial Secretariat for Science and Technological Development, Autonomous Province of Vojvodina, grant "Lattice methods and applications".

References

[1] J.A. Goguen. *l*-fuzzy sets. *J. Math. Anal. Appl.*, 18:145–174, 1967.

[2] C.V. Negoita and D.A. Ralescu. *Applications of fuzzy sets to system analysis*. Basel, Birkhäuser Verlag, 1975.

[3] R. Bělohlávek. *Fuzzy Relational Systems*. Kluwer Academic Publishers, 2002.

[4] S. Díaz, B.De Baets, and S. Montes. Additive decomposition of fuzzy pre-orders. *Fuzzy Sets and Systems*, 158:830–842, 2007.

[5] S. Díaz, S. Montes, and B.De Baets. Transitivity bounds in additive fuzzy preference structures. *IEEE Trans. Fuzzy Systems*, 15:275–286, 2007.

[6] L.A. Zadeh. Similarity relations and fuzzy orderings. *Inform. Sci.*, pages 177–200, 1971.

[7] B. Borchardt, A. Maletti, B. Šešelja, A. Tepavčević, and H. Vogler. Cut sets as recognizable tree languages. *Fuzzy Sets and Systems*, 157:1560–1571, 2006.

[8] L.M. Brown and E. Ertürk. Fuzzy sets as texture spaces, i. representation theorems. *Fuzzy Sets and Systems*, 110:227–235, 2000.

[9] M. Chechik, B. Devereux, S. Easterbrook, A.Y.C. Lai, and V. Petrovykh. Efficient multiple-valued model-checking using lattice representations. *Lecture Notes in Computer Science*, 2154:451–465, 2001.

[10] Y.M. Liu and M.K. Luo. *Fuzzy Topology*. World Scientific, 1997.

[11] B. Šešelja. *l*-fuzzy covering relation. *Fuzzy Sets and Systems*, 158:2456–2465, 2007.

[12] B. Šešelja and A. Tepavčević. A note on natural equivalence relation on fuzzy power set. *Fuzzy Sets and Systems*, 148:201–210, 2004.

[13] B. Šešelja, A. Tepavčević, and H. Vogler. A note on cut-worthiness of recognizable tree series. *Fuzzy Sets and Systems*, 159:3087–3090, 2008.

[14] A. Tepavčević and G. Trajkovski. *l*-fuzzy lattices: an introduction. *Fuzzy Sets and Systems*, 123:209–216, 2001.

[15] R. Fuentes-González. Down and up operators associated to fuzzy relations and t-norms: A definition of fuzzy semi-ideals. *Fuzzy Sets and Systems*, 117:377–389, 2001.

[16] G. Birkhoff. *Lattice Theory*. Amer. Math. Soc. Providence R.I., 1967.

[17] B.A.Davey and H.A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, 1992.

[18] B. Šešelja and A. Tepavčević. Completion of ordered structures by cuts of fuzzy sets, an overview. *Fuzzy Sets and Systems*, 136:1–19, 2003.

[19] B. Šešelja and A. Tepavčević. Representing ordered structures by fuzzy sets, an overview. *Fuzzy Sets and Systems*, 136:21–39, 2003.

[20] M. Gorjanac-Ranitović and A. Tepavčević. General form of lattice-valued fuzzy sets under the cutworthy approach. *Fuzzy Sets and Systems*, pages 1213–1216, 2007.

[21] B. Šešelja and A. Tepavčević. Representation of lattices by fuzzy sets. *Information Sciences*, 79:171–180, 1993.

[22] D. Sinha. A general theory of fuzzy arithmetic. *Fuzzy Sets and Systems*, 36:339–363, 1990.

[23] A. Jaballah and F.B. Saidi. Uniqueness results in the representation of families of sets by fuzzy sets. *Fuzzy Sets and Systems*, 157:964–975, 2006.

ISBN: 978-989-95079-6-8

Fuzzy Intersection and Difference Model for Topological Relations

Ahed ALBOODY[1] Florence SEDES[2] Jordi INGLADA[3]

1 Université Paul Sabatier (UPS)
Toulouse, 118 Route de Narbonne, F-31062-CEDEX 9, France
2 Institut de Recherche en Informatique de Toulouse (IRIT), Université Paul Sabatier (UPS),
Toulouse, 118 Route de Narbonne, F-31062-CEDEX 9, France
3 Centre National d'Etudes Spatiales (CNES)
Toulouse, 18 Avenue Edouard Belin, 31401- CEDEX 9, France
Email: albahed@yahoo.fr, Florence.Sedes@irit.fr, jordi.inglada@cnes.fr

Abstract—*Topological relations have played important roles in spatial query, analysis and reasoning in Geographic Information Systems (GIS) and spatial databases. The topological relations between crisp and fuzzy spatial objects based upon the 9-intersections topological model have been identified. However the formalization of the topological relations between fuzzy regions needs more investigation. The paper provides a theoretical framework for modelling topological relations between fuzzy regions based upon a new fuzzy topological model called the Fuzzy Intersection and Difference (FID) Model. A novel topological model is formalized based on Fuzzy Topological Space (FTS). In order to derive all fuzzy topological relations between two fuzzy spatial objects, the fuzzy spatial object (A) is decomposed in four components: the Interior, the Interior's Boundary, the Object's Boundary, and the Exterior's Boundary of A. By use of this definition of fuzzy spatial object, new 4*4-Intersection and Fuzzy Intersection and Difference (FID) models are proposed as a qualitative model for the identification of all topological relations between two simple fuzzy regions. These two new models are compared with other fuzzy models studied in the literature. Examples are provided to illustrate the use of these two models presented in this paper with results which can be applied for modeling GIS and geospatial databases.*

Keywords— Fuzzy Intersection and Difference Model, 4*4-Intersection Matrix, Fuzzy Objects, Topological Relations, GIS

1 Introduction

Topological relations have an important significance in GIS modelling since they are the basis for spatial modelling, spatial query, analysis and reasoning. How to identify the topological relations between spatial objects is a critical point in GIS modelling.

During recent years, topological relations have been much investigated in the crisp and fuzzy topological space. The well-known *4-intersection* approach described in [1, 2], as well as the *9-intersection* approach as discussed in [3], and the *Intersection and Difference* (*ID*) model studied in [4, 5, 6], were proposed to formalize topological relations between two simple regions in the *Crisp Topological Space* (*CTS*).

The *4-intersection* model is extended in [7] to deal with the topological relations between spatial objects with holes.

Geographical phenomena in GIS with uncertain boundaries can be modelled by *Regions with Broad Boundaries* (*BBRs*) as in [8]. A region with broad boundary is an extension of a region with a crisp boundary (refer to simple regions with holes as in [7]). Objects with broad boundaries as defined in [8] are spatial objects, whose crisp boundaries are replaced by an area expressing the boundary's uncertainty. The *9-intersection model* is extended in [9] to describe topological relations between *BBRs* by replacing the crisp boundary in the *9-intersection* with the *broad boundary*. Another method called *4-tuple representation* of topological relations between *BBRs* is used in [10] to infer new topological information. The *4-tuple representation* can distinguish the same topological relations as identified by the *extended 9-intersection*. The *4-tuple*, however, can be applied to the reasoning of topological relations between *BBRs* [10], because it uses the composition of topological relations between crisp regions to determine those between uncertain and vague regions.

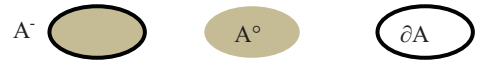
More recently fuzzy spatial objects have been emphasized since there are spatial features which are not always crisp. Fuzzy spatial objects are those with indeterminate boundaries. For fuzzy boundaries, that is, boundaries that are by nature not crisp, the broad boundary represents their minimum and maximum extent.

In order to derive the topological relations between fuzzy spatial objects, the *9-intersection approach* was updated into the *3*3-intersection approach* in the fuzzy topological space [11, 12]. Furthermore, in these two works, a *4*4-intersection matrix* was built up by using the topological properties of fuzzy sets, and then a *5*5-intersection matrix* can be built up based on certain conditions.

In the next section, we compare these models.

2 Related Work about Topological Relation Models and Spatial Objects

Crisp spatial objects have been formally defined in GIS. Point, line and polygon are three primitives in GIS. Fig. 1 represents the closure, interior and boundary of a closed disk as crisp spatial objects [2]. The *4-intersection* and *9-intersection matrix* are well-known approaches to identifying topological relation models between these two crisp spatial objects using the concept of the interior, boundary and exterior.



Closure Interior Boundary

Figure 1: Closure, Interior and Boundary of a Crisp Object

By using some topological invariants of the intersection such as the empty/non empty contents, the topological

ISBN: 978-989-95079-6-8

relations between two crisp spatial objects can be identified. This approach implies the following facts in *CTS*: (1) the interior, boundary and the exterior of a subset are topological invariants; (2) these topological invariants are mutually disjoint in *CTS*; and (3) the empty/non-empty contents of the intersections between these three topological parts of two subsets are topological invariants. Then the *4-intersection* and *9-intersection models* are defined as:

$$I_4(A,B)=\begin{bmatrix} A^\circ \cap B^\circ & A^\circ \cap \partial B \\ \partial A \cap B^\circ & \partial A \cap \partial B \end{bmatrix}$$

$$I_9(A,B)=\begin{bmatrix} A^\circ \cap B^\circ & A^\circ \cap \partial B & A^\circ \cap B^- \\ \partial A \cap B^\circ & \partial A \cap \partial B & \partial A \cap B^- \\ A^- \cap B^\circ & A^- \cap \partial B & A^- \cap B^- \end{bmatrix}$$

Eight topological relations of the spatial reasoning system (Region Connection Calculus) *RCC8* (*DC*, *EC*, *EQ*, *PO*, *TPP*, *TPPi*, *NTPP* and *NTPPi*) have been identified between two simple regions by using these two models in [2, 3].

In *ID model* [4, 5, 6], a crisp spatial object is defined by its interior and boundary; two intersection sets are $A^\circ \cap B^\circ$ and $\partial A \cap \partial B$; two difference sets are A - B and B - A. This model can also distinguish the eight topological relations. This crisp topological model is represented by:

$$ID(A,B)=\begin{bmatrix} A^\circ \cap B^\circ & A-B \\ B-A & \partial A \cap \partial B \end{bmatrix}$$

The main difference between the *4-intersection model* and *ID model* is that both intersection sets $A^\circ \cap \partial B$ and $\partial A \cap B^\circ$ of 4-intersection model are replaced by two differences A - B and B - A.

However, the fact (2) in CTS cannot hold in Fuzzy Topological Space (FTS). That means the interior; the boundary and the exterior of a fuzzy set may not be disjoint with each other. Therefore the *4-intersection, 9-intersection* and *ID models* cannot be directly applied for the identification of relations between two fuzzy sets.

The *9-intersection model* is extended in [8, 9] to spatial objects with broad boundary as simple fuzzy regions. It's expressed by the following matrix:

$$M = I_9(A,B)=\begin{bmatrix} A^\circ \cap B^\circ & A^\circ \cap \Delta B & A^\circ \cap B^- \\ \Delta A \cap B^\circ & \Delta A \cap \Delta B & \Delta A \cap B^- \\ A^- \cap B^\circ & A^- \cap \Delta B & A^- \cap B^- \end{bmatrix}$$

Using this model, *44 relations* between two simple fuzzy regions by using the *3*3-intersection matrix* are possible. For composite regions with broad boundaries, there are *14 additional* topological relations [8]. Fig. 2 represents a region with broad boundary as simple fuzzy spatial object.



Figure 2: Region with a Broad Boundary

In [14], they investigated a special space and formalized the 9-intersection in *Crisp FTS*. They proved that the *Crisp FTS* whose open sets are crisp is able to meet the above conditions.

The *9-intersection matrix* can be formalized as:

$$I_9(A,B)=\begin{bmatrix} A^\circ \cap B^\circ & A^\circ \cap \partial B & A^\circ \cap B^e \\ \partial A \cap B^\circ & \partial A \cap \partial B & \partial A \cap B^e \\ A^e \cap B^\circ & A^e \cap \partial B & A^e \cap B^e \end{bmatrix}$$

The above *3*3-intersection matrix* is derived based on the interior, boundary, and exterior of a simple fuzzy spatial object. By use of *9-intersection matrix*, *44 topological relations* are identified between two simple fuzzy regions (See Appendix 1. in [14]). Fig. 3 represents a region with indeterminate boundary as simple fuzzy spatial object.



Figure 3: Two Fuzzy Spatial Objects

In [14], it was shown that fuzzy spatial objects can be decomposed into four parts: the interior, the boundary of the boundary $\partial(\partial A)$, the interior of the boundary $(\partial A)^\circ$ and the exterior, which are mutually disjoint as in Fig. 4.

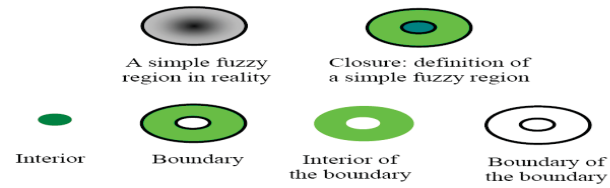


Figure 4: Interior, Boundary, Interior of the Boundary and Boundary of the Boundary of a Simple Fuzzy Region (After X. Tang and W. Kainz in [14])

Therefore, they introduced a *4*4-Intersection matrix* between two simple fuzzy spatial objects as follows:

$$I_{4*4}=\begin{bmatrix} A^\circ \cap B^\circ & A^\circ \cap \partial B & A^\circ \cap (\partial B)^\circ & A^\circ \cap B^e \\ \partial(\partial A) \cap B^\circ & \partial(\partial A) \cap \partial B & \partial(\partial A) \cap (\partial B)^\circ & \partial(\partial A) \cap B^e \\ (\partial A)^\circ \cap B^\circ & (\partial A)^\circ \cap \partial B & (\partial A)^\circ \cap (\partial B)^\circ & (\partial A)^\circ \cap B^e \\ A^e \cap B^\circ & A^e \cap \partial B & A^e \cap (\partial B)^\circ & A^e \cap B^e \end{bmatrix}$$

Under certain conditions, *152 relations* are identified by using the *4*4-intersection* approach (See Appendix 2. in [14] for more details).

After investigation about the topological relations between two simple fuzzy regions compared with these models studied in the literature, we can see that some topological relations can't be identified by these models. Here are some relations presented in Fig. 5.

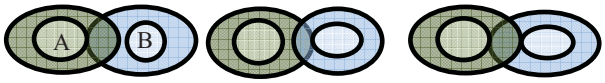


Figure 5: Some of Topological Relations to Identify

The question is how many topological relations there are exactly between two simple fuzzy objects? To answer this question, we will extend the *4-intersection* and the *ID models* with a new definition of the fuzzy boundary for fuzzy regions.

The disadvantage of the *3*3-intersection* and *4*4-intersection* models is that the *intersection operator* ($\cap$) is the most expensive one in terms of computation. In order to reduce the computational cost of the *3*3-intersection* and *4*4-intersection* models; and to reduce the computational complexity by avoiding spatial operations between

ISBN: 978-989-95079-6-8

topological components with different dimensions (1-D and 2-D), we will try to reduce the number of intersections by introducing the *difference operator* (-).

Finally, the motivation of the paper is trying to build a topological model for identification of all fuzzy topological relations between two fuzzy regions.

The structure of the paper is as follows. The fuzzy topological space is defined in section 3. Section 4 is our contribution which consists of a new definition of simple fuzzy regions, fuzzy boundaries and their properties. The novel major contribution of our study, proposed in section 5, is a new form of *4*4- intersection model* and *the Fuzzy Intersection and Difference (FID) model*. Section 6 shows the identification of fuzzy topological relations between two simple fuzzy regions by using two models (*4*4- intersection and FID*) based on empty/non-empty contents. We end with a discussion about results and a conclusion.

3 Fuzzy Topological Space (FTS)

Fuzzy topology is constructed based on fuzzy sets. It is an extension of general (crisp) topology.

Let A be a fuzzy subset of an ordinary (crisp) set X, and $\wp(X)=[0,1]^X$ be the fuzzy power set of X. $[0,1]^X$ can be viewed as a lattice in which a supremum (or *join*) is denoted by $\vee$ and an infimum (or *meet*) by $\wedge$, conventionally. They correspond to the *union* and the *intersection*, respectively. $\forall\delta\in\wp(X)$ if $(1)\,\Phi, X\subseteq\delta$, $(2)\,\forall A_i\in\delta,\ \vee A_i\in\delta,(3)$ $\forall U,V\in\delta,\ U\wedge V\in\delta$, then δ is called a *fuzzy topology* on *X* (*i*∈*I* is an index set). (*X, δ*) is called a *Fuzzy Topological Space (FTS)* as defined in [15, 16]. Every element of δ is called an open (fuzzy) set in (X, δ). A set A is a closed (fuzzy) set if its complement A^c is open. The union of all open sets contained in A is the interior of A, denoted by A°. The intersection of all closed sets containing A is called the closure of A, denoted by A^-. The exterior of A is the complement of A^- and is denoted by A^e. Obviously, it is an open set. The boundary ∂A of a subset A is the intersection of the closure of A with the closure of the complement of A.

The boundary of a subset may also have its interior and its boundary of the boundary. On the other hand, the interior and the closure of a subset also have their boundaries. For example, the boundary of the boundary of a fuzzy set A is the union of the boundary of the closure and the boundary of the interior of a fuzzy set [14].

Based on this information, one can define a maximum of four other different areas from that defined in [14] for each object: Interior, Boundary of the Object, Interior of the Boundary, and Exterior of the Boundary in the next section.

4 Definition of Simple Fuzzy Regions and a Fuzzy Boundary

In this section, we will develop a definition of simple fuzzy region, fuzzy boundary and their properties.

4.1 Definition of Simple Fuzzy Region

A crisp region is defined in *CTS*. Correspondingly, a fuzzy region should be defined in *FTS*. We now define a simple fuzzy region in *FTS*.

A simple fuzzy region is made up of two regions A_1 and A_2 with $A_1\subset A_2$ (see Fig. 6), where: (1) the interior of A is the interior of A_1 $A^\circ=(A_1)^\circ$ and A° is an open subset and connected; (2) the interior's boundary of A is the boundary of A_1 as $A^i=\partial(A^\circ)=\partial A_1$, and A^i is a closed subset and connected; (3) the boundary of A is ∂A defined as the interior of the difference between A_1 and A_2 as $\partial A=(A_2-A_1)^\circ$, and ∂A is an open subset and connected; (4) the exterior's boundary of A is the boundary of ∂A_2 as $A^e=\partial^e(\partial A)=\partial A_2$, and A^e is a closed subset and connected; and (5) the intersection of all closed sets containing A is called the closure of A, denoted by A^-. Fig. 6 shows the four components of simple fuzzy regions.

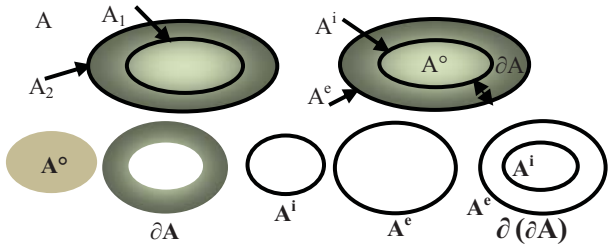


Figure 6: Interior, Boundary, Interior's Boundary, and Exterior's Boundary of a Simple Fuzzy Region

This definition is considered as the de decomposition of the boundary in [14, 15, 16] into disjoint subsets such as the interior boundary of the boundary, the exterior boundary of the boundary and the interior of the boundary with condition that is the interior of the boundary couldn't be a non-empty set.

We called the boundary (∂A) of A by the fuzzy boundary. This definition of fuzzy regions is very interesting to identify all topological relations between two simple fuzzy regions that other models [8, 9, 14] can't identify. We will prove that in the next section by developing a *4*4-intersection* matrix. In the next part, we will define the properties of the fuzzy boundary.

4.2 Properties of Fuzzy Boundary

Let A be a fuzzy set in *FTS* (X, δ). Based on the definition of simple fuzzy regions, we find the properties of the fuzzy boundary (∂A) as follows: (1) ∂A is an open subset of A; (2) ∂A = ¬ Ø is a non-empty subset; (3) the boundary of A (∂A) is the interior of the difference $\partial A=(A^e-A^i)^\circ$ between the exterior's boundary (A^e) and the interior's boundary (A^i) of A; (4) the union $A^i\cup\partial A\cup A^e$ of the interior's boundary (A^i), the boundary (∂A) and the exterior's boundary (A^e) of A is a closed subset of A; (5) the interior boundary of the boundary (∂A) of A $\partial^i(\partial A)=A^i=\partial(A^\circ)$ is the interior's boundary (A^i); and (6) the exterior boundary of the boundary (∂A) of A $\partial^e(\partial A)=A^e=\partial^e(A)$ is the exterior's boundary (A^e).

We also find that the intersections between A^o, A^i, ∂A, A^e are, respectively, always empty, and the union of these parts

ISBN: 978-989-95079-6-8

is equal to A as follows: $A^{\circ} \cap A^{i} = \emptyset$; $A^{i} \cap \partial A = \emptyset$; $\partial A \cap A^{e} = \emptyset$; $A^{\circ} \cap A^{i} \cap \partial A \cap A^{e} = \emptyset$; and $A^{\circ} \cup A^{i} \cup \partial A \cup A^{e} = A$.

It also can be easily proven by the above intersections that the interior (A^{o}), the interior's boundary (A^{i}), the boundary ∂A and the exterior's boundary (A^{e}) of a simple fuzzy region (A) are mutually disjoint.

In order to identify all possible topological relations, the condition of the mutual disjointness of these four parts of simple fuzzy regions is important to propose and construct a new method to form the intersection matrix. In the next section, a new *4*4-intersection matrix* and *fuzzy intersection and difference model* are proposed based upon this definition of simple fuzzy region.

5 Contributions: 4*4 - Intersection and Fuzzy Intersection and Difference (FID) Models

In this section, we will develop two models to identify the topological relations between two simple fuzzy regions.

Supposing there are two simple fuzzy objects A and B in the *FTS*, we adopt the interior, boundary, interior's boundary, and exterior's boundary to formalize two new topological models as in the next.

*5.1 Contribution 1: 4*4 - Intersection Model*

In the first contribution, the first model is a new *4*4 - intersection matrix* which uses the operator (∩) of intersection. Between these two simple fuzzy spatial regions A and B, the *4*4-intersection matrix* will be as presented in Table 1:

Table 1: 4*4 - Intersection Matrix

∩	B°	B^{i}	∂B	B^{e}
A°	A°∩ B°	A°∩ B^{i}	A°∩ ∂B	A°∩ B^{e}
A^{i}	A^{i} ∩B°	A^{i} ∩ B^{i}	A^{i} ∩∂B	A^{i} ∩ B^{e}
∂A	∂A ∩B°	∂A ∩ B^{i}	∂A ∩∂B	∂A ∩ B^{e}
A^{e}	A^{e} ∩B°	A^{e} ∩ B^{i}	A^{e} ∩ ∂B	A^{e} ∩ B^{e}

And the *4*4-inetrsection model* applied to simple fuzzy objects is expressed by the following expression:

$$I_{4*4}(A,B) = \begin{bmatrix} A^{\circ} \cap B^{\circ} & A^{\circ} \cap B^{i} & A^{\circ} \cap \partial B & A^{\circ} \cap B^{e} \\ A^{i} \cap B^{\circ} & A^{i} \cap B^{i} & A^{i} \cap \partial B & A^{i} \cap B^{e} \\ \partial A \cap B^{\circ} & \partial A \cap B^{i} & \partial A \cap \partial B & \partial A \cap B^{e} \\ A^{e} \cap B^{\circ} & A^{e} \cap B^{i} & A^{e} \cap \partial B & A^{e} \cap B^{e} \end{bmatrix}$$

This new *4*4- intersection matrix* (I_{4*4}) is considered as an extension of the *4-intersection* model [1, 2] for simple fuzzy spatial regions.

The *intersection operator* (∩) is perhaps the most expensive one in terms of computation. In order to reduce the computational cost of this *4*4-intersection* model, we will extend this model to the fuzzy intersection and difference model in the next part.

5.2 Contribution 2: Fuzzy Intersection and Difference (FID) Model

In this model, we will introduce the *difference operator* (-). In order to avoid spatial operations between topological components with different dimensions (A°, B°, ∂A, ∂B as 2-D; and A^{i}, B^{i}, A^{e}, B^{e} as 1-D), we will replace the intersection between the terms A°∩ B^{i}, A°∩ B^{e}, A^{i} ∩B°, A^{i} ∩∂B, ∂A ∩ B^{i}, ∂A ∩ B^{e}, A^{e} ∩B° and A^{e} ∩ ∂B in the *4*4-intersection* model by the differences as in the matrix (see Table 2).

The four intersections (A°∩B°, A°∩∂B, ∂A∩B°, ∂A∩∂B) with topological components with dimension 2-D and the four intersections (A^{i}∩B^{i}, A^{i}∩B^{e}, A^{e}∩B^{i}, A^{e}∩B^{e}) with topological components with dimension 1-D remain unchanged as similar as in [4, 5, 6] for the ID model.

Table 2: 4*4 - Intersection and Difference Matrix

∩	B°	B^{i}	∂B	B^{e}	-
A°	A°∩ B°	A^{i} - B^{i}	A°∩ ∂B	A^{i} - B^{e}	A°
A^{i}	B^{i} - A^{i}	A^{i} ∩ B^{i}	B^{i} - A^{i}	A^{i} ∩ B^{e}	A^{i}
∂A	∂A ∩B°	A^{e} - B^{i}	B^{e} – A^{i} ∂A ∩∂B	A^{e} - B^{e}	∂A
A^{e}	B^{i} - A^{e}	A^{i} - B^{i} A^{e} ∩ B^{i}	B^{i} – A^{e} B^{e} - A^{e}	A^{i} - B^{e} A^{e} ∩ B^{e}	A^{e}

By simplification and arrangement of this *4*4 - Intersection and Difference matrix*, we obtain two intersection matrices and two difference matrices (Table 3):

Table 3: Intersection and Difference Matrices

∩	B°	∂B
A°	A°∩ B°	A°∩ ∂B
∂A	∂A ∩B°	∂A ∩∂B

∩	B^{i}	B^{e}
A^{i}	A^{i}∩ B^{i}	A^{i}∩ B^{e}
A^{e}	A^{e} ∩ B^{i}	A^{e} ∩ B^{e}

-	B^{i}	B^{e}
A^{i}	A^{i} - B^{i}	A^{i} - B^{e}
A^{e}	A^{e} - B^{i}	A^{e} - B^{e}

-	A^{i}	A^{e}
B^{i}	B^{i} - A^{i}	B^{i} - A^{e}
B^{e}	B^{e} - A^{i}	B^{e} - A^{e}

At the end, the *Fuzzy Intersection and Difference (FID) model* is written as follows:

$$FID_{4*4}(A,B) = \begin{bmatrix} A^{\circ} \cap B^{\circ} & A^{\circ} \cap \partial B & A^{i} - B^{i} & A^{i} - B^{e} \\ \partial A \cap B^{\circ} & \partial A \cap \partial B & A^{e} - B^{i} & A^{e} - B^{e} \\ A^{i} \cap B^{i} & A^{i} \cap B^{e} & B^{i} - A^{i} & B^{i} - A^{e} \\ A^{e} \cap B^{i} & A^{e} \cap B^{e} & B^{e} - A^{i} & B^{e} - A^{e} \end{bmatrix}$$

The *FID* model combines two different operators (intersection and difference). The *FID* has two advantages: first, it reduces the computational complexity by avoiding spatial operations between topological components with different dimensions, e.g., A°∩ B^{i}, A°∩ B^{e}, A^{i} ∩B°, A^{i} ∩∂B, ∂A ∩ B^{i}, ∂A ∩ B^{e}, A^{e} ∩B° and A^{e} ∩ ∂B, with A°, B°, ∂A, ∂B as 2-D, and A^{i}, B^{i}, A^{e}, B^{e} as 1-D; and second, it reduces the computational cost due to only eight intersections in the matrix of *FID* model.

This *FID* model is considered as an extension of the *ID* model [4, 5, 6] for simple fuzzy spatial regions.

In general, there are 2^{16}= 65536 relations between two fuzzy regions by using the *4*4-intersection* matrix and *FID* model. For GIS applications, some conditions will limit the number of these relations as in [7, 8, 14]. However, how to find all possible topological relations between two simple fuzzy regions needs more investigation. It's done in the next section for the *4*4-intersection* matrix and *FID* model.

ISBN: 978-989-95079-6-8

6 Identification of Topological Relations based upon 4*4 - Intersection and FID Models

In this section, we focus on the identification of all fuzzy topological relations between two simple fuzzy regions by *4*4-intersection* and *FID* models.

Let A and B be two simple fuzzy regions. For relation identification, each intersection or difference in *4*4-intersection* and *FID matrices* takes value of either empty ($\emptyset$) or non-empty ($\neg\emptyset$). Every different set of *4*4-intersection* and *FID* matrices describes a different topological relation. Some values of these two matrices have no sense on the topological relation.

*6.1 Identification by 4*4 - Intersection Model*

For identification by *4*4-intersection model*, by respecting the definition in section 3, we scan all possible configurations for A and B in two different steps as following: (1) If the exterior's boundary of A intersects with the exterior's boundary of B ($A^e \cap B^e = \neg\emptyset$), then, we search all possible topological relations between A and B, we find *105 relations*; (2) If the exterior's boundary of A doesn't intersect with the exterior's boundary of B ($A^e \cap B^e = \emptyset$), then, we search all topological relations between A and B, we find *47 relations*. The total number of topological relations identified between A and B is *152 relations*. Some of these topological relations are not identified and determined in [8, 9, 14]. Here are some examples in Fig .7 and Fig .8.

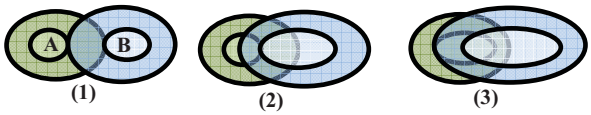


Figure 7: Examples for $A^e \cap B^e = \neg\emptyset$

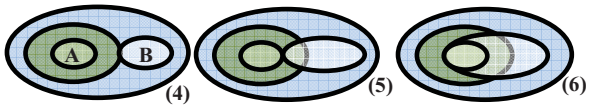


Figure 8: Examples for $A^e \cap B^e = \emptyset$

The *4*4-intersection matrices* correspondent for (1), (2) and (3) in Fig .7, and for (4), (5) and (6) in Fig .8 are given, respectively, by:

$$(1)\begin{bmatrix}\phi & \phi & \phi & \phi\\ \phi & \phi & \phi & \neg\phi\\ \phi & \phi & \neg\phi & \neg\phi\\ \phi & \neg\phi & \neg\phi & \neg\phi\end{bmatrix}\ (2)\begin{bmatrix}\phi & \phi & \neg\phi & \neg\phi\\ \phi & \neg\phi & \neg\phi & \neg\phi\\ \neg\phi & \neg\phi & \neg\phi & \neg\phi\\ \neg\phi & \neg\phi & \neg\phi & \neg\phi\end{bmatrix}\ (3)\begin{bmatrix}\neg\phi & \neg\phi & \neg\phi & \phi\\ \neg\phi & \neg\phi & \neg\phi & \neg\phi\\ \neg\phi & \neg\phi & \neg\phi & \neg\phi\\ \neg\phi & \neg\phi & \neg\phi & \neg\phi\end{bmatrix}$$

$$(4)\begin{bmatrix}\phi & \phi & \neg\phi & \phi\\ \phi & \phi & \neg\phi & \phi\\ \phi & \phi & \neg\phi & \phi\\ \phi & \neg\phi & \neg\phi & \phi\end{bmatrix}\ (5)\begin{bmatrix}\phi & \phi & \neg\phi & \phi\\ \phi & \neg\phi & \neg\phi & \phi\\ \neg\phi & \neg\phi & \neg\phi & \phi\\ \neg\phi & \neg\phi & \neg\phi & \phi\end{bmatrix}\ (6)\begin{bmatrix}\neg\phi & \phi & \phi & \phi\\ \neg\phi & \neg\phi & \phi & \phi\\ \neg\phi & \neg\phi & \neg\phi & \phi\\ \neg\phi & \neg\phi & \neg\phi & \phi\end{bmatrix}$$

We note that the separateness between these relations can't be realized by other models studied in [8, 14]. In the next part, we will identify all relations by the *FID model*.

6.2 Identification by FID Model

For identification by *FID model*, by respecting the definition in section 3, we apply the same two steps in previous part (6.1) for $A^e \cap B^e = \neg\emptyset$ and $A^e \cap B^e = \emptyset$. 152 topological relations can be identified by using the *FID model*. These relations identified by *FID model* is the same relations by *4*4-intersection model*. We give some examples in Fig .9 and Fig .10.

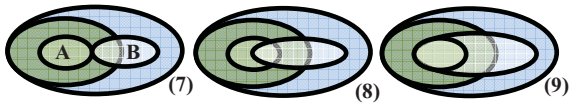


Figure 9: Examples for $A^e \cap B^e = \neg\emptyset$

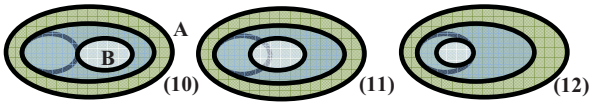


Figure 10: Examples for $A^e \cap B^e = \emptyset$

The *FID* matrices correspondent to (7), (8) and (9) in Fig .9, and to (10), (11) and (12) in Fig .10 are given, respectively, by:

$$(7)\begin{bmatrix}\phi & \neg\phi & \neg\phi & \phi\\ \neg\phi & \neg\phi & \neg\phi & \phi\\ \neg\phi & \phi & \neg\phi & \neg\phi\\ \neg\phi & \neg\phi & \neg\phi & \neg\phi\end{bmatrix}\ (8)\begin{bmatrix}\neg\phi & \neg\phi & \neg\phi & \phi\\ \neg\phi & \neg\phi & \neg\phi & \phi\\ \neg\phi & \phi & \neg\phi & \neg\phi\\ \neg\phi & \neg\phi & \neg\phi & \neg\phi\end{bmatrix}\ (9)\begin{bmatrix}\neg\phi & \phi & \phi & \phi\\ \neg\phi & \neg\phi & \neg\phi & \phi\\ \neg\phi & \phi & \neg\phi & \neg\phi\\ \neg\phi & \neg\phi & \neg\phi & \neg\phi\end{bmatrix}$$

$$(10)\begin{bmatrix}\phi & \neg\phi & \neg\phi & \phi\\ \neg\phi & \neg\phi & \neg\phi & \neg\phi\\ \neg\phi & \neg\phi & \neg\phi & \phi\\ \phi & \phi & \neg\phi & \phi\end{bmatrix}\ (11)\begin{bmatrix}\neg\phi & \neg\phi & \neg\phi & \phi\\ \neg\phi & \neg\phi & \neg\phi & \neg\phi\\ \neg\phi & \neg\phi & \neg\phi & \phi\\ \phi & \phi & \neg\phi & \phi\end{bmatrix}\ (12)\begin{bmatrix}\neg\phi & \neg\phi & \neg\phi & \phi\\ \phi & \neg\phi & \neg\phi & \neg\phi\\ \neg\phi & \neg\phi & \phi & \phi\\ \phi & \phi & \neg\phi & \phi\end{bmatrix}$$

To find and extract these relations by the *4*4-intersection* and *FID models*, we have developed and implemented these two steps on MATLAB.

7 Applications and Discussion

For GIS applications, satellite images and spatial databases, these two models (*4*4-intersection* and *FID*) can determine the topological relations between simple fuzzy regions. For that, we need to know how to generate fuzzy spatial objects from satellite images and to find the four components of our definition (interior, interior's boundary, boundary and exterior's boundary) for satellite images and GIS objects. We can adopt processed data such as classification or segmentation results of satellite images. For example, Land Use and Land Cover (LULC), most of which is obtained from the classification results of satellite images, may be a good example of a fuzzy spatial object as in [17]. In principle, a fuzzy spatial object can also be generated by other methods as in [14, 18].

These two models can be used in order to evaluate the change detection process (for Land Cover changes)of geographical objects (Beach, Forest, Residential Area…) represented in GIS and satellite imagery (TM and SPOT images) databases as in [14, 17, 19, 20].

Comparing between the relations identified in [8, 9, 14] and our models (*4*4-intersection and FID*), we find that there are *100 new relations* which can't be discriminated and identified by other models. Some of these *100 new topological relations* identified by *4*4-intersection and FID* models are presented in the appendix of this paper.

8 Conclusion

In this paper, we have proposed a new definition of simple fuzzy spatial region by decomposing the spatial region into

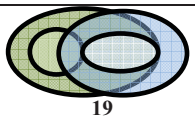
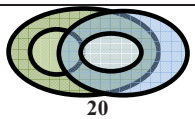
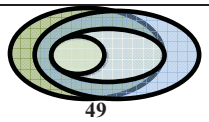
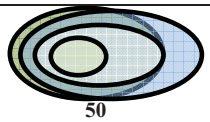
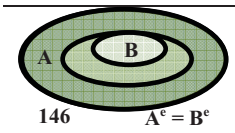
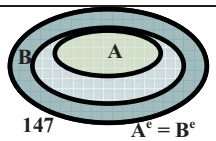
ISBN: 978-989-95079-6-8

four components: interior, interior's boundary, boundary and exterior's boundary. Based upon these four components, a *4*4-intersection matrix* is introduced to identify all topological relations between two simple fuzzy regions. Then, in order to reduce the computational complexity of the *4*4-intersection model,* the *Fuzzy Intersection and Difference (FID) model* is developed based on the *4*4-intersection matrix*. The main contribution of this work is these two models *4*4-intersection and FID. 152* fuzzy topological relations can be identified by using the *4*4-intersection and FID models*. Among these *152 relations*, *100 new relations* can't be discriminated or identified as different relations by other models [8, 9, 14].

In our future work, we will try to find more topological relations which can be identified if we adopt in the definition $A^e \cap A^i = \neg\emptyset$, $\partial A = \neg\emptyset$ and A^i don't intersect with the exterior of A. Then, we will try to classify these fuzzy topological relations identified by I4*4 and FID models to be grouped into the eight relations DC, EC, PO, TPP and TPPi, NTPP and NTPPi, and EQ of the spatial reasoning system *RCC8*.

Appendix

Some Topological Relations between Two Simple Fuzzy Regions by Using the *4*4-Intersection Model (I_{4*4} Matrix)* and the *Fuzzy Intersection and Difference (FID Matrix) Model*

Illustration	I_{4*4} Matrix	FID Matrix	Illustration	I_{4*4} Matrix	FID Matrix
 19	(19) $\begin{bmatrix} \phi & \phi & \neg\phi & \neg\phi \\ \phi & \neg\phi & \neg\phi & \neg\phi \\ \neg\phi & \neg\phi & \neg\phi & \neg\phi \\ \phi & \neg\phi & \neg\phi & \neg\phi \end{bmatrix}$	(19) $\begin{bmatrix} \phi & \neg\phi & \neg\phi & \neg\phi \\ \neg\phi & \neg\phi & \neg\phi & \neg\phi \\ \neg\phi & \neg\phi & \neg\phi & \phi \\ \neg\phi & \neg\phi & \neg\phi & \neg\phi \end{bmatrix}$	 20	(20) $\begin{bmatrix} \phi & \phi & \neg\phi & \neg\phi \\ \phi & \neg\phi & \neg\phi & \neg\phi \\ \neg\phi & \neg\phi & \neg\phi & \neg\phi \\ \phi & \phi & \neg\phi & \neg\phi \end{bmatrix}$	(20) $\begin{bmatrix} \phi & \neg\phi & \neg\phi & \neg\phi \\ \neg\phi & \neg\phi & \neg\phi & \neg\phi \\ \neg\phi & \neg\phi & \neg\phi & \phi \\ \phi & \neg\phi & \neg\phi & \neg\phi \end{bmatrix}$
 49	(49) $\begin{bmatrix} \neg\phi & \phi & \phi & \phi \\ \neg\phi & \neg\phi & \phi & \phi \\ \neg\phi & \neg\phi & \neg\phi & \neg\phi \\ \phi & \neg\phi & \neg\phi & \neg\phi \end{bmatrix}$	(49) $\begin{bmatrix} \neg\phi & \phi & \phi & \phi \\ \neg\phi & \neg\phi & \neg\phi & \neg\phi \\ \neg\phi & \phi & \neg\phi & \phi \\ \neg\phi & \neg\phi & \neg\phi & \neg\phi \end{bmatrix}$	 50	(50) $\begin{bmatrix} \neg\phi & \phi & \phi & \phi \\ \neg\phi & \phi & \phi & \phi \\ \neg\phi & \neg\phi & \neg\phi & \neg\phi \\ \phi & \neg\phi & \neg\phi & \neg\phi \end{bmatrix}$	(50) $\begin{bmatrix} \neg\phi & \phi & \phi & \phi \\ \neg\phi & \neg\phi & \neg\phi & \neg\phi \\ \phi & \phi & \neg\phi & \phi \\ \neg\phi & \neg\phi & \neg\phi & \neg\phi \end{bmatrix}$
 146 $A^e = B^e$	(146) $\begin{bmatrix} \neg\phi & \phi & \neg\phi & \phi \\ \phi & \neg\phi & \neg\phi & \phi \\ \phi & \phi & \neg\phi & \phi \\ \phi & \phi & \phi & \neg\phi \end{bmatrix}$	(146) $\begin{bmatrix} \neg\phi & \neg\phi & \neg\phi & \phi \\ \phi & \neg\phi & \neg\phi & \phi \\ \neg\phi & \phi & \phi & \phi \\ \phi & \neg\phi & \neg\phi & \phi \end{bmatrix}$	 147 $A^e = B^e$	(147) $\begin{bmatrix} \neg\phi & \phi & \phi & \phi \\ \neg\phi & \neg\phi & \phi & \phi \\ \neg\phi & \neg\phi & \neg\phi & \phi \\ \phi & \phi & \phi & \neg\phi \end{bmatrix}$	(147) $\begin{bmatrix} \neg\phi & \phi & \phi & \phi \\ \neg\phi & \neg\phi & \neg\phi & \phi \\ \neg\phi & \phi & \neg\phi & \phi \\ \phi & \neg\phi & \neg\phi & \phi \end{bmatrix}$

References

[1] M.J. Egenhofer, A formal definition of binary topological relationships, *The Third International Conference of Foundations of Data Organization and Algorithms (FODO)*, Paris, France, Springer-Verlag, pp. 452-472, 1989.

[2] M.J. Egenhofer, and R. Franzosa, Point-set topological spatial relations, *International Journal of Geographic Information Systems,* 5(2): 161-174, 1991.

[3] M.J. Egenhofer, and J.R. Herring, A mathematical framework for the definition of topological relationships, *Proceedings of the 4th International Symposium on Spatial Data Handling,* Zurich, Columbus, OH, International Geographical Union, pp. 803-813, 1990.

[4] M. Deng, T. Cheng, X. Chen and Z. Li, Multi-level topological relations between spatial regions based upon topological invariants, *Springer Science, GeoInformatica*, 11(2): 239-267, 2007.

[5] A. Alboody, J. Inglada and F. Sèdes, Enriching the spatial reasoning system RCC8, *16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM GIS 2008)*, *The SIGSPATIAL Special,* 1(1): 14-20, 2009.

[6] A. Alboody, F. Sèdes and J. Inglada, Multi-level topological relations of the spatial reasoning system RCC-8, *The First International Conference on Advances in Databases, Knowledge, and Data Applications (DBKDA'09), IEEE Computer Society*, pp. 13–21, 2009.

[7] M.J. Egenhofer, E. Clementini and P. D. Felice, Topological relations between regions with holes, *International Journal of Geographic Information Systems*, 8(2): 129-144, 1994.

[8] E. Clementini, Objects with broad boundaries, *Book of S. Shekhar and H. Xiong: Encyclopedia of GIS*, Springer US, pp. 793-799, 2008.

[9] E. Clementini and P. Di Felice, An algebraic model for spatial objects with indeterminate boundaries, *P. Burrough, A. Frank (Eds.), Geographic Objects with Indeterminate Boundaries*, Taylor & Francis, London, pp. 155–169, 1996.

[10] S. Du, Q. Qin, Q. Wang and H. Ma, Reasoning about topological relations between regions with broad boundaries, *International Journal of Approximate Reasoning,* 47(2): 219-232, 2008.

[11] X. Tang, Y. Fang and W. Kainz, Fuzzy topological relations between fuzzy spatial objects, *LNCS in Springer Berlin, Fuzzy Systems and Knowledge Discovery*, 4223: 324-333, 2006.

[12] X. Tang and W. Kainz, Analysis of topological relations between fuzzy regions in a general fuzzy topological space, *Proceedings of Symposium on Geospatial Theory, Processing and Applications,* Canada, Ottawa, 2002.

[13] X. Tang, Y. Fang, and W. Kainz, Topological matrices for topological relations between fuzzy regions, *Proceedings of the 4th International Symposium on Multispectral Image Processing and Pattern Recognition (SPIE),* 6045: 604524, Wuhan, China, 2005.

[14] X. Tang and W. Kainz, Spatial object modelling in fuzzy topological spaces with applications to Land Cover change, *Ph.D Thesis, 9 January* 2004. www.itc.nl/library/Papers_2004/phd/xinming.pdf

[15] C. L. Chang, Fuzzy topological spaces, *Journal of Math. Anal. Appl*. 24: 182–190, 1968.

[16] X. Tang, W. Kainz and H. Zhang, Some topological invariants and a qualitative topological relation model between fuzzy regions, *4th International Conference on Fuzzy Systems and Knowledge Discovery, IEEE*, 1(24-27): 241- 246, 2007.

[17] A. Bassiri, A. Alesheikh and M. R. Malek, Spatio-temporal object modelling in fuzzy topological space, *The International Archives of the Photogrammetric, Remote Sensing and Spatial Information Sciences,* XXXVII (B2): 131-134, 2008.

[18] W. Shi and K. Liu, A fuzzy topology for computing the interior, boundary, and exterior of spatial objects quantitatively in GIS, *Computers & Geosciences, Elsevier,* 33: 898–915, 2007.

[19] T. Cheng, Fuzzy objects: their changes and uncertainties, *The Journal of the American Society for Photogrammetric Engineering and Remote Sensing*, 2 (1): 41 - 49, 2002.

[20] T. Cheng, P. Fisher and Z. Li, Double vagueness: uncertainty in multi-scale fuzzy assignments of duneness, *The Revue of Geo-Spatial Information Science*, *Springer*, 7(1): 58-66, 2004.

ISBN: 978-989-95079-6-8

T-norm-based fuzzy logics and logics for reasoning under vagueness

Thomas Vetterlein, Klaus-Peter Adlassnig

Institute for Medical Expert and Knowledge-Based Systems, Medical University of Vienna
Vienna, Austria
Email: {Thomas.Vetterlein,Klaus-Peter.Adlassnig}@meduniwien.ac.at

Abstract— *We contrast the concept underlying t-norm-based propositional fuzzy logics with the problem to whose solution fuzzy logics are frequently suggested as helpful – namely, to find a model of reasoning with vague information. We argue that fuzzy logics are useful as long as truth values can be identified with the meaning of the considered propositions. This, however, is rarely the case in practice; hence we see the need to broaden the concept underlying this important class of logics and try fresh approaches. In particular, we should flexibilise the formalism to allow that propositions do not arise in the same context, but are just known to be related in some way.*

We tackle the problem tentatively. We define a set of rules which, as we assume, are minimally required to enable us to argue about vague propositions whose content is not taken into account. Our choice of rules reflects the practical requirements of a certain expert system on which we work.

Although we deal here with fuzzy logic in a very direct sense, we arrive at calculi completely different from the t-norm-based ones. Without incorporating truth degrees explicitly, we are led to Belnap's logic, which can, but need not, be endowed with a semantics based on graded truth degrees. When formalising also truth degrees, we get a logic which can be based on what we call metric De-Morgan lattices.

Keywords— t-norm-based fuzzy logics, reasoning under vagueness, medical expert system, De-Morgan lattices, metric De-Morgan lattices

1 T-norm-based fuzzy logic for reasoning about vague information – a trap

Fuzzy logics are distinguished from classical logic by the incorporation of an extended set of truth values. In the standard case, the value 0 is used to express falsity, the value 1 is used to express trueness, and all remaining real values in between these two limit points are added in order to cope with the fact that objects may fulfil a property to an intermediate degree. We arrive naturally at the idea to evaluate propositions in the real unit interval, whose most basic feature is its linear order.

The connectives used in fuzzy logics consequently need to be interpreted by operations on $[0, 1]$. Typically, a conjunction is present, which is typically interpreted by a left-continuous t-norm. Sometimes, an involutive negation is present as well, conveniently interpreted by the standard negation $1 - \cdot$. A further connective, which logicians, in contrast to engineers, generally consider as the most basic one, is the implication, which is typically interpreted by the residuum belonging to the t-norm. Finally, the logic may or may not offer the possibility to express explicitly to which degree a proposition holds.

Based on this approach, more than just a few logics have been defined and intensively studied. Monographs of basic importance include [14, 13] as well as, as regards the explicitation of truth degrees, [18]. We remark that we certainly do not address all logics which have been called "fuzzy" in the literature; this would be impossible as it seems that nowadays any logic has a "fuzzy" counterpart. Here, we just speak about logics of the indicated type.

We wish to address in this note the peculiar relationship between t-norm-based fuzzy logics and a problem which is regularly mentioned in discussions on the fundamentals of fuzzy logics: how to formalise reasoning when the referred information is possibly vague. The discussion on the very nature of fuzzy logics is old. An aim has been to develop fuzzy logics, as they are, from clear, meaningful principles; see, among many papers, e.g. [9, 10, 21, 22]. Here, we want to approach once the subject from the other side, namely, from the point of view of a specific application: an appropriate formalisation of justifiable reasoning can simply be a practical need.

Let us first try to formulate our concern in an abstract way. We assume to be given a set of propositions whose content does not matter. This means that for the derivation of consequences, the meaning is not taken into account and can be assumed to be unknown. We just know that the propositions refer to the observable or unobservable properties of somebody or something, describable in, possibly scientific, natural language. Important for us, the propositions express the presence of some property which can be vague, where, as usual, vagueness is characterised by the possibility of borderline cases. Moreover, we assume to have some knowledge about the mutual relationships. These relationships may express that some property is more general than another one, or a causal implication based on experience; again, we do not require the relationships to hold necessarily strict. What we are finally interested in, is to find a formal framework which tells us how to derive new information from the one we have to our disposal. Since the outlined situation is very general, the framework must be very general as well; we wonder, so to say, about a minimal logic for reasoning under vagueness.

The described problem is not purely academic and in particular not part of any "ivory-tower" philosophical theory, but has a practical background. We work towards an appropriate formalisation of a medical expert system [7]. The system which we are to analyse is called Cadiag-2, the second generation of the expert systems Cadiag – "*C*omputer-*A*ssisted *Dia*gnosis", and aims at the differential diagnostic decision support in patient care [1, 2]. Cadiag-2 processes both vague and uncertain data; we restrict here to the first case only. The procession of uncertain data calls for a probabilistic logic and causes completely different problems than those discussed here; for probability theory, well-founded calculi exist, whose unpleas-

ISBN: 978-989-95079-6-8

ant feature, however, can be a too high complexity as well as incomprehensible inference rules.

As we intend to formalise propositions disregarding their content, our problem is clearly a case for a propositional logic. As mentioned, many propositional fuzzy logics have been proposed in the past, based on different ways how to endow the real unit interval with a structure. Moreover, a wide range of logics has been introduced related to the problem which we address here, in particular different versions of "logics of argumentation". For a comprehensive overview and a large collection of references, we recommend the handbook [12]. Here, we just mimic the first steps towards the generally much more sophisticated and often more specialised systems found in the literature. What we have in mind is to concentrate exclusively on the aspect of vagueness, to proceed in a way which can hardly be further generalised, and to see how the result relates to t-norm-based fuzzy logics.

The set of eligible propositional logics is much restricted by one basic requirement, dictated by the intended application: all constituents of the formal logic need to have a counterpart on the informal level. Specifically, there must exist a plausible way how to think about each connective which appears in the logic's language, most easily obtained by a clear correlation with a natural-language expression. Moreover, if a proposition is provable from others, there must exist a proof in a proof system such that each step is comprehensible as a plausible argument, rather than a pure manipulation of strings; in the best case, each step can easily be translated to an explanation in natural language, exhibiting the causal or logical relationship on which the argument is based.

A fuzzy logic of the above-indicated kind does not meet any of these requirements. To see the problem, let us assume that we actually can find a logic fitting to our needs. We actually feel that it is natural to assume so; after all, we wish to formalise possibly vague statements, and these are appropriately evaluated on a linear continuous scale. So let us see how the logic could look like. We need a conjunction $\wedge$ and a negation $\sim$; the interpretation by the infimum and the standard negation, respectively, will do in our case. We note that the interpretation of truth values is not our subject here, and our particular choices for the connectives provide just an example. Moreover, we need to express truth degrees explicitly; to this end, we add constants $\bar{r}$ for each rational $r \in [0,1]$. So far, we do not encounter problems.

The first serious problem comes with the implication. Let us tentatively add the connective $\rightarrow$ interpreted by the residuum belonging to $\wedge$. Note that we then arrive at the logic $\mathsf{RGL}_\sim$, the Gödel logic enriched with the standard negation and truth constants [8]. Now, there is a natural way how to think about a statement "$\alpha \rightarrow \beta$": we interpret it as "α implies β". However, this clarity disappears as soon as we nest implications on the left side, like in "$(\alpha \rightarrow \beta) \rightarrow \gamma$". If our reference, a set of propositions, has a priori the structure of a residuated lattice, we can say that $\alpha \rightarrow \beta$ denotes the weakest element which, together with α, implies β. But in our case, there is no such structure available; we recall that we do not wish to assume any a priori structure as we would have to analyse the propositions by content then.

We note that this critics is related to the discussion out of which relevance logics arose [3]. Furthermore, a discussion in which the role of the implication connective in fuzzy logics is opposed to the needs of certain applications, can be found in [4].

For us, there is a reasonable way how to proceed: to drop the implication as a generally applicable connective. To this end, we consider in [7] a logic in which the implication always appears at the only place at which it can be appropriately called an implication: on the outermost level. Namely, we consider pairs of implication-free formulas, with the intended meaning that the left one denotes a proposition which is stronger then the right one. Thus we leave the area of t-norm-based logics and enter the field of lattice word problems. Namely, our model is the following algebra K, a Kleene algebra with added constants: $([0,1]; \wedge, \vee, \sim, (\bar{r})_{r \in \mathbb{Q}[0,1]})$.

We arrive at a logic, which we call GZL, whose formal expressions possess straightforward interpretations. However, we encounter the second and even more serious problem when considering a proof system for GZL. Following the lines of [5], we have in [7] presented a proof system based on sequents-of-relations. Unfortunately, our requirement that proof steps should be comprehensible in an informal way, is far from being fulfilled. Consider the K-tautology

$$\alpha \wedge \sim\alpha \;\rightarrow\; \beta \vee \sim\beta. \tag{1}$$

In its proof, we have to make necessarily use of the possibility to use multisets of relations, namely,

$$\alpha \Rightarrow \beta \;\mid\; \beta \Rightarrow \alpha \tag{2}$$

will appear in the proof. However, α and β refer to arbitrary facts, and the tentative translation "α implies β, or β implies α" is nonsense. Note that the problem concerns the proved result as well; (1) cannot be interpreted as a statement which anybody would ever tell.

The deeper reason for this difficulty is the semantics. An element of $[0,1]$ is intended to be the truth degree of a proposition; but it is treated like its meaning. What we might think as being associated to a property, telling that the property does not fully apply, is already the property itself. As a consequence, when using $[0,1]$ as a model, we may be led to the situation that we compare something by strength what by content would never be comparable. A valid statement of the form "$\alpha \rightarrow \beta$" is not really translatable to "from α we can conclude β", but only that α is under all circumstances assigned a smaller truth value than β, and based on this interpretation, (2) becomes indeed meaningful and just expresses the linear order of the truth degrees. However, this interpretation is not what we want.

The interpretation in the linearly ordered set of reals may certainly be useful at other places. A requirement comes into play which we have frequently argued for: to put fuzzy logics on firm grounds, we need first to be aware of the nature of what we reason about. In case of t-norm-based fuzzy logic, we reason about a set of propositions which has the internal structure of a residuated lattice, as it is the case for universes of fuzzy sets. The same, by the way, applies to classical propositional logic, which exactly reasons about a collection of propositions endowed a priori with the structure of a Boolean algebra; the popular claims about a "general validity" of this logic are meaningless.

ISBN: 978-989-95079-6-8

For us, the only way out is to restrict the calculus for GZL to those inferences which are not in conflict with our intended interpretation: the sequent $\alpha \Rightarrow \beta$ should mean that α is a statement stronger than β. We can achieve this is by not allowing multisets of relations, but only single relations. The interesting observation in [7] is that by means of this restriction, we get – not exactly but, say – very close to the logic which is actually used for the expert system which we examine.

The observation that the logic which we need arises by a certain restriction of a t-norm-based fuzzy logic, might be considered interesting, but not really satisfying. It rather suggests that the conceptual differences between the logic underlying systems like Cadiag-2 or similar expert systems on the one hand, and fuzzy logics on the other hand, cannot be bridged.

2 A minimal logic for reasoning under vagueness, without explicit degrees

The problem how to formalise ways to argue about vague propositions of unspecified content and their mutual interrelations, calls for alternative solutions. Let us opt for the syntactical approach; we will assemble some inference rules which translate to argumentation steps in a straightforward way. We will then check if some semantics with a reasonable interpretation can be found ex post, taking all imaginable possibilities into account and in particular not restricting ourselves to structures known from fuzzy logics or fuzzy set theory.

We note that this procedure seems to be in sharp contrast to the guiding principles of mathematical modelling which we have defended earlier, namely the principle that prior to any formalisation, the structure of reference needs to be specified first, in a way that the meaning of all its constituents is clear. However, in the present case, we do not do metamathematics, we do not examine ways how structures of a certain type are generally examined in a sound way; we do mathematics. Namely, it is the way of reasoning itself which is our object of investigation, and we do not share the opinion that rules for proper argumentation are fixed and thus can be derived from some higher-level truth. Intuitively acceptable inference rules will rather constitute a structure over a set of atomic propositions, and we do not assume a canonical answer how it may look like. In any case, we examine a logic as a mathematical object, the notion "logic" just being a name for it.

As indicated, there is not really a canonical way to select rules. One may argue against certain rules shown below, or feel that there is something missing. A discussion would not be fruitful if no guidelines were provided. We keep with the application in medicine; the rules shown below are extracted from those essential for the medical expert system with whose formalisation we are concerned.

In this section, we consider the case that we do not deal with truth values explicitly. We define the propositional logic DML as follows.

Definition 2.1 The *propositions* of DML are built up from a set of symbols $\varphi_1, \varphi_2, \ldots$ and the two constants $\bar{0}, \bar{1}$ by means of the binary connectives $\wedge, \vee$ and the unary connective $\sim$; the set of propositions is denoted by $\mathcal{F}_L$. The *implications* of DML are ordered pairs of propositions, denoted by $\alpha \to \beta$, where $\alpha, \beta \in \mathcal{F}_L$; the set of implications is denoted by $\mathcal{F}_I$.

Moreover, a *sequent* is an ordered pair of a non-empty finite set of propositions and a single proposition, notated by $\gamma_1, \ldots, \gamma_k \Rightarrow \delta$. The axioms and rules of DML are the following, for any propositions α, β, γ and sequent Γ:

$$\bar{0} \Rightarrow \alpha \quad \alpha \Rightarrow \alpha \quad \alpha \Rightarrow \bar{1}$$

$$\frac{\Gamma \Rightarrow \alpha \quad \alpha \Rightarrow \beta}{\Gamma \Rightarrow \beta} \qquad \frac{\Gamma \Rightarrow \alpha}{\Gamma, \beta \Rightarrow \alpha}$$

$$\frac{\Gamma \Rightarrow \alpha \quad \Gamma \Rightarrow \beta}{\Gamma \Rightarrow \alpha \wedge \beta} \qquad \frac{\Gamma, \alpha, \beta \Rightarrow \gamma}{\Gamma, \alpha \wedge \beta \Rightarrow \gamma}$$

$$\frac{\Gamma, \alpha \Rightarrow \gamma \quad \Gamma, \beta \Rightarrow \gamma}{\Gamma, \alpha \vee \beta \Rightarrow \gamma} \qquad \frac{\Gamma \Rightarrow \alpha}{\Gamma \Rightarrow \alpha \vee \beta} \qquad \frac{\Gamma \Rightarrow \beta}{\Gamma \Rightarrow \alpha \vee \beta}$$

$$\frac{\alpha \Rightarrow \beta}{\sim\beta \Rightarrow \sim\alpha} \qquad \frac{\sim\alpha \Rightarrow \beta}{\sim\beta \Rightarrow \alpha} \qquad \frac{\alpha \Rightarrow \sim\beta}{\beta \Rightarrow \sim\alpha}$$

The notion of a proof of a sequent from a finite set of sequents is defined in the expected way. A *theory* of DML is a finite set of implications. An implication $\alpha \to \beta$ is called provable from $\mathcal{T} = \{\alpha_1 \to \beta_1, \ldots, \alpha_n \to \beta_n\}$ if there is a proof of $\alpha \Rightarrow \beta$ from $\{\alpha_1 \Rightarrow \beta_1, \ldots, \alpha_n \Rightarrow \beta_n\}$, in signs $\mathcal{T} \vdash \alpha \to \beta$.

As to be expected, a sequent $\gamma_1, \ldots, \gamma_k \Rightarrow \delta$ is meant to express that "γ_1 and ... and γ_k imply δ".

Note that this logic again does not contain the implication as a connective, and it does contain a negation. Moreover, the calculus is sound, but not complete, with respect to classical two-valued interpretations; again (1) is not derivable. For the lack of the implication connective, also the comparison with intuitionistic logic is (probably) not well possible.

But a semantics is easily found, since DML is Belnap's logic of De-Morgan lattices; DML differs only slightly from the calculus presented in [11]. A structure $(M; \wedge, \vee, \sim, 0, 1)$ is a De-Morgan lattice if (i) $(M; \wedge, \vee, 0, 1)$ is a distributive lattice and (ii) $\sim$ is an order-reversing and involutive unary operation. De-Morgan lattices are subalgebras of direct products of the algebra $(\mathcal{M}_4; \wedge, \vee, \sim, 0, 1)$, where $(\mathcal{M}_4; \wedge, \vee, 0, 1)$ is the four-element Boolean lattice and $\sim$ maps each of the two atoms to itself [16]. It follows that we can provide a semantics based on $\mathcal{M}_4$; we note that assigning one of the four truth values to a proposition φ is usually interpreted as that φ is known to be true, false, neither true nor false, both true and false, respectively.

Most remarkably, fuzziness does not appear. However, the algebra $\mathcal{M}_4$ possesses a natural "fuzzified" extension, and we may alternatively base DML on a kind of fuzzy semantics. To make the comparison possible, we consider the pair V_c and V_f; note that the algebra V_c is isomorphic to $\mathcal{M}_4$.

Definition 2.2 Let $V_c = \{(s,t) : s,t \in \{0,1\}\}$, endowed with the componentwise natural order and the operation $\sim$ defined by

$$\sim(s,t) = (1-t, 1-s) \tag{3}$$

for $s,t \in \{0,1\}$. A *crisp evaluation* of DML is a mapping $v\colon \mathcal{F}_L \to V_c$ preserving $\wedge, \vee, \sim$ and mapping $\bar{0}$ to $(0,0)$ and $\bar{1}$ to $(1,1)$. An implication $\alpha \to \beta$ is then said to be *satisfied* by v if $v(\alpha) \leq v(\beta)$. A theory $\mathcal{T}$ is said to *crisply entail* an implication $\alpha \to \beta$ if the latter is satisfied by all crisp evaluations satisfying every element of $\mathcal{T}$; we write $\mathcal{T} \models_c \alpha \to \beta$ in this case.

ISBN: 978-989-95079-6-8

Furthermore, let $V_f = \{(s,t)\colon\ s,t \in [0,1]\}$, endowed with the componentwise natural order and the operation $\sim$, which is again defined by (3), where however this time $s,t \in [0,1]$. We define *fuzzy evaluations*, *satisfaction*, and *fuzzy entailment* similarly as above.

Theorem 2.3 *Let $\mathcal{T}$ be a theory and $\alpha \to \beta$ an implication of* DML. *Then $\mathcal{T} \vdash \alpha \to \beta$ if and only if $\mathcal{T} \models_c \alpha \to \beta$ if and only if $\mathcal{T} \models_f \alpha \to \beta$.*

Proof. Completeness in the indicated sense, but with respect to arbitrary De-Morgan lattices, holds due to [11, Theorem 4.11, (A1)]. But any De-Morgan lattice is a subalgebra of a direct product of copies of V_c, or alternatively, of a direct product of copies of V_f. □

As regards the interpretation of the semantics which we have proposed for DML, the situation is surprising with respect to the fuzzy variant. There is a close connection of DML to a logic which has been proposed in the context of decision making, based on the observation that we often consider separately the arguments in favour and the arguments against a possible decision [19].

On the other hand, the fact that we can work, without the need to change the inference rules, with crisp truth values as well, is somewhat disillusioning. The reasoning is the same if we assume our propositions to refer to vague or crisp properties. The situation certainly changes when we include truth degrees, as to be done next.

3 A minimal logic for reasoning under vagueness, with explicit degrees

The logic DML discussed in the last section is intended to be useful for reasoning about relationships between statements involving vagueness. The vagueness, however, cannot be addressed directly; and the logic can actually equally well considered as a logic not concerning vagueness. Moreover, a possible non-strictness of the relationships themselves is not expressible.

In applications, it can be desirable to have the possibility to denote a vague property by one single symbol, to which the degree of presence is explicitly attached. In this chapter, we attempt to formulate a calculus similarly to DML, but with explicit reference to truth degrees.

This problem is much more involved. We believe that there are many possibilities, and that the decision which is the best is even more difficult than in the above case.

According to a common procedure, we could enrich the language by truth constants. We will not follow this way; apart from the fact that we have not succeeded to produce a reasonable result concerning a logic DML enriched with truth constants, the idea is actually not well in accordance with the guidelines formulated above: truth constants should not be mixed with the meaning of a proposition. So we will use the two sharp truth constants only, representing falsity and trueness.

Truth constants should appear on a separate level. We propose to make a graded implication the basic syntactical constituent:

$$\alpha \xrightarrow{t} \beta,$$

with the intended meaning that α implies β to a degree $\geq t \in [0,1]$, where α and β are propositions of DML. As an example, let α denote a crisp proposition like "having a body temperature of $37.8°C$", and let β denote "having fever"; then the statement would hold with, say, $t = 0.8$.

In general, α and β are meant to refer to any vague property. Then either t refers to the compatibility of α with β. Alternatively, we can mean that α and β are causally related; then the smaller t is, the less strict is this relationship. Finally, we may also deal with single properties. Namely, the expressions

$$\bar{1} \xrightarrow{t} \delta \quad \text{and} \quad \delta \xrightarrow{t} \bar{0} \tag{4}$$

may serve to express that δ holds to the degree t, or δ is refused with the degree t, respectively.

To formulate inference rules is not straightforward; in the present context, our aim can only be plausibility. The basic question is which truth degree is, by tendency, assumed after two successive inference steps which are both based on a non-strict relationship. To examine this problem is even more difficult than to make a reasonable choice with regard to the truth values themselves, a problem which has been studied numerous times, see e.g. [15, 17]. It would clearly be desirable to have methods at hand to examine also the present situation empirically, in analogy to the procedure followed in [15].

Only the case of implications of the form $\alpha \xrightarrow{1} \beta$ is clear; they are supposed to express strict relationships, and for them, the rules of DML should be applicable. Furthermore, the degrees are assumed to be lower bounds, hence an implication is the less expressive the smaller the indicated truth degree is. Statements of the form $\alpha \xrightarrow{0} \beta$ do not express anything.

So let us see how the set of rules for our refined logic could look like. For implications $\alpha \xrightarrow{1} \beta$, we will use the rules of DML. For implications of the form $\alpha \xrightarrow{t} \beta$, where $t < 1$, the rules of DML introducing $\wedge$ or $\vee$ are not generalisable though; the degree of the conclusion cannot be assumed to be calculable from the degrees of the assumptions. However, what we should be able to say is, if we replace α by a stronger proposition, or β by a weaker proposition, then the relationship between α and β should be characterised by a higher truth value, so that t will still be a lower bound. Next, assume that we have proved $\alpha \xrightarrow{s} \beta$ and $\beta \xrightarrow{t} \gamma$; then $\alpha \xrightarrow{u} \gamma$ will be derivable as well, and we have to offer a way to calculate the degree u from s and t. As mentioned above, a well-founded decision is impossible, hence just like in case of the design of a fuzzy logic, a pragmatic solution is needed here. We opt for the operation dual to the truncated addition: we take here the Łukasiewicz t-norm $\odot\colon\ [0,1]^2 \to [0,1],\ \ (a,b) \mapsto (a+b-1) \vee 0$.

We specify the propositional logic ArgL as follows.

Definition 3.1 The set $\mathcal{F}_L$ of *propositions* is defined like for DML. An *implication* of ArgL consists of ordered triples of two propositions and a rational value $t \in [0,1]$; we write $\alpha \xrightarrow{t} \beta$, where $\alpha, \beta \in \mathcal{F}_L$; the set of implications is denoted by $\mathcal{F}_I$.

Moreover, a *sequent* is an ordered triple consisting of a non-empty finite set of propositions, a single proposition, and a rational value $t \in [0,1]$; we write

$$\gamma_1, \ldots, \gamma_k \overset{t}{\Rightarrow} \delta.$$

ISBN: 978-989-95079-6-8

The *crisp* rules of ArgL are those of DML, the symbol $\Rightarrow$ being replaced at all places by $\stackrel{1}{\Rightarrow}$.

The *fuzzy* rules of ArgL are the following:

$$\frac{\Gamma \stackrel{s}{\Rightarrow} \alpha \quad \alpha \stackrel{t}{\Rightarrow} \beta}{\Gamma \stackrel{s \odot t}{\Rightarrow} \beta} \qquad \frac{\Gamma \stackrel{t}{\Rightarrow} \alpha}{\Gamma \stackrel{s}{\Rightarrow} \alpha}, \text{ where } s \leq t$$

$$\frac{\Gamma, \alpha \stackrel{t}{\Rightarrow} \delta}{\Gamma, \alpha \wedge \beta \stackrel{t}{\Rightarrow} \delta} \qquad \frac{\Gamma \stackrel{t}{\Rightarrow} \alpha}{\Gamma \stackrel{t}{\Rightarrow} \alpha \vee \beta}$$

The notion of a *proof*, a *theory*, the *provability* of an implication from a theory, is defined similarly like for DML.

To associate to this calculus a reasonable semantics, is the next challenge. Only one point seems to be certain – we are not led to fuzzy sets. The only remarkable fact is that a t-norm is involved; for connections between t-norms and a somewhat similar setting, see [6, 20].

Let us consider the following structures, so-to-say the algebraic counterpart of ArgL. Here, $\oplus : [0,1]^2 \to [0,1]$, $(a,b) \mapsto (a+b) \wedge 1$ is the t-conorm associated to $\odot$.

Definition 3.2 A structure $(A; \wedge, \vee, \sim, 0, 1, d)$ is called a *metric De-Morgan lattice* if $(A; \wedge, \vee, \sim, 0, 1)$ is a De-Morgan lattice and $d\colon A \times A \to [0,1]$ is such that (i) $d(a,b) = 0$ if and only if $a \leq b$ and (ii) $d(a,c) \leq d(a,b) \oplus d(b,c)$.

An *evaluation* of ArgL in a metric De-Morgan lattice A is a mapping $v : \mathcal{F}_L \to A$ preserving $\wedge, \vee, \sim$ and the constants. An implication $\alpha \stackrel{t}{\to} \beta$ is *satisfied* by an evaluation v if $d(v(\alpha), v(\beta)) \leq 1 - t$. Semantic entailment is defined as usual.

Let us consider the following instructive example. Let $(A; \wedge, \vee, \sim, 0, 1)$ be a Boolean algebra, and let $\mu\colon A \to [0,1]$ be a strictly positive submeasure on A, meaning that, for $a, b \in A$, (i) $\mu(0) = 0$, (ii) $\mu(a) > 0$ if $a > 0$, (iii) $a \leq b$ implies $\mu(a) \leq \mu(b)$, (iv) $\mu(1) = 1$, and (v) $\mu(a \vee b) \leq \mu(a) \oplus \mu(b)$. Furthermore, put $d(a,b) = \mu(a \wedge \sim b)$. Then we may check that $(A; \wedge, \vee, \sim, 0, 1, d)$ is a metric De-Morgan lattice.

Theorem 3.3 *Let $\mathcal{T}$ be a theory and $\alpha \stackrel{t}{\to} \beta$ an implication of* ArgL*, where $t \in (0,1]$. Then $\mathcal{T} \vdash \alpha \stackrel{t}{\to} \beta$ if and only if $\mathcal{T} \models \alpha \stackrel{t}{\to} \beta$.*

Proof. The "only if" part is easy; just define the satisfaction of sequents by identifying the set on the left side with its conjunction.

For the "if" part, assume that $\mathcal{T}$ does not prove the implication $\gamma \stackrel{s}{\to} \delta$.

Let A be the quotient of $\mathcal{F}_L$ w.r.t. the equivalence relation $\Leftrightarrow$, where $\alpha \Leftrightarrow \beta$ if $\alpha \stackrel{1}{\to} \beta$ and $\beta \stackrel{1}{\to} \alpha$ are provable from $\mathcal{T}$. Then A is naturally endowed with the structure of a De-Morgan lattice.

For $\alpha, \beta \in \mathcal{F}_L$, let $d([\alpha],[\beta]) = 1 - t$, where $t \in [0,1]$ is maximal such that $\mathcal{T} \vdash \alpha \stackrel{t}{\to} \beta$. Endowed with d, A is a metric De-Morgan lattice. By construction, the elements of $\mathcal{T}$ are satisfied by the natural embedding of $\mathcal{F}_L$ into A, but not $\gamma \stackrel{s}{\to} \delta$. □

Needless to comment, our formalism comes closer to measure theory than to fuzzy set theory.

From the interpretational point of view, the semantics based on metric De-Morgan lattices is to be clarified though. Note that only the special case of a Boolean algebra with a submeasure offers an intuitively well comprehensible picture.

4 Conclusion

T-norm-based propositional fuzzy logics are frequently discussed as a suitable tool to model reasoning under vagueness. We have stressed that this is the case as long as the propositions which are formalised share the same reference; namely, they must be modellable by a system of fuzzy sets over a common universe.

If propositions are arbitrary, we run into difficulties when trying to apply techniques of fuzzy logics. This is, for example, the case for the medical expert system Cadiag-2, whose knowledge base contains information on logical and causal relationships between entities which are processed regardless of their meaning. We have risen the question how to define, in this general setting, a minimal frame for what we could call formalised argumentations. As a proposal, we have designed two minimal, but for our needs fully sufficient, systems; we have done so purely syntactically, allowing only rules with a clear interpretation.

The first version concerns reasoning without explicit reference to truth degrees; what comes out is the De-Morgan logic, which allows for interpretations without any connection to fuzziness. The second version incorporates truth values, but the calculus which comes out, still is by no means related to t-norm-based fuzzy logic. The semantics which can be defined ex post are De-Morgan lattices endowed with a non-symmetric distance function.

Our calculi are qualified in that they allow to reproduce the inference mechanism of Cadiag-2. The further elaboration on details of the calculus is work to be done, as well as the analysis of the newly introduced notion of a metric De-Morgan lattice.

References

[1] K.-P. Adlassnig, G. Kolarz, W. Scheithauer, H. Effenberger, G. Grabner, Cadiag: Approaches to computer-assisted medical diagnosis, *Comput. Biol. Med.*, 15:315–335, 1985.

[2] K.-P. Adlassnig, G. Kolarz, W. Scheithauer, G. Grabner, Approach to a hospital-based application of a medical expert system, *Med. Inform.*, 11:205–223, 1986.

[3] A. R. Anderson, N. D. Belnap, J. M. Dunn, "Entailment. The logic of relevance and necessity", Vol. I and II, Princeton University Press, Princeton 1975 and 1992.

[4] L. Běhounek, On the difference between traditional and deductive fuzzy logic, *Fuzzy Sets Syst.* 159:1153–1164, 2008.

[5] M. Baaz, A. Ciabattoni, C. G. Fermüller, Cut-Elimination in a Sequents-of-Relations Calculus for Gödel Logic, in: Proceedings of the 31st International

ISBN: 978-989-95079-6-8

Symposium on Multiple-Valued Logic (Warsaw 2001), IEEE Computer Society; 181–186, 2001.

[6] B. De Baets, R. Mesiar, Pseudo-metrics and $\mathcal{T}$-equivalences, *J. Fuzzy Math.*, 5:471–481, 1997.

[7] A. Ciabattoni, T. Vetterlein, On the (fuzzy) logical content of Cadiag-2, submitted.

[8] F. Esteva, Ll. Godo, P. Hájek, M. Navara, Residuated fuzzy logics with an involutive negation, *Arch. Math. Logic*, 39:103–124, 2000.

[9] C. G. Fermüller, Theories of vagueness versus fuzzy logic: Can logicians learn from philosophers? *Neural Network World*, 13:455–466, 2003.

[10] C. G. Fermüller, Revisiting Giles's game, in: O. Majer, A.-V. Pietarinen, T. Tulenheimo (eds.), *Games: Unifying Logic, Language, and Philosophy*, Springer-Verlag, Berlin 2009; 209–228.

[11] J. M. Font, Belnap's four-valued logic and De Morgan lattices, *Log. J. IGPL*, 5:413–440, 1997.

[12] D. M. Gabbay, R. H. Johnson, H. J. Ohlbach, J. Woods, *Handbook of the logic of argument and inference. The turn towards the practical*, North-Holland, Amsterdam 2002.

[13] S. Gottwald, *A Treatise on Many-Valued Logics*, Research Studies Press, Baldock 2001.

[14] P. Hájek, *Metamathematics of Fuzzy Logic*, Kluwer, Dordrecht 1998.

[15] H. M. Hersh, A. Caramazza, A fuzzy set approach to modifiers and vagueness in natural language, *Journal of Experimental Psychology: General*, 105:254–276, 1976.

[16] J. A. Kalman, Lattices with involution, *Trans. Am. Math. Soc.*, 87:485–491, 1058.

[17] A. Matoušková, V. Novák, J. Kulka, Psycholinguistic Research of the Meaning of Linguistic Hedges, draft.

[18] V. Novák, I. Perfilieva, J. Močkoř, *Mathematical principles of fuzzy logic*, Kluwer, Dordrecht 1999.

[19] M. Öztürk, A. Tsoukiàs, Modelling uncertain positive and negative reasons in decision aiding, *Decision Support Systems* 43:1512–1526, 2007.

[20] T. Vetterlein, T-norms induced by metrics on Boolean algebras, *Soft Comput.*, 10:995–1000, 2006.

[21] T. Vetterlein, Fuzzy logic as a logic of the expressive strength of information, *Soft Comput.*, 12:479–485, 2008.

[22] T. Vetterlein, A way to interpret Łukasiewicz Logic and Basic Logic, *Studia Logica*, 90:407–423, 2008.

ISBN: 978-989-95079-6-8

On the representation theorem of multi-adjoint concept lattices*

Jesús Medina[1] Manuel Ojeda-Aciego[2]

1. Dept. Matemáticas, Universidad de Cádiz. Spain
2. Dept. Matemática Aplicada, Universidad de Málaga. Spain
Email: jesus.medina@uca.es, aciego@uma.es

Abstract— *Formal concept analysis has become an important and appealing research topic. There exist a number of different fuzzy extensions of formal concept analysis and of its representation theorem, which gives conditions for a complete lattice in order to be isomorphic to a concept lattice. In this paper we concentrate on the study of operational properties of the mappings α and β required in the representation theorem.*

Keywords— Formal concept analysis, multi-adjoint framework.

1 Introduction

Formal concept analysis [12] has become an important and appealing research topic both from a theoretical perspective [18, 29, 32] and from the applicative one. Regarding applications, we can find papers ranging from ontology merging [10, 27], to applications to the Semantic Web by using the notion of concept similarity [11], and from processing of medical records in the clinical domain [14] to the development of recommender systems [8].

Soon after the introduction of "classical" formal concept analysis, a number of different approaches for its generalization were introduced and, nowadays, there are works which extend the theory with ideas from fuzzy set theory [3, 21, 22] or fuzzy logic reasoning [2, 4, 9] or from rough set theory [20, 30, 33] or some integrated approaches such as fuzzy and rough [31], or rough and domain theory [19].

In this paper we concentrate on the fuzzy extensions of formal concept analysis, for which a number of different approaches have been presented. To the best of our knowledge, the first one was given in [6], although they did not advance much beyond the basic definitions, probably due to the fact that they did not use residuated implications. Later, in [3, 28] the authors independently used complete residuated lattices as structures for the truth degrees; for this approach, a representation theorem was proved directly in a fuzzy framework in [5], setting the basis of most of the subsequent direct proofs.

In [23, 24] as a new general approach to formal concept analysis multi-adjoint concept lattices were introduced, in which the philosophy of the multi-adjoint paradigm [15, 26] to formal concept analysis is applied. With the idea of providing a general framework in which the different approaches stated above could be conveniently accommodated, the authors worked in a general non-commutative environment; and this naturally led to the consideration of adjoint triples, also called implication triples [1] or bi-residuated structures [25] as the main building blocks of a multi-adjoint concept lattice.

*Partially supported by the Spanish Science Ministry under grant TIN 2006-15455-C03-01 and by Junta de Andalucía under grant P06-FQM-02049.

The representation (or fundamental) theorem gives conditions for a complete lattice in order to be isomorphic to a concept lattice. This theorem is proved in the classical case [12] and in the fuzzy paradigms [5, 13, 16, 23, 28]. As a consequence, to obtain the isomorphism it is necessary to search two mappings α and β that satisfy some properties, one of these relate the mappings with the relation. In this paper, we present a characterization of this last property, which is more efficient than the actual from the computationally point of view. Moreover, some other interesting properties of mappings α and β are introduced.

The structure of the paper is as follows: in Section 2 we recall the definition of the multi-adjoint concept lattices and, in particular, the mappings α and β required in the definition of lattice representing a multi-adjoint concept lattice. Then, in Section 3, we prove some new results concerning α and β. Finally, some concluding remarks are added.

2 Multi-adjoint concept lattices

In this section we will recall the more important definitions and results from [23]. The first definition introduces the basic building blocks of the multi-adjoint concept lattices, the *adjoint triples*, which are generalisations of the notion of adjoint pair under the hypothesis of having a non-commutative conjunctor.

The lack of commutativity of the conjunctor, directly provides two different ways of generalising the well-known adjoint property between a t-norm and its residuated implication, depending on which argument is fixed in the conjunction.

Definition 1 *Let $(P_1, \leq_1)$, $(P_2, \leq_2)$, $(P_3, \leq_3)$ be posets and $\&\colon P_1 \times P_2 \to P_3$, $\swarrow\colon P_3 \times P_2 \to P_1$, $\nwarrow\colon P_3 \times P_1 \to P_2$ be mappings, then $(\&, \swarrow, \nwarrow)$ is an* adjoint triple *with respect to P_1, P_2, P_3 if:*

1. *$\&$ is order-preserving in both arguments.*
2. *$\swarrow$ and $\nwarrow$ are order-preserving in the consequent and order-reversing in the antecedent.*
3. *$x \leq_1 z \swarrow y$ iff $x \,\&\, y \leq_3 z$ iff $y \leq_2 z \nwarrow x$, where $x \in P_1$, $y \in P_2$ and $z \in P_3$.*

Note that in the domain and codomain of the considered conjunctor we have three (in principle) different sorts, thus providing a more flexible language to a potential user. Furthermore, notice that no boundary condition is required, in difference to the usual definition of multi-adjoint lattice [26] or implication triple [1]. Nevertheless, some boundary conditions follow from the definition, specifically, from the adjoint property (condition (3) above) [23].

ISBN: 978-989-95079-6-8

Lemma 1 *If $(P_1, \leq_1)$, $(P_2, \leq_2)$, $(P_3, \leq_3)$ have bottom element and $(\&, \swarrow, \nwarrow)$ is an adjoint triple, then $(P_1, \leq_1)$ and $(P_2, \leq_2)$ have top element and for all $x \in P_1, y \in P_2$ and $z \in P_3$ the following properties hold:*

1. $\bot_1 \,\&\, y = \bot_3, \qquad x \,\&\, \bot_2 = \bot_3.$
2. $z \nwarrow \bot_1 = \top_2, \qquad z \swarrow \bot_2 = \top_1.$

In order to provide more flexibility into our language, we will allow the existence of several adjoint triples for a given triplet of posets. Notice, however, that since these frames will be used as the underlying structures of our generalization of concept lattice, it is reasonable to require the lattice structure on some of the posets in the definition of adjoint triple.

Definition 2 *A* multi-adjoint frame *$\mathcal{L}$ is a tuple*

$$(L_1, L_2, P, \preceq_1, \preceq_2, \leq, \&_1, \swarrow^1, \nwarrow_1, \ldots, \&_n, \swarrow^n, \nwarrow_n)$$

where $(L_1, \preceq_1)$ and $(L_2, \preceq_2)$ are complete lattices, $(P, \leq)$ is a poset and, for all $i = 1, \ldots, n$, $(\&_i, \swarrow^i, \nwarrow_i)$ is an adjoint triple with respect to L_1, L_2, P.

For short, a multi-adjoint frame will be denoted as $(L_1, L_2, P, \&_1, \ldots, \&_n)$.

Following the usual approach to formal concept analysis, given a frame, a *multi-adjoint context* is a tuple consisting of sets of objects and attributes and a fuzzy relation among them; in addition, the multi-adjoint approach also includes a function which assigns an adjoint triple to each object (or attribute). This feature is important in that it allows for defining subgroups of objects or attributes in terms of different degrees of preference, see [23]. Formally, the definition is the following:

Definition 3 *Let $(L_1, L_2, P, \&_1, \ldots, \&_n)$ be a multi-adjoint frame, a* context *is a tuple (A, B, R, σ) such that A and B are non-empty sets (usually interpreted as attributes and objects, respectively), R is a P-fuzzy relation $R\colon A \times B \to P$ and $\sigma\colon B \to \{1, \ldots, n\}$ is a mapping which associates any element in B with some particular adjoint triple in the frame.*[1]

Once we have fixed a multi-adjoint frame and a context for that frame, we can define the following mappings $^{\uparrow_\sigma}\colon L_2^B \longrightarrow L_1^A$ and $^{\downarrow^\sigma}\colon L_1^A \longrightarrow L_2^B$ which can be seen as generalisations of those given in [4, 17]:

$$g^{\uparrow_\sigma}(a) = \inf\{R(a,b) \swarrow^{\sigma(b)} g(b) \mid b \in B\} \quad (1)$$

$$f^{\downarrow^\sigma}(b) = \inf\{R(a,b) \nwarrow_{\sigma(b)} f(a) \mid a \in A\} \quad (2)$$

These two arrows, $(^{\uparrow_\sigma}, ^{\downarrow^\sigma})$, generate a Galois connection [23]. For the sake of self-containment, this concept is defined below:

Definition 4 *Let $(P_1, \leq_1)$ and $(P_2, \leq_2)$ be posets, and $^{\downarrow}\colon P_1 \to P_2$, $^{\uparrow}\colon P_2 \to P_1$ mappings, the pair $(^{\uparrow}, ^{\downarrow})$ forms a* Galois connection *between P_1 and P_2 whenever the following conditions hold:*

1. *$^{\uparrow}$ and $^{\downarrow}$ are order-reversing.*
2. *$x \leq_1 x^{\downarrow\uparrow}$ for all $x \in P_1$.*
3. *$y \leq_2 y^{\uparrow\downarrow}$ for all $y \in P_2$.*

Proposition 1 ([23]) *Let $(L_1, L_2, P, \&_1, \ldots, \&_n)$ be a multi-adjoint frame and (A, B, R, σ) be a context, then the pair $(^{\uparrow_\sigma}, ^{\downarrow^\sigma})$ is a Galois connection between L_1^A and L_2^B.*

As usual in the different frameworks of formal concept analysis, a *multi-adjoint concept* is a pair $\langle g, f\rangle$ satisfying that $g \in L_2^B$, $f \in L_1^A$ and that $g^{\uparrow_\sigma} = f$ and $f^{\downarrow^\sigma} = g$; with $(^{\uparrow_\sigma}, ^{\downarrow^\sigma})$ being the Galois connection defined above.

Definition 5 *The* multi-adjoint concept lattice *associated to a multi-adjoint frame $(L_1, L_2, P, \&_1, \ldots, \&_n)$ and a context (A, B, R, σ) is the set*

$$\mathcal{M} = \{\langle g, f\rangle \mid g \in L_2^B, f \in L_1^A \text{ and } g^{\uparrow_\sigma} = f, f^{\downarrow^\sigma} = g\}$$

where the ordering is defined by $\langle g_1, f_1\rangle \preceq \langle g_2, f_2\rangle$ if and only if $g_1 \preceq_2 g_2$ (equivalently $f_2 \preceq_1 f_1$).

The ordering just defined above actually provides $\mathcal{M}$ with the structure of a complete lattice [23]. This follows from proposition 1 (the arrows $(^{\uparrow_\sigma}, ^{\downarrow^\sigma})$ forms a Galois connection) and the theorem below.

Theorem 1 ([7]) *Let $(L_1, \preceq_1)$, $(L_2, \preceq_2)$ be complete lattices, let $(^{\uparrow}, ^{\downarrow})$ be a Galois connection between L_1, L_2 and consider $\mathcal{C} = \{\langle x, y\rangle \mid x^{\uparrow} = y, x = y^{\downarrow}; x \in L_1, y \in L_2\}$; then $(\mathcal{C}, \preceq)$ is a complete lattice, where*

$$\bigwedge_{i\in I}\langle x_i, y_i\rangle = \langle \bigwedge_{i\in I} x_i, (\bigvee_{i\in I} y_i)^{\downarrow\uparrow}\rangle;$$

$$\bigvee_{i\in I}\langle x_i, y_i\rangle = \langle (\bigvee_{i\in I} x_i)^{\uparrow\downarrow}, \bigwedge_{i\in I} y_i\rangle$$

and $\langle x_1, y_1\rangle \preceq \langle x_2, y_2\rangle$ if and only if $x_1 \preceq_1 x_2$.

From now on, we will fix a multi-adjoint frame $(L_1, L_2, P, \&_1, \ldots, \&_n)$ and context (A, B, R, σ). Moreover, to improve readability, we will write $(^{\uparrow}, ^{\downarrow})$ instead of $(^{\uparrow_\sigma}, ^{\downarrow^\sigma})$ and $\swarrow^b$, $\nwarrow_b$ instead of $\swarrow^{\sigma(b)}$, $\nwarrow_{\sigma(b)}$.

In the next section, we will present some new properties about the functions α and β involved in the representation (or fundamental) theorem for the multi-adjoint framework presented in [23]. In order to do this, we will recall some necessary definitions.

Definition 6 *Given a complete lattice L, a subset $K \subseteq L$ is* infimum-dense *(resp.* supremum-dense*) if and only if for all $x \in L$ there exists $K' \subseteq K$ such that $x = \inf(K')$ (resp. $x = \sup(K')$).*

A multi-adjoint concept lattice is said to be represented by a complete lattice provided there is a pair of functions, α and β, satisfying the conditions stated in the definition below:

Definition 7 *A multi-adjoint concept lattice*[2] *$(\mathcal{M}, \preceq)$ is* represented *by a complete lattice $(V, \sqsubseteq)$ if there exists a pair of mappings $\alpha\colon A \times L_1 \to V$ and $\beta\colon B \times L_2 \to V$ such that:*

[1] A similar theory could be developed by considering a mapping $\tau\colon A \to \{1, \ldots, n\}$ which associates any element in A with some particular adjoint triple in the frame.

[2] Recall that we are considering a multi-adjoint concept lattice on a fixed frame $(L_1, L_2, P, \&_1, \ldots, \&_n)$ and context (A, B, R, σ).

ISBN: 978-989-95079-6-8

1a) $\alpha[A \times L_1]$ *is infimum-dense;*

1b) $\beta[B \times L_2]$ *is supremum-dense; and*

2) For all $a \in A$, $b \in B$, $x \in L_1$, $y \in L_2$*:*

$$\beta(b, y) \sqsubseteq \alpha(a, x) \quad \textit{if and only if} \quad x \,\&_b\, y \leq R(a, b)$$

From the definition of representability above the following properties follow:

Proposition 2 *Given a complete lattice* $(V, \sqsubseteq)$ *which represents a multi-adjoint concept lattice* $(\mathcal{M}, \preceq)$*, and mappings* $f \in L_1^A$ *and* $g \in L_2^B$*, we have:*

1. β *is order-preserving in the second argument.*
2. α *is order-reversing in the second argument.*
3. $g^{\uparrow}(a) = \sup\{x \in L_1 \mid v_g \sqsubseteq \alpha(a, x)\}$*, where* $v_g = \sup\{\beta(b, g(b)) \mid b \in B\}$.
4. $f^{\downarrow}(b) = \sup\{y \in L_2 \mid \beta(b, y) \sqsubseteq v_f\}$*, where* $v_f = \inf\{\alpha(a, f(a)) \mid a \in A\}$.
5. *If* $g_v(b) = \sup\{y \in L_2 \mid \beta(b, y) \sqsubseteq v\}$*, then* $\sup\{\beta(b, g_v(b)) \mid b \in B\} = v$.
6. *If* $f_v(a) = \sup\{x \in L_1 \mid v \sqsubseteq \alpha(a, x)\}$*, then* $\sup\{\alpha(a, f_v(a)) \mid a \in A\} = v$.

Finally, the fundamental theorem for multi-adjoint concept lattices presented in [23] is the following.

Theorem 2 *A complete lattice* $(V, \sqsubseteq)$ *represents a multi-adjoint concept lattice* $(\mathcal{M}, \preceq)$ *if and only if* $(V, \sqsubseteq)$ *is isomorphic to* $(\mathcal{M}, \preceq)$.

3 New results about the mappings α and β

In this section, we introduce some new interesting properties about the mappings α and β. So, let us assume a complete lattice $(V, \sqsubseteq)$ which represents a multi-adjoint concept lattice $(\mathcal{M}, \preceq)$ and the mappings $\alpha\colon A \times L_1 \to V$, $\beta\colon B \times L_2 \to V$.

We will restate below the isomorphism constructed in fundamental theorem, based on both the α and β functions, since these expressions will be used later.

Proposition 3 ([23]) *If a complete lattice* $(V, \sqsubseteq)$ *represents a multi-adjoint concept lattice* $(\mathcal{M}, \preceq)$*, then there exists an isomorphism* $\varphi\colon \mathcal{M} \to V$ *and two mappings* $\beta\colon B \times L_2 \to V$, $\alpha\colon A \times L_1 \to V$*, such that:*

$$\begin{aligned} \varphi(\langle g, f\rangle) &= \sup\{\beta(b, g(b)) \mid b \in B\} \\ &= \inf\{\alpha(a, f(a)) \mid a \in A\} \end{aligned}$$

for all concept $\langle g, f\rangle \in \mathcal{M}$.

The following result shows continuity-related properties of α and β in their second argument.

Proposition 4 *The mappings* $\beta\colon B \times L_2 \to V$ *and* $\alpha\colon A \times L_1 \to V$ *satisfy that:*

1. *For all indexed set* $Y = \{y_i\}_{i \in I} \subseteq L_2$ *and* $b \in B$*:*

$$\beta(b, \sup\{y_i \mid i \in I\}) = \sup\{\beta(b, y_i) \mid i \in I\}$$

2. *For all indexed set* $X = \{x_i\}_{i \in I} \subseteq L_1$ *and* $a \in A$*:*

$$\alpha(a, \sup\{x_i \mid i \in I\}) = \inf\{\alpha(a, x_i) \mid i \in I\}$$

Proof: 1. Consider $b \in B$ and $Y = \{y_i\}_{i \in I} \subseteq L_2$, as $\alpha[A \times L_1]$ is infimum-dense and $\beta(b, \sup Y) \in V$, there exists an indexing set Λ such that $\beta(b, \sup Y) = \inf\{\alpha(a_j, x_j) \mid j \in \Lambda\}$; as a result $\beta(b, \sup Y) \sqsubseteq \alpha(a_j, x_j)$ for every $j \in \Lambda$. From proposition 2(1), we obtain that $\beta(b, y_i) \sqsubseteq \alpha(a_j, x_j)$, for every $i \in I$ and $j \in \Lambda$, and hence $\sup\{\beta(b, y_i) \mid i \in I\} \sqsubseteq \alpha(a_j, x_j)$ for every $j \in \Lambda$, then

$$\begin{aligned} \sup\{\beta(b, y_i) \mid i \in I\} &\sqsubseteq \inf\{\alpha(a_j, x_j) \mid j \in \Lambda\} \\ &= \beta(b, \sup Y) \end{aligned}$$

For the other inequality, let us consider $\sup\{\beta(b, y_i) \mid i \in I\}$ and, as $\alpha[A \times L_1]$ is infimum-dense, there exists an indexing set Λ' such that $\sup\{\beta(b, y_i) \mid i \in I\} = \inf\{\alpha(a_j, x_j) \mid j \in \Lambda'\}$. Now, for all $i \in I$ and $j \in \Lambda'$ we obtain that $\beta(b, y_i) \sqsubseteq \alpha(a_j, x_j)$, therefore, from Definition 7(2), $x_j \,\&_b\, y_i \leq R(a_j, b)$. Now, as $(\&_b, \swarrow^b, \nwarrow_b)$ is an adjoint triple, we have the following chain of equivalent statements:

$$\begin{aligned} x_j \,\&_b\, y_i &\leq R(a_j, b) && \text{for all } i \in I \\ y_i &\preceq_2 R(a_j, b) \nwarrow_b x_j && \text{for all } i \in I \\ \sup Y &\preceq_2 R(a_j, b) \nwarrow_b x_j \\ x_j \,\&_b \sup Y &\leq R(a_j, b) \end{aligned}$$

so, $\beta(b, \sup Y) \sqsubseteq \alpha(a_j, x_j)$ for every $j \in \Lambda'$, and thus

$$\begin{aligned} \beta(b, \sup Y) &\sqsubseteq \inf\{\alpha(a_j, x_j) \mid j \in \Lambda'\} \\ &= \sup\{\beta(b, y_i) \mid i \in I\} \end{aligned}$$

2. This proof is analogous using that $\beta[B \times L_2]$ is supremum-dense. □

We continue below by proving some boundary conditions fulfilled by α and β.

Proposition 5 *The two mappings* $\alpha\colon A \times L_1 \to V$ *and* $\beta\colon B \times L_2 \to V$ *are such that* $\alpha(a, \bot_1) = \top_V$ *and* $\beta(b, \bot_2) = \bot_V$ *for all* $b \in B$ *and* $a \in A$.

Proof: Given $a \in A$, let us prove that $\alpha(a, \bot_1) = \top_V$. Firstly, recall that lemma 1 implies that $\bot_1 \,\&_b\, y \leq R(a, b)$ for all $b \in B$ and $y \in L_2$; now, from Definition 7(2) we obtain that $\beta(b, y) \sqsubseteq \alpha(a, \bot_1)$ for all $b \in B$ and $y \in L_2$, that is, $\alpha(a, \bot_1)$ is an upper bound of the set of elements $\beta(b, y)$ for all $b \in B$ and $y \in L_2$. Now, as β is supremum-dense, there is an indexing set Λ such that $\top_V = \sup\{\beta(b_i, y_i) \mid i \in \Lambda\}$, therefore, we have that: $\top_V \sqsubseteq \alpha(a, \bot_1)$. Hence, $\top_V = \alpha(a, \bot_1)$.

The other equality follows similarly. □

From the propositions above, we obtain the following corollary which states the behaviour of α and β regarding suprema of any set (either empty or non-empty).

Corollary 1 *The mappings* $\beta\colon B \times L_2 \to V$, $\alpha\colon A \times L_1 \to V$ *satisfy that:*

ISBN: 978-989-95079-6-8

1. $\beta(b, \sup Y) = \sup\{\beta(b, y) \mid y \in Y\}$, *for all* $Y \subseteq L_2$ *and* $b \in B$.

2. $\alpha(a, \sup X) = \inf\{\alpha(a, x) \mid x \in X\}$, *for all* $X \subseteq L_1$ *and* $a \in A$.

As a consequence of the property above we have the following result, which gives us a more efficient form to write Property (2) in Definition 7 to check if a lattice is isomorphic to a concept lattice, that is, in order to apply Theorem 2.

Proposition 6 *Given* $a \in A$, $b \in B$, *the applications* $\beta_b: L_2 \to V$, $\alpha_a: L_1 \to V$ *have residuated mappings, that is, there exist* $\beta'_b: L_2 \to V$, $\alpha'_a: L_1 \to V$ *such that:*

$$\begin{array}{lll} \beta_b(y) \sqsubseteq v & \textit{if and only if} & y \preceq_2 \beta'_b(v) \\ v \sqsubseteq \alpha_a(x) & \textit{if and only if} & x \preceq_1 \alpha'_a(v) \end{array}$$

for all $x \in L_1$, $y \in L_2$ *and* $v \in V$.

Proof: If we define $\beta'_b(v) = \sup\{y \in L_2 \mid \beta_b(y) \sqsubseteq v\}$ and, similarly, $\alpha'_a(v) = \sup\{x \in L_1 \mid v \sqsubseteq \alpha_a(x)\}$, we obtain the result straightforward from Corollary 1. □

The following proposition states a necessary and sufficient condition for the mappings α and β to fulfill the second condition in the definition of representable lattice.

Proposition 7 *The mappings* α *and* β *satisfy Property (2) in Definition 7 if and only if, for all* $a \in A$, $b \in B$, $x \in L_1$, $y \in L_2$, *any of the following equalities holds:*

$$\begin{array}{lcl} \beta'_b(\alpha_a(x)) & = & R(a,b) \nwarrow x \\ \alpha_a(\beta'_b(y)) & = & R(a,b) \swarrow y \end{array}$$

Proof: Firstly, we assume that the mappings α and β satisfy Property (2) in Definition 7. The first equality is given from the following chain of equivalences, given $a \in A$, $b \in B$, $x \in L_1$, $y \in L_2$:

$$\begin{array}{lcl} y \preceq_2 \beta'_b(\alpha_a(x)) & \iff & \beta(b,y) \sqsubseteq \alpha(a,x) \\ & \iff & x \,\&_b\, y \leq R(a,b) \\ & \iff & y \preceq_2 R(a,b) \nwarrow x \end{array}$$

if we substitute y by $R(a,b) \nwarrow x$ in the first sentence and y by $\beta'_b(\alpha_a(x))$ in the last one. The second equality follows similarly.

Now, we assume that $\beta'_b(\alpha_a(x)) = R(a,b) \nwarrow x$, hence

$$\begin{array}{lcl} \beta(b,y) \sqsubseteq \alpha(a,x) & \iff & y \preceq_2 \beta'_b(\alpha_a(x)) \\ & \overset{(*)}{\iff} & y \preceq_2 R(a,b) \nwarrow x \\ & \iff & x \,\&_b\, y \leq R(a,b) \end{array}$$

where $(*)$ is given from the hypothesis. □

As a result of the previous proposition, we obtain a straightforward mechanism to obtain the mappings α and β in order to check whether a lattice is isomorphic to a concept lattice.

Finally, the following property shows that any subset of $A \times L_1$ or of $B \times L_2$ is related to a concept via α and φ, or β and φ, respectively.

Proposition 8 *Consider a multi-adjoint concept lattice* $(\mathcal{M}, \preceq)$ *represented by a complete lattice* $(V, \sqsubseteq)$ *and the mappings* $\alpha: A \times L_1 \to V$, $\beta: B \times L_2 \to V$, *then for each* $K \subseteq A \times L_1$, *there exists a unique concept* $\langle g, f\rangle \in \mathcal{M}$ *such that*

$$\inf\{\alpha(a,x) \mid (a,x) \in K\} \quad = \quad \varphi(\langle g, f\rangle)$$

Analogously, for each $K' \subseteq B \times L_2$, *there exists a unique concept* $\langle g, f\rangle \in \mathcal{M}$ *such that*

$$\sup\{\beta(b,y) \mid (b,y) \in K'\} \quad = \quad \varphi(\langle g, f\rangle)$$

Proof: Given $K \subseteq A \times L_1$, let us consider the sets $K_a = \{x \mid (a,x) \in K\}$, and the function $h: A \to L_1$ defined as $h(a) = \sup K_a$.

By Corollary 1, we have that, for all $a' \in A$, the following equality holds

$$\alpha(a', h(a')) = \inf\{\alpha(a', x) \mid x \in K_a\}$$

Therefore:

$$\begin{array}{l} \inf\{\alpha(a', h(a')) \mid a' \in A\} = \\ \qquad = \quad \inf\{\inf\{\alpha(a', x) \mid x \in K_{a'}\} \mid a' \in A\} \\ \qquad = \quad \inf\{\alpha(a', x) \mid (a', x) \in K\} \end{array}$$

Finally, we obtain the following chain of equalities:

$$\begin{array}{lcl} \inf\{\alpha(a,x) \mid (a,x) \in K\} & = & \inf\{\alpha(a', h(a')) \mid a' \in A\} \\ & \overset{(1)}{=} & \sup\{\beta(b', h^{\downarrow}(b')) \mid b' \in B\} \\ & \overset{(2)}{=} & \varphi(\langle h^{\downarrow}, h^{\downarrow\uparrow}\rangle) \end{array}$$

where (1) follows by Proposition 2, (2) by Proposition 3. This means that the concept whose existence is postulated in the statement is $\langle h^{\downarrow}, h^{\downarrow\uparrow}\rangle$.

Now, the uniqueness follows from the isomorphism φ:
If there would exist another concept $\langle g, f\rangle$ such that $\sup\{\beta(b, g(b)) \mid b \in B\} = \inf\{\alpha(a,x) \mid (a,x) \in K\}$, we would have:

$$\begin{array}{lcl} \varphi(\langle h^{\downarrow}, h^{\downarrow\uparrow}\rangle) & = & \sup\{\beta(b, h^{\downarrow}(b)) \mid b \in B\} \\ & = & \inf\{\alpha(a,x) \mid (a,x) \in K\} \\ & = & \sup\{\beta(b, g(b)) \mid b \in B\} \\ & = & \varphi(\langle g, f\rangle) \end{array}$$

Thus, $\langle h^{\downarrow}, h^{\downarrow\uparrow}\rangle = \langle g, f\rangle$.

The second statement follows similarly. □

4 Conclusions

The representation theorem is one of the most important results in the theory of formal concept analysis, since it provides conditions in order to determine whether a given lattice is isomorphic to some concept lattice. In this paper, an analytic expression for the mappings α and β involved in the representation theorem of t-concept lattices is provided, together with some interesting properties.

ISBN: 978-989-95079-6-8

References

[1] A. Abdel-Hamid and N. Morsi. Associatively tied implicacions. *Fuzzy Sets and Systems*, 136(3):291–311, 2003.

[2] C. Alcalde, A. Burusco, R. Fuentes-González, and I. Zubia. Treatment of L-fuzzy contexts with absent values. *Information Sciences*, 179:1–15, 2009.

[3] R. Belohlávek. Lattice generated by binary fuzzy relations (extended abstract). In *4th Intl. Conf. on Fuzzy Sets Theory and Applications*, page page 11, 1998.

[4] R. Belohlávek. Concept lattices and order in fuzzy logic. *Annals of Pure and Applied Logic*, 128:277–298, 2004.

[5] R. Belohlávek. Lattices of fixed points of fuzzy Galois connections. *Mathematical Logic Quartely*, 47(1):111–116, 2004.

[6] A. Burusco and R. Fuentes-González. The study of L-fuzzy concept lattice. *Mathware & Soft Computing*, 3:209–218, 1994.

[7] B. Davey and H. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, second edition, 2002.

[8] P. du Boucher-Ryana and D. Bridge. Collaborative recommending using formal concept analysis. *Knowledge-Based Systems*, 19(5):309–315, 2006.

[9] S.-Q. Fan, W.-X. Zhang, and W. Xu. Fuzzy inference based on fuzzy concept lattice. *Fuzzy Sets and Systems*, 157(24):3177–3187, 2006.

[10] A. Formica. Ontology-based concept similarity in formal concept analysis. *Information Sciences*, 176(18):2624–2641, 2006.

[11] A. Formica. Concept similarity in formal concept analysis: An information content approach. *Knowledge-Based Systems*, 21(1):80–87, 2008.

[12] B. Ganter and R. Wille. *Formal Concept Analysis: Mathematical Foundation*. Springer Verlag, 1999.

[13] G. Georgescu and A. Popescu. Concept lattices and similarity in non-commutative fuzzy logic. *Fundamenta Informaticae*, 53(1):23–54, 2002.

[14] G. Jiang, K. Ogasawara, A. Endoh, and T. Sakurai. Context-based ontology building support in clinical domains using formal concept analysis. *International Journal of Medical Informatics*, 71(1):71–81, 2003.

[15] P. Julian, G. Moreno, and J. Penabad. On fuzzy unfolding: A multi-adjoint approach. *Fuzzy Sets and Systems*, 154(1):16–33, 2005.

[16] S. Krajci. The basic theorem on generalized concept lattice. In V. Snásel and R. Belohlávek, editors, *International Workshop on Concept Lattices and their Applications, CLA 2004*, pages 25–33, 2004.

[17] S. Krajci. A generalized concept lattice. *Logic Journal of IGPL*, 13(5):543–550, 2005.

[18] S. O. Kuznetsov. Complexity of learning in concept lattices from positive and negative examples. *Discrete Applied Mathematics*, 142:111–125, 2004.

[19] Y. Lei and M. Luo. Rough concept lattices and domains. *Annals of Pure and Applied Logic*, 2009. Article in press (http://dx.doi.org/10.1016/j.apal.2008.09.028).

[20] M. Liu, M. Shao, W. Zhang, and C. Wu. Reduction method for concept lattices based on rough set theory and its application. *Computers & Mathematics with Applications*, 53(9):1390–1410, 2007.

[21] X. Liu, W. Wang, T. Chai, and W. Liu. Approaches to the representations and logic operations of fuzzy concepts in the framework of axiomatic fuzzy set theory I. *Information Sciences*, 177(4):1007–1026, 2007.

[22] X. Liu, W. Wang, T. Chai, and W. Liu. Approaches to the representations and logic operations of fuzzy concepts in the framework of axiomatic fuzzy set theory II. *Information Sciences*, 177(4):1027–1045, 2007.

[23] J. Medina, M. Ojeda-Aciego, and J. Ruiz-Calviño. Formal concept analysis via multi-adjoint concept lattices. *Fuzzy Sets and Systems*, 160(2):130–144, 2009.

[24] J. Medina, M. Ojeda-Aciego, and J. Ruiz-Calviño. On multi-adjoint concept lattices: denition and representation theorem. *Lect. Notes in Artificial Intelligence*, 4390:197–209, 2007.

[25] J. Medina, M. Ojeda-Aciego, A. Valverde, and P. Vojtáš. Towards biresiduated multi-adjoint logic programming. *Lect. Notes in Artificial Intelligence*, 3040:608–617, 2004.

[26] J. Medina, M. Ojeda-Aciego, and P. Vojtáš. Similarity-based unification: a multi-adjoint approach. *Fuzzy Sets and Systems*, 146:43–62, 2004.

[27] V. Phan-Luong. A framework for integrating information sources under lattice structure. *Information Fusion*, 9:278–292, 2008.

[28] S. Pollandt. *Fuzzy Begriffe*. Springer, Berlin, 1997.

[29] K.-S. Qu and Y.-H. Zhai. Generating complete set of implications for formal contexts. *Knowledge-Based Systems*, 21:429–433, 2008.

[30] M.-W. Shao, M. Liu, and W.-X. Zhang. Set approximations in fuzzy formal concept analysis. *Fuzzy Sets and Systems*, 158(23):2627–2640, 2007.

[31] L. Wang and X. Liu. Concept analysis via rough set and afs algebra. *Information Sciences*, 178(21):4125–4137, 2008.

[32] X. Wang and W. Zhang. Relations of attribute reduction between object and property oriented concept lattices. *Knowledge-Based Systems*, 21(5):398–403, 2008.

[33] Q. Wu and Z. Liu. Real formal concept analysis based on grey-rough set theory. *Knowledge-Based Systems*, 22(1):38–45, 2009.

ISBN: 978-989-95079-6-8

A Neuro-fuzzy System for Fraud Detection in Electricity Distribution

C. Muniz, M. Vellasco, R. Tanscheit, K. Figueiredo

Computational Intelligence Lab., Department of Electrical Engineering,
Pontifical Catholic University of Rio de Janeiro
Rio de Janeiro, Brazil
Email: cyroassis@gmail.com, {marley, ricardo, karla}@ele.puc-rio.br

Abstract *— The volume of energy loss that Brazilian electrical utilities have to deal with has been ever increasing. Electricity distribution companies have suffered significant and increasing losses in the last years, due to theft, measurement errors and other irregularities. Therefore there is a great concern to identify the profile of irregular customers, in order to reduce the volume of such losses. This paper presents a combined approach of a neural networks committee and a neuro-fuzzy hierarchical system intended to increase the level of accuracy in the identification of irregularities among low voltage consumers. The data used to test the proposed system are from Light S.A., the distribution company of Rio de Janeiro. The results obtained presented a significant increase in the identification of irregular customers when compared to the current methodology employed by the company.*

Keywords— neural nets, hierarchical neuro-fuzzy systems, binary space partition, electricity distribution, fraud detection.

1 Introduction

Commercial electricity loss due to irregularities (energy theft and measurement errors) reach about R$ 5 billion a year in Brazil, which is around 5% of the total amount spent in energy consumption. A great part of that loss occurs in the area operated by Light S.A.: 3.8 million customers in 28 cities in the state of Rio de Janeiro. Electricity bills are, in some cases, up to 17% more expensive than they should be, so as to compensate for energy thefts and other fraudulent practices [1].

The company makes use of heuristics to indicate a set of low voltage customers who may practice some kind of irregularity. This set of customers, initially classified as suspects, is then analyzed by an expert, who selects a certain number to be directly inspected. The inspection is performed by an employee that visits each residence to confirm (or not) the irregularity. Based on this simple approach, Light S.A. has reached a low identification rate. It is evident that the adopted procedure is not effective and that a more accurate system must be developed.

Some researches have already tackled this problem, confirming the difficulty of correctly identifying irregular consumption. In [2], Rough Sets have been used to fraud detection among electrical energy consumers. The results obtained are promising but accuracy is still low (around 20%) due to the presence of noisy information in the company's database. In [3] and [4] a similar approach is presented, where only past consumption information is considered in the analysis. Due to the temporal nature of the data, time series analysis methods [5] are applied in order to get new invariant characteristics. However, other relevant variables, such as information about the measurement device and the average temperature are not considered.

This paper describes a new intelligent methodology, based on neural networks and fuzzy systems, that takes into consideration historical consumption as well as other important variables. The main objective is to increase the accuracy level when identifying irregular low voltage consumers, selected from a group of suspicious clients.

The proposed methodology is composed of two basic modules: Filtering and Classification, the former comprising an ensemble of five artificial neural networks, while the latter consists of a Neuro-fuzzy Hierarchical System [6]. In the Filtering module, each neural network indicates whether customers are suspected of irregularities or not. As in any ensemble-based methodology, the use of a committee of neural networks aims at improving accuracy by producing a consensus decision that is potentially more accurate than individual neural networks [7]. The Filtering Module was developed to better extract actual normal and irregular customers for training the Classification Module. This module is introduced in the expectation of improving even further the detection of irregularities. It should be noted that the database contains untrustworthy data, due to incorrect information filled in by some visiting employees.

The paper is organized in four additional sections: Section 2 describes the current methodology used by Light S.A. for detecting irregularities, Section 3 describes the proposed methodology, Section 4 presents the case studies considered, and, finally, Section 5 concludes the work.

2 Current Methodology for Fraud Detection

Currently, Light S.A. makes use of diverse methodologies for selecting low voltage customers that present irregularity indications. The *Quarterly Indicator* methodology consists of comparing the past three months' consumption. The *Annual Indicator* consists of evaluating clients' consumption during the last 24 months. The *Adjustment Factor* methodology compares last month consumption to that in the corresponding month of the preceding year. Finally, the *Tendency Factor* methodology consists of comparing clients' previous month's consumption with that expected in the present month. In all cases, if the reduction in consumption is higher than expected, the client is considered a suspect.

All customers identified by those methodologies form

ISBN: 978-989-95079-6-8

the set of suspects, which is analyzed by an expert; some are then recommended for direct inspection. Through this methodology, Light has attained an average *Positive Predictive Value* (PPV) of 25%, which is the proportion of proven irregular customers among all those who have been classified as suspects of some irregularity and have been inspected. Table 1 shows a typical confusion matrix [8] for a two-class problem (*normal* and *irregular* consumers).

Table 1: Suspected x Inspected Customers

System Output	Inspection Result	
	Irregular	Normal
Irregular	A	B
Normal	C	D

The PPV is computed by:

$$PPV = A / (A + B)$$

3 Proposed Methodology

The methodology proposed in this work consists of three basic modules: Pre-Processing and Normalization, Filtering and Classification.

The whole methodology is shown in Fig. 1. Fig. 1a shows the system's configuration used during the learning phase, which contains all three modules; Fig. 1b, on the other hand, presents a simpler configuration, used for the recall phase, which does not include the Filtering Module.

The first module, present in both training and recall phases and called Pre-processing and Normalization, is where all attributes are selected, normalized and coded.

It should be pointed out, however, that the available inspection database is, unfortunately, very noisy, containing unreliable information due to (intentional or not) incorrect indication of an irregular consumer as a normal one during inspection. Therefore, to create a more reliable database for training the Classification Module, a Filtering Module was conceived.

The Filtering Module makes use of a committee of five neural networks, which select *irregular* and *normal* consumers that better characterize the two different classes. The resulting filtered database is then employed for training the Classification Module, which eventually identifies a customer as *normal* or *irregular*. Therefore, future classifications of customers (recall phase) will employ only the Pre-processing and Classification modules. The Filtering Module is only necessary if the system needs to be retrained.

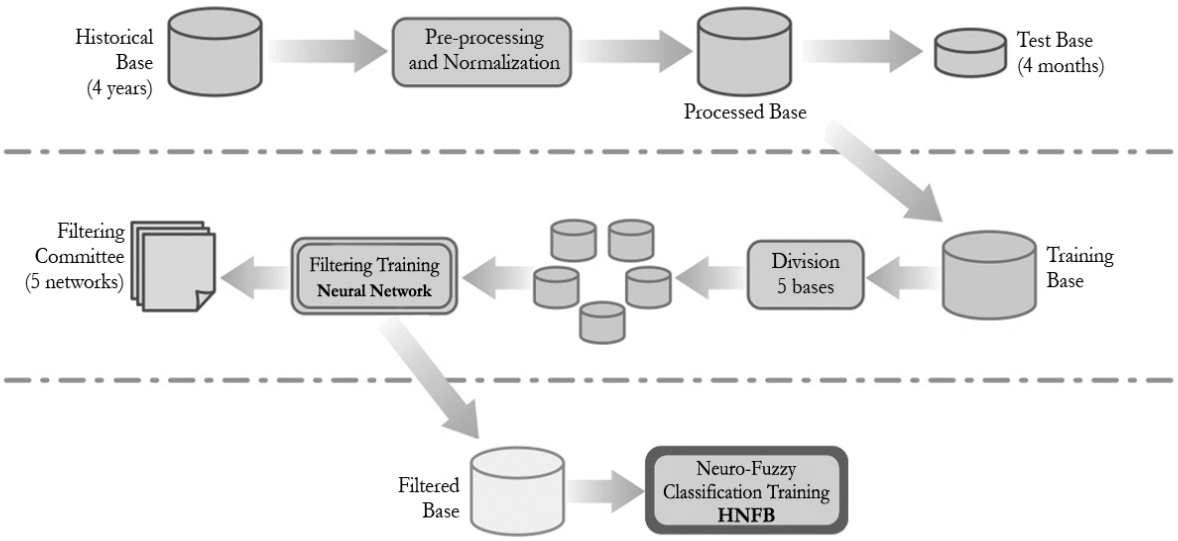


Figure 1(a): Complete methodology of the Learning Phase – Pre-processing, Filtering and Classification

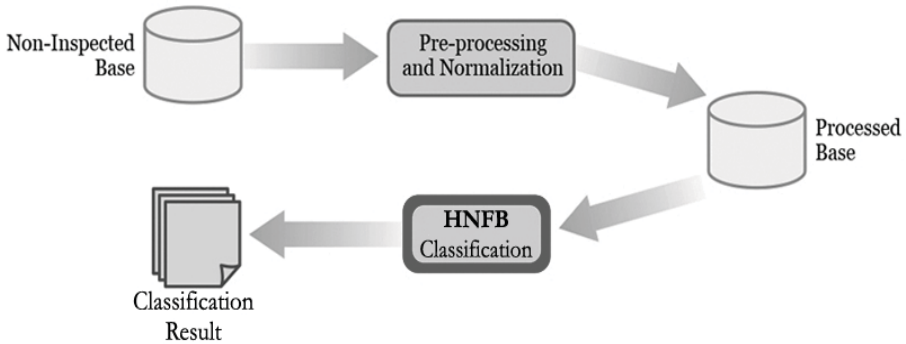


Figure 1(b): Methodology for the Recall Phase - Pre-processing and Classification modules.

ISBN: 978-989-95079-6-8

The following sub-sections present in more detail each of the modules.

3.1 Pre-Processing and Normalization Module

The data pre-processing module is composed of the following phases: selection of attributes, data cleaning and normalization. The database contains general information on both commercial/industrial and residential consumers in the so-called East Regional of Rio de Janeiro. It is well known that the behaviours of residential and non-residential clients are very distinct. Therefore, the database was divided into residential and non-residential clients, resulting in a specific Module for each class of clients.

After the evaluation of the entire available attributes, those which refer to customers' consumption for the last 4 years were selected, as well as some attributes that characterize the installation made at the consumer unit. Table 2 presents the attributes selected in this phase.

Table 2: Attributes selected from the database

Attribute	Description
Local	Consumer Unity Identification code
Inspection date	Date when the inspection was made
Related Month	Month used as a reference. Previous month to the inspection.
Origin	Main reason for the inspection
Measurer	Measuring device identifier
MonoBiTri	Indicates if the installation is mono, bi or triphasic
Consumption irregularity code	Last detected consumption irregularity
Reading irregularity code	Last confirmed reading irregularity
Month consumption	Consumer unity consumption during the related month
Previous year consumption	Registered consumption one year before the related month
Quarterly Indicator	Variation in the past three months' consumption, one by one
Adjustment Factor	Variation from the current month consumption to the same month from the previous year
Annual Indicator	Clients' consumption during the last 24 months
Tendency Factor	Clients' previous month's consumption with the consumption expected in the present month
Fraud indicator	Indicates whether the inspector found an irregularity or not

The first three attributes in the database identify, respectively, the installation where the inspection was made, the date of inspection and the month used as a reference for data acquisition. Those three attributes are presented only to identify consumer registers and are not used as inputs to the system. The five following attributes describe in detail the technical aspects of the installation and the customer's behavior. These are nominal attributes, coded as follows:

Origin – 60 different origins, divided into 4 classes (encoded in 2 binary inputs):

1. Intelligence: information extracted from the database
2. System: automatically identified information by the system of Light S.A
3. Report: complaints made by other customers or readers
4. Grouping: inspection without fraud suspicion

Measuring device – 90 models of measuring devices, grouped in 6 classes (encoded in 3 binary inputs):

1. MMD – direct mechanical measurer
2. OMMD – direct mechanical measurer (obsolete)
3. MMI – indirect mechanical measurer
4. OMMI – indirect mechanical measurer (obsolete)
5. MED – direct electronic measurer
6. OMED – direct electronic measurer (obsolete)

MonoBiTri – three different classes: monophasic, biphasic and triphasic (encoded by 2 binary inputs).

Reading Irregularity Code and *Consumption Irregularity Code* – coded in such a way that if one customer presents any kind of irregularity, his attribute will be 1; otherwise, it is zero.

The next six attributes – *consumption in a given month*, *previous year consumption* and the indicators explained below – define the consumption profile of the customer. Since these are numerical attributes, they have been normalized through:

$$value_{NORM} = \frac{v_{MAX} - value}{v_{MAX} - v_{MIN}}$$

where *value* is the attribute to be normalized, $value_{NORM}$ is the normalized attribute, v_{MAX} is the maximum attribute value registered in the last two years and v_{MIN} is the minimum attribute value registered in the last 2 years.

Four additional indicators have been computed from the monthly consumption information of each customer: *three-months moving average*, *six-months moving average*, *last 12 months average* and the *previous year average* (from the 24th previous month to the 12th previous month).

The *Fraud Indicator* attribute provides the result of the inspection, and indicates if any irregularity was found. This attribute was used as output (target) in training.

Finally, the following variables were also added to the set of attributes: *minimum* and *maximum monthly temperatures*, according to the customer's geographic area, and the *amount of consumption of all customers in the month* under analysis.

All data have been filtered (spurious data, redundant and incomplete were eliminated) and normalized [9]. Through normalization all attributes present the same interval of variation, which is important for training.

ISBN: 978-989-95079-6-8

3.2 Filtering Module

As already mentioned, the available database is rather noisy, due to some unreliable information in the Fraud Indicator attribute. In order to extract patterns from the database that are more reliable examples of actual *normal* and *irregular* customers, the Filtering Module was proposed. This consists of a committee of five neural network, all of them Multi-Layer Perceptrons (MLP) [10][11], with 21 inputs, one hidden layer and one neuron in the output layer, with sigmoidal threshold function (Fig. 2). The output neuron represents two classes: *irregular* customers (measurement or technical irregularities) and *normal* customers (those who do not show any irregularity).

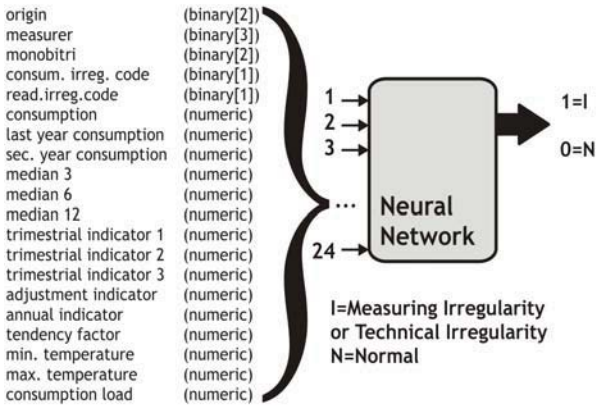


Figure 2: MLP Neural Network topology with one output

As the network output is a continuous value between zero and one, a threshold level of 0.5 was defined so as to distinguish *irregular* (output value ≥ 0.5) from *normal* customers (output value < 0.5). It should be mentioned that different threshold limits, between [0.3, 0.8] with steps of 0.05 were also evaluated, but the best performance has been achieved for 0.5.

The complete database is divided into five disjoint sets in order to train each of the networks that form the committee. After training, the complete database is processed by the neural networks. The selection of more reliable training patterns (filtering process) is carried out as follows: a pattern that is labeled as *irregular* in the dataset and is indicated by three or more votes in the committee evaluation as also *irregular* is taken as a positive true pattern, that is, a customer that actually characterizes an *irregular* pattern. On the other hand, patterns specified as *normal* in the database and indicated by at least three votes from the committee as being *normal* are called negative true patterns. These patterns are then considered as reliable examples of normal customers (see Fig. 3).

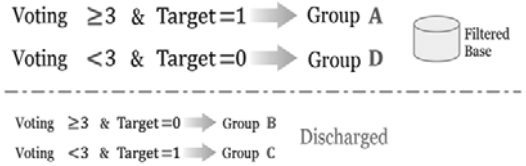


Figure 3: Limits for the Filtering Committee

A new filtered database is then formed, containing all positive and negative true patterns, which will then be used for training the Classification Module. The positively false registers (those normal patterns that receive three or more votes as being irregular from the committee) and the negatively false registers (irregular patterns that receive at least three votes as being normal) are discarded as noisy patterns and are not used to train the Classification Module.

3.3 Classification Module

As mentioned above, the main difference between this and the filtering module is the database used for training. Now only the filtered database, obtained from the filtering module, is used to adjust the classifier parameters. Once training is completed, the Classification Module is ready for use with new input patterns. In this phase, the Filtering Module is no longer necessary (see Fig. 1b), and will only be used again if retraining becomes necessary.

The structure used in this stage is a Hierarchical Neuro-fuzzy with Binary Space Partition (BSP) system [6] [12], briefly described below.

Hierarchical Neuro-Fuzzy Systems have been devised to overcome limitations of traditional neuro-fuzzy systems, which in general have a limited capacity for creating their own structure and rules. Additionally, most of the models employ grid partition of the input space, which, due to the rule explosion problem, are more adequate for applications with a smaller number of inputs. When a greater number of input variables are necessary, the system's performance deteriorates.

The Neuro-fuzzy Hierarchical BSP (HNFB) system used here makes use of basic cells. An HNFB cell is a neuro-fuzzy mini-system that performs fuzzy binary partitioning of the input space. The HNFB cell generates a crisp output after a defuzzification process. Fig. 4 illustrates the cell's functionality, where x represents the input variable, $\rho(x)$ and $\mu(x)$ are the membership functions *low* and *high*, respectively, which generate the antecedents of the two fuzzy rules, and y is the output. The linguistic interpretation of the mapping implemented by the HNFB cell is given by the following rules:

- If $x \in \rho$ then $y = d_1$
- If $x \in \mu$ then $y = d_2$

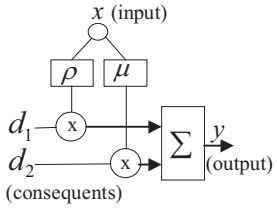


Figure 4: Neuro-Fuzzy BSP cell

Each rule corresponds to one of the two partitions generated by BSP. Each partition can in turn be subdivided into two parts by means of another HNFB

ISBN: 978-989-95079-6-8

cell. The profiles of membership functions $\rho(x)$ and $\mu(x)$ are complementary logistic functions.

The output y of an HNFB cell (defuzzification process) is given by the weighted average. Due to the fact that the membership function $\rho(x)$ is the complement to 1 of $\mu(x)$, the following equation applies:

$$y = \rho(x) * d_1 + \mu(x) * d_2 \quad \text{or} \quad y = \sum_{i=1}^{2} \alpha_i d_i$$

where α_i symbolizes the firing level of the rule in partition i and are given by $\alpha_1=\rho(x)$ and $\alpha_2=\mu(x)$. Each d_i corresponds to one of the possible consequents below:

- A singleton: d_i = constant.
- A linear combination of inputs: $d_i = \sum_{k=1}^{n} w_k x_k + w_0$

  where: x_k is the k-th input, w_k represent the weight associated with input x_k, n is the total number of inputs and w_0 is a constant value.
- The output of a generic cell j: $d_i = y_j$

The HNFB model may be described as a system that is made up of interconnections of HNFB cells, as illustrated in Fig. 5, along with the respective partitioning of the input space. In the system presented in Fig. 5, the initial partitions 1 and 2 ('BSP0' cell) have been subdivided; hence, the consequents of its rules are the outputs of BSP1 and BSP2, respectively. In turn, these subsystems have, as consequents, values d_{11}, y_{12}, d_{21} and d_{22}, respectively. Consequent y_{12} is the output of the 'BSP12' cell. The output of the system in Fig. 3(b) is:

$$y = \alpha_1(\alpha_{11}d_{11} + \alpha_{12}(\alpha_{121}d_{121} + \alpha_{122}d_{122})) + \alpha_2(\alpha_{21}d_{21} + \alpha_{22}d_{22})$$

Although each BSP cell divides the input space only into two fuzzy sets (*low* and *high*), the complete HNFB architecture divides the universe of each variable into as many partitions as necessary. The number of partitions is determined during the learning process. In Fig. 5, for instance, the upper left part of the input space (partition 12) has been further subdivided by the horizontal variable x_1, resulting in three fuzzy sets for the complete universe of discourse for this specific variable.

The training algorithm makes use of the gradient descent method for learning the structure of the model and, consequently, linguistic rules. The parameters that define the the membership functions of the antecedents and consequents are regarded as fuzzy weights of the neuro-fuzzy system.

A tuning parameter δ, called decomposition rate prevents the structure from growing indefinitely.

4 Experimental Results

This section presents the results obtained for Filtering, the Neural Networks Committee, and Classification, performed by the HNFB system.

4.1 Filtering

After processing the selected data three databases were obtained. These were composed of 4159 customers with non technical irregularity, 3754 customers with technical irregularity and 14405 normal customers.

Data were divided into Training, Validation and Testing databases. Training and Validation sets refer to the period between September 2002 and June 2006, with the exception of February and March 2006. The data for these two months, together with those for July and August 2006 were used to form the test set. These months were selected to evaluate the system's performance in different seasons (summer and winter in the southern hemisphere). Due to the unbalanced characteristic of the database, with many more examples of *normal* customers than the available number of *irregular* patterns (technical and non-technical irregularities), the training and validation sets were equalized to avoid a biased behaviour of the neural networks. The datasets actually used in the experiments consisted of 3754 customers with technical irregularities, 3754 randomly selected customers with non-technical irregularities and 7508 normal ones, also randomly chosen. For training, 75% of the balanced samples have been considered; the rest was left for validation.

Five different databases were created, each to train one of the five neural networks that form the filtering committee. For each neural network, 10 training processes were realized and the best performing network from the validation set has been selected to form the committee. The early stopping methodology was used during each training process, which specifies the optimal number of epochs to avoid over fitting and attain best generalization performance.

As the main objective of the proposed system was to increase the PPV of irregularity detection, two experiments were carried out: one considering the PPV as the error metric for the neural networks training, and another considering the general classification error as the metric used to select the best neural network. It was observed that the latter provided better results, since when the classification error for both classes is minimized, the PPV is indirectly maximized. Thus, in the following experimental results, the general classification error metric was used.

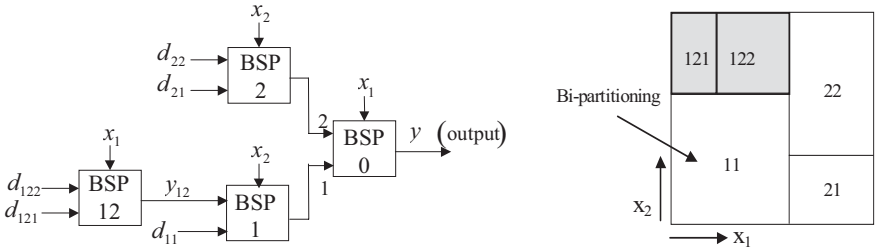


Figure 5: Example of an HNFB system and respective input space Partitioning

ISBN: 978-989-95079-6-8

After all neural networks had been trained and the committee members had been selected, a test was carried out with the data of customers investigated during the months of February, March, July and August 2006. That test base has 4663 registers of inspected customers.
Table 3 illustrates filtering results for one specific ZEI (Elementary Intervention Zones) of East Regional, in the same fashion of Table 1.

Table 3: Results with the filtering committee

	Committee	
	Irregular	Normal
Irregular	960	1164
Normal	572	1970
	PPV_{NN}:	45.2%
	Class. Error:	37.2%
	PPV_{Light}:	32.8%

As can be observed, by using only the Filtering Module the PPV has already increased from 32.8% to 45.2%.

4.2 Classification

The same inputs considered for filtering were initially considered for classification, performed by the HNFB system. After variable selection by the least-square estimator (LSE) [13], nine were effectively used as inputs to the system. The test database was identical to that used for testing in the filtering stage.

It can be seen from the classification results shown in Table 4 that the VPP has improved even further, reaching 51.2%. This attests the advantage of the proposed methodology.

Table 4: Results for classification

	Committee	
	Irregular	Normal
Irregular	991	943
Normal	541	2191
	PPV_{HNFB}:	51.2%
	Class. Error:	31.8%
	PPV_{Light}:	32.8%

In addition to performing classification, the HNFB also produces fuzzy rules as a result. In this experiment 45 rules have been generated, of which an example is:
If *Origin* is *low* and *minimum temperature* is *low* then class is *normal*.

5 Conclusions

This paper presented an intelligent system, based on a neural networks committee and on a Hierarchical Neuro-fuzzy BSP system, to identify consumers' frauds in an electrical distribution company of Rio de Janeiro, Brazil.

The system is formed by two modules, one for filtering the database and another for actual classification of a consumer. The use of this methodology greatly improved the current fraud detection rate.

Besides increasing the performance regarding the correct identification of fraudulent consumers, the system has the additional advantage of ranking irregular costumers as a result of the number of votes the neural networks committee provides to identify a customer as irregular.

The proposed methodology is being applied to other areas covered by the company. Additionally, a clustering algorithm will be applied to non-residential consumers in order to group them in classes with similar consumption behaviour.

References

[1] J. E. C. Junior, *Detecção de Fraudes em Unidades Consumidoras de Energia Elétrica Utilizando Rough Sets*, M.SC. Dissertation, UFMS (in Portuguese), 2005.

[2] J. E. Cabral, E. M. Gontijo, J. O. P. Pinto and J.R. Filho, Fraud Detection in Electrical Energy Consumers using Rough Sets, *IEEE International Conference on Systems, Man and Cybernetics*, 4: 3625-3629, 2004.

[3] T. Rauber, I. Drago, F. Varejão and R. Queiroga, Extracão e Seleção de Características na Identificação de Perdas Comerciais na Distribuição de Energia Elétrica, *XXV Congress of the Brazilian Computation Society* (in Portuguese), 2005.

[4] J. Rong, H. Tagaris, A. Lachsz and M. Jeffrey, Wavelet based Feature Extraction and Multiple Classifiers for Electricity Fraud Detection, *Transmission and Distribution Conference and Exhibition*, 3: 2251-2256, 2002.

[5] D. S. G. Pollock, *A Handbook of Time-Series Analysis, Signal Processing and Dynamics*. Academic Press, New York, 1999.

[6] M. Vellasco, Pacheco, K. Figueiredo, F. Souza, Hierarchical Neuro-Fuzzy Systems - Part I, *Encyclopedia of Artificial Intelligence*, Juan R. Rabuñal, Julián Dorado & Alejandro Pazos (Eds.), Information Science Reference, 2008.

[7] R. Santos, M. Vellasco, F. Artola and S. Fontoura. Neural Net Ensembles for Lithology Recognition, *4th International Workshop on Multiple Classifier Systems*, LCNS 2709 – Multiple Classifier Systems, Springer-Verlag, 2003.

[8] F. Provost, T. Fawcett, and R. Kohavi, (1998), The Case Against Accuracy Estimation for Comparing Induction Algorithms, *15th International Conference on Machine Learning*, 445-453, 1998.

[9] M. Dash and H. Liu, Feature Selection for Classification, *Intelligent Data Analysis*, Elsevier, 1(3): 131 - 156, 1997.

[10] S. Haykin (1999), *Neural Networks - a Comprehensive Foundation*, Prentice-Hall, New Jersey, 1999.

[11] C.M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, 1995.

[12] L.B. Gonçalves, M.M.B.R. Vellasco, M.A.C. Pacheco, F. J. de Souza, Inverted Hierarchical Neuro-Fuzzy BSP System: A Novel Neuro-Fuzzy Model for Pattern Classification and Rule Extraction in Databases, *IEEE Transactions on Systems, Man & Cybernetics, Part C: Applications and Review*, 36(2): 236-248, March 2006.

[13] F.-L. Chung and J.-C. Duan, On Multistage Fuzzy Neural Network Modeling, *IEEE Transactions on Fuzzy Systems*, 8(2): 125-142, 2000.

ISBN: 978-989-95079-6-8

A Fuzzy System for the Assessment of Human Reliability

D.M. Szwarcman, J. Domech, R. Tanscheit, M. Vellasco, M. Pacheco, N. P. Ribeiro

Computational Intelligence Lab., Department of Electrical Engineering,
Pontifical Catholic University of Rio de Janeiro
Rio de Janeiro, Brazil
Email: {dilza, , ricardo, marley, marco, nick}@ele.puc-rio.br; jesus.more@estacio.br

Abstract *— This work presents a methodology for the characterization of human reliability based on fuzzy sets concepts, which has been implemented in an innovative decision support system, providing managers with an intelligent computational tool for reducing the possibility of human errors in industrial activities. Considering that such activities can be described as operational, maintenance or inspection processes, which are composed of a set of procedures, the methodology is carried out in two levels: the process level and the procedure level. The proposed system provides a human reliability index, which allows the identification of problems that may constitute causes of human errors, as well as the indication of possible strategies for the control of potentially adverse impacts of interactions that add uncertainty and complexity to processes.*

Keywords— fuzzy sets, human reliability, industrial process safety.

1 Introduction

Human reliability has received much attention in areas such as nuclear, aviation and petrochemical industries. Although resources have been historically applied mostly in equipment reliability and process optimization, it has been realized that priority should be given to the study of human reliability, focusing on the adaptation of the equipment and the working environment to the capabilities, limitations and needs of the human being. Human error usually arises from inadequacies of the system design, such as task complexity and error-likely situations. Errors are likely to occur when the task requirements exceed human limitations regarding perception, attention, remembering, etc. Some situational characteristics may predispose operators to errors, such as inadequate workspace and training procedures, as well as poor supervision. Errors may also reflect individual differences, related to human attributes such as abilities and attitudes. Important individual factors are susceptibility to stress and inexperience, for example, which may increase tenfold the possibility of occurrence of a human error.

Studies in human reliability are generally divided into two generations [1]. First-generation methods are characterized by comparing human performance to that of a machine, associating probabilities of success or of a fault to the operators' actions. Second-generation methods extend the analysis of human reliability to cognitive systems, by considering decision levels, diagnosis processes, dexterity, knowledge and organizational factors. These are probability-based methods, which makes it difficult to establish a precise model for human fault prediction, since a large quantity of data is needed for mapping all the uncertainties inherent to human behavior.

Probabilistic analysis is used for analyzing system reliability objectively and assumes that an equipment or human failure occurs at random. A failure of a single component may occur at random; a human error, however, does not necessarily occur in that way, since a human factor is composed of a large number of attributes (or performance shaping factors) and its functional structure is very complex. By using the probabilistic approach, where the equipment and procedure are qualified, it is assumed that the operator correctly implements all the procedure's provisions and thus isolates the human factor elements.

Other works [2] have contributed to human reliability research employing Fuzzy Set Theory and the concept of possibility of failure instead of probability of failure. Lian and Wang [3] used fuzzy relations to estimate the *fuzzy probability*. Nowakowski [4][5] observed that, when applying the fuzzy approach, probabilistic interpretations of human reliability are abandoned; instead, human reliability is defined in terms of possibility measures. Wang [6] used the concept of fuzziness to evaluate human performance in an inspection task. Onisawa [7] analyzed human reliability in the events that preceded the nuclear accident in Chernobyl and showed the importance of considering the possibility approach in human reliability analysis. More recently, Onisawa [8] created a model that integrates the subjectivity in specialists' opinions to reliability analysis.

This work presents a fuzzy-based decision support system for the analysis of human reliability in operation, maintenance and inspection activities in industrial and production processes, where the human error may have a great impact on safety and on the environment. As operational, maintenance and inspection processes consist of a set of procedures, the human reliability characterization method is carried out at both process and procedure levels.
This paper comprises four additional sections. Section 2 presents the methodology, describing in detail the human reliability characterization at both process and procedure levels. Section 3 presents a decision support system developed to provide managers with computational tools for the application of the human reliability characterization method, helping them to choose the most adequate strategic actions to be taken. Section 4 discusses a case study in the oil industry and Section 5 concludes the work.

2 Proposed Methodology

This work deals with a two-level analysis of human reliability in industrial activities. On a first level, each of the operation, maintenance and inspection processes are

ISBN: 978-989-95079-6-8

analyzed as a whole; then, on a second level, human reliability is considered for every procedure that composes each of the processes.

The following sub-sections describe in details the two-level analysis contemplated by the proposed methodology.

2.1 Process Level Characterization

This module is an improved version of the process level characterization first introduced in [9].

The characterization of human reliability in a process has the objective of establishing the degree of attendance of Performance Shaping Factors (PSFs), which can be of human, technical or environmental types. Such characterization may be carried out in any category: operation, maintenance or inspection.

The method for determining the degree of attendance of a set of attributes (see Fig. 1) begins with the selection of PSFs that affect the human being. After that, experts specify, through a questionnaire, the influence of each of these factors on human reliability. By aggregating the experts' opinions about the PSFs' influences, weighed by the degree of importance of each expert, a standard degree of influence is established for each PSF. By considering the degrees for all PSFs, a Quality Standard is obtained. In a second stage, workers' opinions on the attendance of each of the PSFs are collected, also through a questionnaire, producing the degree of attendance of each PSF [9].

For human PSFs that are related to cognitive characteristics, namely intelligence, communication ability, sociability, attention, stress and anticipation, the attendance degree is determined by psychometric tests applied to each operator. It is considered that, for such type of factors, the operators opinion would correspond to a self judgment, which could result in biased information about the attendance degree. Psychometric tests are likely to provide more reliable results, since they are constructed to determine individual psychosomatic characteristics in an unbiased manner. The enhanced process level characterization method is depicted in the diagram of Fig. 1, which highlights the Psychometrics tests module, one of the significant contributions of this work.

2.2 Procedure Level Characterization

The human reliability characterization at procedure level focuses on two main objectives: establish a failure possibility index for each procedure in an operation, maintenance or inspection process, and for each of the human actions that are performed to accomplish it; and secondly, establish the degree of attendance for each worker with respect to the cognitive demands of each action within the procedure.

The characterization method requires that experts specify the failure expectation for each action, the degree of dependency of each action on the preceding one, the influence that a failure in one action may have on subsequent ones and the cognitive demands of each action. All this information is acquired through the application of questionnaires, where experts give linguistic answers such as *low*, *medium*, or *high*. The experts' opinions are aggregated and weighed by the degree of importance of each expert so as to produce, for each action, a failure possibility index and standard demand degrees with respect to the cognitive factors considered. The overall failure index for the procedure can be obtained by combining the actions indexes. By submitting the workers to psychometric tests, the degree of attendance of each cognitive factor for all operators can be determined. The method is depicted in the diagram of Fig. 2.

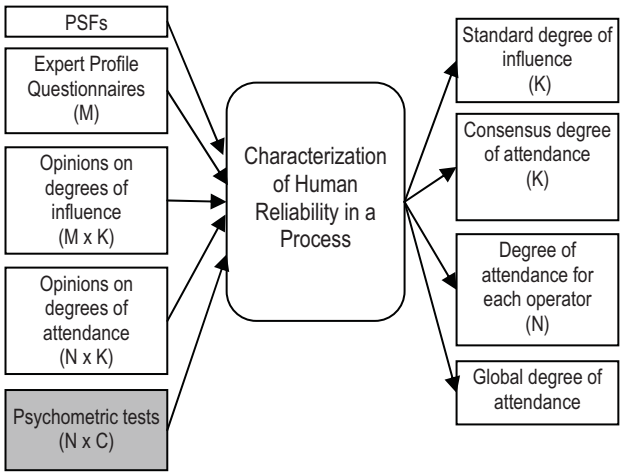


Figure 1: Characterization of Human Reliability in a Process, with K PSFs, C cognitive factors, M experts and N operators.

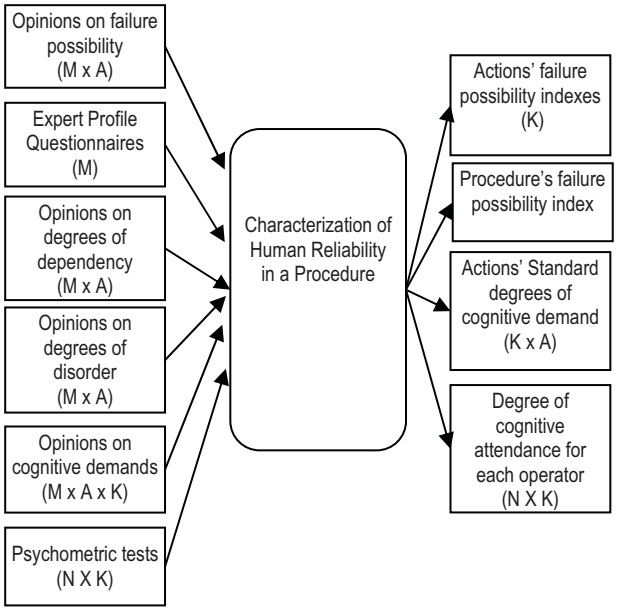


Figure 2: Characterization of Human Reliability in a Procedure, with A actions, K cognitive factors, M experts and N operators.

The procedure level characterization methodology consists of the following stages.

Establishment of a committee of decision-makers

This is one of the most important steps in the methodology, since the quality of information will depend on the experts' proficiency.

Establishment of the relative importance of each expert

This step is accomplished through an Expert Profile Identification Questionnaire (EPIQ), which consists of a set

ISBN: 978-989-95079-6-8

of questions for the evaluation of each expert's relative importance and thus assigning him a weight [10]. This weight will establish the influence of the expert's opinions on the final standard degrees and failure possibility indexes.

Choice of linguistic values for the evaluation of actions

This stage consists of choosing linguistic terms, or values, for the evaluation (by experts) of the expectation of failure of an action, its degree of dependency on the preceding one, its influence on subsequent ones and its cognitive demands. For instance, the terms used for the evaluation of the expectation of failure usually are: *very high, high, medium, low,* and *very low***.** All terms are associated to triangular fuzzy sets, defined by three parameters.

Acquirement of experts' opinions

This step consists of obtaining from the selected experts, through questionnaires, their opinions on the expectation of failure of each action, its degree of dependency on the preceding action, its influence on succeeding ones and its cognitive demands.

Application of psychometric tests to operators

Psychometric tests are applied to operators in order to obtain the degree of attendance of cognitive demands for each worker.

Fuzzy treatment of the data provided by experts and of psychometric tests results

Here, the individual prognoses from each expert for failure expectation, dependency, influence on future actions and cognitive demands are aggregated, generating a consensus for each evaluated attribute. The Hsu and Chen's model [11] is used to pool the expert's opinions: the similarity aggregation method (SAM) is used for combining the opinions of each expert. The opinion of an expert i is expressed by a fuzzy set denoted by $\tilde{A}_i$. The agreement degree (or similarity measure) $S(\tilde{A}_i, \tilde{A}_j)$ between two experts i and j can be determined by the proportion of the consistent area to the total area:

$$S(\tilde{A}i,\ \tilde{A}j) = \frac{\int_x \min(\mu_{\tilde{A}_i}(x), \mu_{\tilde{A}_j}(x))dx}{\int_x \max(\mu_{\tilde{A}_i}(x), \mu_{\tilde{A}_j}(x))dx} \quad (1)$$

Once all agreement degrees between experts are measured, an agreement matrix (AM) can be built, giving an insight into the agreement between the experts.

$$AM = \begin{bmatrix} 1 & S_{12} & S_{13} & ..... & S_{1n} \\ S_{21} & 1 & S_{23} & ..... & S_{2n} \\ . & . & . & . & . \\ . & . & . & . & . \\ S_{n1} & S_{n2} & S_{n3} & ..... & 1 \end{bmatrix} \quad (2)$$

The average agreement degree AAD_i of expert E_i (i = 1 to n) is given by:

$$AAD_i = \frac{1}{n-1}\sum_{\substack{j=1 \\ j\neq i}}^{n} S_{ij} \quad (3)$$

The relative agreement degree RAD_i of expert E_i (i = 1 to n) is given by:

$$RAD_i = \frac{AAD_i}{\sum_{i=1}^{n} AAD_i} \quad (4)$$

Finally, by weighing the relative agreement degree of each expert k by the degree of importance Gi_k (obtained through a questionnaire), the consensus coefficient for that expert can be calculated by equation (5).

$$CC_k = \frac{RAD_k * Gi_k}{\sum_{i=1}^{n}(RAD_i * Gi_i)} \quad (5)$$

The aggregated opinions are given by:

$$\tilde{N} = \sum_{i=1}^{n} CC_i \times \tilde{A}_i \quad (6)$$

In order to obtain a single value for each of the evaluated attributes – representing the failure expectation index, the dependency and adverse influence degrees, and the cognitive demands for each action, the *max* defuzzification method is applied [12].

For the case of cognitive demand, after normalization, and considering all cognitive demands for all actions, a Cognitive Demand Standard is built.

The overall failure possibility index for an action i (P_i) is obtained from its failure expectation index (F_i), its dependency degree on the previous action ($D_{i,(i-1)}$) and the adverse influence degree caused by failure of preceding actions ($T_{j,i}$; where $j< i$):

$$P_i = (1-(1-F_i)*(1-D_{i-(i-1)})*\prod_{j<i}(1-T_{j,i})) \quad (7)$$

The overall failure possibility for a procedure is obtained from the actions indexes:

$$P_{procedure} = (1-\prod_{i}(1-P_i)) \quad (8)$$

Finally, a degree of attendance of the operators to the Cognitive Demand Standard of each action is obtained. For that matter, the degree of attendance (OP_k) to each cognitive demand is multiplied by its demand standard (DS_k) and a weighted average R is then calculated:

$$R = \frac{\sum_{k=1}^{n} DS_k * OP_k}{\sum_{k=1}^{n} DS_k} \quad (9)$$

3 Decision Support System

The decision support system for human reliability analysis was developed as an Intranet application, maintaining all information in a centralized database and being accessible to all professionals involved in a characterization project: managers, experts, operators and psychologists.

The system is totally configurable to any company or business area and can de used for characterization and analysis at process and procedure levels, allowing for:

- human reliability characterization trials or *projects*, as they are called in the system, for different units of the company and different periods of time;
- the complete configuration of the PSFs to be considered, as well as the corresponding questionnaires to be applied to experts and workers, so

ISBN: 978-989-95079-6-8

that questions can be set up in accordance with the personnel's profile and company's characteristics;
- the complete configuration of the sequence of actions that make up the procedures to be analyzed;
- the on-line answering of EPIQ, PSF and procedure evaluation questionnaires by experts;
- the on-line answering of PSF questionnaires by operators;
- the on-line input of the psychometric test results for all workers by an authorized psychologist;
- the association of reference files to each project, such as manuals or norms that must be followed on the given human reliability characterization trial.

The system provides two sets of reports that give managers effective support on strategic decisions regarding the improvement of human reliability, and, consequently, the minimization of process failure.

The first set of reports, called *Process Reports*, comprises results from the process level characterization, including:

- Degree of importance of *PSFs*:
  The opinions of all experts with respect to the degree of importance of all PSFs are presented in a comparative chart.
- Average degree of attendance of PSFs:
  The average degrees of attendance of all PSFs, as seen by each operator, are presented in a comparative chart. A degree close to 1 indicates that, in the environment being evaluated, the operator considers that, in average, the quality standard is being met. On the other hand, a degree closer to 0 indicates that the operator evaluates that, in average, the quality standard is not being attended to. In this case, the possibility of failure for this worker is high.
- Distance to quality standard for all PSFs, as seen by a given operator:
  The quality standard for PSFs, established by the experts, represents the PSF *demand*, while the attendance degrees specified by a worker represent the PSF *offer*. This report shows graphically and in tabular form the distances between offer and demand. The greater the distance between the attendance degree and the quality standard, the more efforts and investments should the company dedicate to the given PSF to increase its attendance level. Factors with distances larger than 0.5 should merit special attention (Fig. 3).
- Inclusion coefficients for all PSFs, as seen by a given operator:
  This report shows the inclusion degree of the fuzzy set *quality standard* in the fuzzy set *attendance*, given by each operator. An inclusion coefficient close to 1 indicates that, in the environment being evaluated, the particular operator assimilation is greater than the quality standard.

The second set of reports, called *Procedure Reports*, present results from the procedure level characterization, including:

- Failure expectation for actions:
  The experts' aggregated expectations with regard to the failure of each action are presented. Failure expectations are presented together with the corresponding aggregated certainty degrees, which indicate the overall degree of sureness of the experts with respect to their opinions.

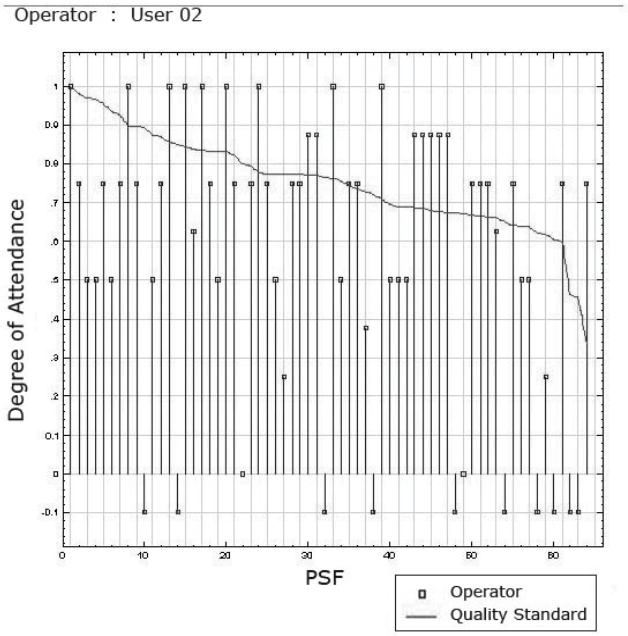


Figure 3: Distance to quality standard for all PSFs, as seen by a given operator.

- Dependency degree of actions:
  The experts' aggregated opinions on the dependency of each action on the previous one are presented. A degree close to 1 indicates that a failure on the previous action will most probably cause a failure on the current action. Actions with degrees higher than 0.6 deserve special attention. The dependency degrees are also presented together with the related aggregated certainty degrees.
- Influence of actions:
  In a matrix form, this report presents the experts' aggregated opinions about the possibility of a failure in each action having an influence on succeeding actions (Fig. 4). Degrees higher than 0.6 should be analyzed.

Action	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	**1.00**	**0.71**	**0.75**	**0.85**	**0.76**	**0.76**	**0.72**	**0.76**	**0.76**	**0.76**	**0.76**	**0.76**	**0.83**	**0.65**	**0.83**	**0.86**	**0.83**	**0.75**
2		**1.00**	0.50	0.54	0.50	0.54	**0.69**	**0.76**	**0.92**	**0.76**	**0.83**	**0.76**	**0.72**	**0.65**	**0.81**	**0.75**	**0.81**	**0.75**
3			**1.00**	0.25	0.27	0.32	0.32	**0.83**	**0.66**	**0.72**	**0.65**	**0.72**	0.41	0.18	0.18	0.19	0.17	0.19
4				**1.00**	0.59	0.06	0.17	0.33	0.33	0.17	0.25	0.33	**0.69**	0.41	0.59	0.41	**0.67**	0.48
5					**1.00**	**0.81**	0.54	**0.76**	**0.75**	0.25	0.16	**0.86**	0.51	0.25	0.33	0.10	0.27	0.10
6						**1.00**	0.42	**0.76**	0.52	0.34	0.16	**0.76**	0.42	0.17	0.27	0.10	0.27	0.17
7							**1.00**	**0.86**	**0.83**	**0.83**	**0.86**	**0.76**	**0.72**	**0.67**	**0.76**	0.58	**0.66**	0.47
8								**1.00**	0.19	0.19	0.19	0.41	0.17	0.25	0.06	0.18	0.13	0.06
9									**1.00**	**0.83**	0.32	0.19	0.27	0.25	0.27	0.18	0.18	0.19
10										**1.00**	0.09	0.09	0.33	0.17	0.50	0.17	0.18	0.18
11											**1.00**	0.41	0.06	0.27	0.25	0.33	0.10	0.02
12												**1.00**	**0.81**	0.27	0.17	0.27	0.17	0.18
13													**1.00**	0.59	0.32	0.32	0.27	0.42
14														**1.00**	0.32	0.32	0.27	0.36
15															**1.00**	0.59	0.42	0.42
16																**1.00**	0.37	0.41
17																	**1.00**	0.52
18																		**1.00**

Figure 4: Influence of actions on subsequent ones

- Cognitive demands of actions:
  This report presents the degree with which a cognitive aspect is demanded for a worker to correctly execute an action (in the consensual opinion of all experts). The

ISBN: 978-989-95079-6-8

degree of attendance of each cognitive factor for each operator is also shown. It indicates on which cognitive aspects the worker should receive training. Cognitive demands higher than 0.6 deserve special attention.

- Procedure overall failure possibility:
  The overall failure possibility indexes for the procedure and its actions are shown.

4 Case Study

An oil refinery was chosen as the unit for the case study, namely REDUC (Refinery of Duque de Caxias). This is located at Campos Elísios district, Rio de Janeiro, Brazil, and produces petrol, lubricants and other oil derivatives. The object of this case study is Procedure 1 (part of the Refinery Operation Process [9]), composed of the actions listed in Fig. 5.

The influence on each action by each cognitive factor was evaluated by 17 experts, who answered the questionnaire. The cognitive factors studied were: Anticipation, Attention, Calculation, Emotional State, Group Identification, Intelligence, Memory, Perception, Personality and Stress.

Taking as a basis the API 770 norm, two PSFs questionnaires, one for operators and another for experts, were configured for the considered unit. Both questionnaires correlate one question to each PSF. By using this specific questionnaire, an expert can evaluate which are the most critical PSFs. The system generates the PSFs quality standard by putting together the answers of all experts and weighing them based on EPIQ. By answering his questionnaire, an operator evaluates each PSF attendance. To evaluate the attendance of PSFs associated to cognitive factors, a questionnaire is used by a specialized psychologist, who analyses, based on standard psychology tests, how each interviewed operator is responding to each cognitive factor.

In total, 51 workers of the Operation Process answered the PSFs questionnaire, giving opinions about their working conditions. The 17 available experts answered the questionnaire in order to determine which PSFs had higher need to be attended. They also answered the EPIQ and the questionnaires about actions of Procedure 1, that is, expectation of failure, degree of dependency, influence on subsequent actions and cognitive demands. After all operators and experts of REDUC answered their respective questionnaires, the final reports could be consulted in the system by *project* administrators and managers.

The reports obtained in this case study are described below.

Report of Failure Expectation of Actions

From the report presented in Fig. 6, it is possible to conclude that the actions which had the higher failure expectation were: Return of Line to The Operation, Release of Line to The Maintenance, Return of Tank to The Operation, Aliments Change, and Area Survey.

Action Dependency Ratio Report

This report, depicted in Fig. 6, demonstrates that the actions which had higher dependency ratio on the preceding action were: Return of Pot Flare (18), Return of Line to The Operation (16), and Product Shipments to CIAS, BR/GEI, PETROFLEX (8).

Action Influence Report (see Fig. 4)

Based on this report, it is possible to conclude that actions Demands (1) and Aliments Change (7) have more influence on subsequent actions than any other. Additionally, Product Shipments to CIAS, BR/GEI, PETROFLEX (8) and Loading of Trucks (12) are most affected than other actions.

	Action	Failure Expectation	Certainty Degree
1	Demands	0.0000	1.0000
2	Area Survey	0.3042	1.0000
3	Measurements and Sampling	0.2370	0.7865
4	Drainages	0.1043	0.7291
5	Pumps Starts	0.0269	0.9490
6	Pumps Stops	0.0269	0.8901
7	Aliments Change	0.4313	0.8901
8	PETROFLEX	0.2500	1.0000
9	Exchange Loads of Units	0.2298	1.0000
10	Exchange Receipt of Units	0.2298	0.9314
11	Gravitations	0.0269	0.9314
12	Loading of Trucks	0.1367	0.8901
13	Release of Equipments to Maintenance	0.2500	0.0784
14	Return of Tank to The Operation	0.4458	0.6542
15	Release of Line to The Maintenance	0.4844	0.6542
16	Return of Line to The Operation	0.4844	0.6542
17	Release of Pot Flare	0.2772	0.8901
18	Return of Pot Flare	0.2772	0.8901

Figure 5: Failure Expectation of Actions for Procedure 1.

Action	Dependency Ratio	Certainty Degree
1	--	1.0000
2	0.5334	1.0000
3	0.3855	0.7865
4	0.4215	0.7291
5	0.3336	0.9490
6	0.3067	0.8901
7	0.2840	0.8901
8	**0.6192**	1.0000
9	0.0303	1.0000
10	0.5510	0.9314
11	0.4388	0.9314
12	0.1043	0.8901
13	0.0378	0.0784
14	0.5055	0.6542
15	0.3623	0.6542
16	**0.7674**	0.6542
17	0.5334	0.8901
18	**0.9334**	0.8901

Figure 6: Dependency ratio between each action and the preceding one

Cognitive Demand Attendance Degree of Actions

ISBN: 978-989-95079-6-8

This report, shown in Fig. 7, is used for comparison between the cognitive demand of actions of Procedure 1 and the attendance of each worker. Therefore, a manager can infer whether an operator is capable of executing the actions he/she is assigned to. For instance, it can be seen in Fig. 7 that the given operator does not attend the Anticipation requirement of actions 9, 13 and 16.

Global Index of Failure Possibility of Actions

This report, shown in Fig. 8, indicates the possibility of failure of each action, considering every influence it experiences (such as influences and dependencies from other actions, failure expectation). The system allows the manager to choose the cut level of the influence to be considered. A higher cut level means that the preceding actions which have low influence will not be considered. As a result, in this case study, a 0.6 cut level returns the mid and final actions as the ones with the highest failure possibilities. In opposition, a 0.9 cut level returns a more disperse failure profile, since it only considers the preceding actions that have large influence on the action itself.

	Anticipation	Attention	Calculation	Emotive State	Group Id	Intelligence	Memory	Perception	Personality	Stress
1	0.600	**0.615**	0.262	**0.645**	0.523	0.466	**0.800**	0.458	**0.724**	0.472
2	**0.700**	**0.779**	0.047	**0.700**	0.600	**0.615**	**0.700**	**0.818**	**0.713**	0.577
3	0.600	**0.949**	**0.770**	**0.665**	0.600	**0.690**	**0.700**	**0.738**	**0.700**	**0.624**
4	0.516	**0.762**	0.431	0.521	0.581	0.566	0.100	**0.636**	0.455	**0.705**
5	**0.614**	**0.716**	0.554	**0.633**	0.414	0.600	0.509	**0.905**	0.570	0.494
6	**0.614**	**0.734**	0.197	0.509	0.413	0.600	0.414	**0.843**	**0.638**	0.410
7	**0.781**	**0.955**	**0.700**	**0.667**	**0.762**	**0.638**	**0.732**	**0.734**	**0.694**	**0.662**
8	**0.800**	**0.941**	**0.769**	**0.667**	**0.669**	**0.609**	**0.613**	**0.764**	**0.686**	**0.641**
9	**0.973**	**1.000**	**0.700**	**0.654**	**0.800**	**0.668**	**0.620**	**0.782**	**0.700**	0.600
10	**0.800**	0.000	0.539	**0.700**	0.000	**0.800**	**0.800**	**0.700**	**0.700**	0.600
11	**0.700**	**0.658**	**0.700**	**0.700**	0.000	0.570	0.558	**0.666**	**0.700**	0.600
12	**0.843**	**0.805**	**0.632**	**0.700**	**0.773**	0.600	0.600	**0.669**	**0.700**	**0.628**
13	**0.881**	**0.863**	0.600	**0.700**	**0.604**	**0.721**	**0.854**	**0.762**	**0.691**	0.403
14	**0.822**	**0.884**	0.390	0.533	**0.613**	**0.679**	**0.939**	**0.790**	**0.689**	0.500
15	**0.858**	**0.812**	**0.700**	**0.632**	**0.690**	**0.700**	**0.868**	**0.800**	**0.666**	0.500
16	**0.973**	**0.900**	**0.700**	**0.632**	**0.690**	**0.679**	**0.895**	**0.810**	**0.689**	0.500
17	**0.869**	**0.939**	**0.647**	0.579	**0.791**	**0.692**	**0.868**	**0.889**	**0.666**	0.513
18	**0.869**	**0.939**	**0.647**	**0.718**	**0.791**	**0.628**	**0.869**	**0.887**	**0.666**	**0.628**
Attendance										
User 35	**0.875**	**0.875**	--	--	**1.000**	0.500	--	--	--	0.125

Figure 7: Cognitive Demand Attendance

5 Conclusions

An innovative system for the evaluation of human reliability in industry has been presented. The approach taken makes use of fuzzy sets, so that experts and operators' opinions can be translated into mathematical terms. As a consequence, a Quality Standard (QS) can be established and the attendance to performance shaping factors can be evaluated.

This system can be used at both process and procedure (consisting of series of actions) levels and produces several reports that help managers to make decisions aimed at reducing the possibility of human errors. A case study considered a unit of a Petrobras (Brazilian Oil Company) refinery, where the system has been actually applied to.

Acknowledgment

The authors would like to thank Petrobras S.A. for the financial support to develop the decision support system.

Action	Global Index of Failure	
	(Cut Level 0,6)	(Cut Level 0,9)
1	0.0000	0.0000
2	0.5334	0.5334
3	0.4632	0.4632
4	0.4495	0.4495
5	0.3416	0.3416
6	0.5036	0.3131
7	0.6444	0.3807
8	0.9759	0.6554
9	0.9083	0.5796
10	0.9841	0.6108
11	0.9049	0.4480
12	0.9309	0.1592
13	0.9586	0.0761
14	0.8948	0.5223
15	0.8946	0.5236
16	0.9210	0.8264
17	0.9217	0.6403
18	0.9703	0.9452

Figure 8: Global Index of Failure Possibility of Actions.

References

[1] Y. Fujita, Human Reliability Analysis: a Human Point of View. *Reliability Engineering and System Safety* 38, 1992.

[2] Y. Nishiwaki, Human Factors and Fuzzy Set Theory for Safety Analysis. In: M.C. Cullingford, S.M. Shah, J.H. Gittus (Eds), *Implications of Probabilistic Risk Assesment*, Elsevier, 253-274, 1987.

[3] G.S. Liang, W.J. Wang, Evaluating Human Reliability using Fuzzy Relation, *Microelectronics and reliability*, 33(1): pp 63-80., 1993.

[4] M. Nowakowski, Possibility Distributions in the Linguistic Theory of Actions, *Int. Journal of Man-Machine Studies*, 12: 229-239, 1980.

[5] M. Nowakowski, The Human Operator: Reliability and Language of Action Analysis. In: W. Karwowski, A. Mital (Eds), *Applications of Fuzzy Set Theory in Human Factors*, Elsevier, 165-177, 1986.

[6] M. J. Wang, J. Sharit, and C.G. Drury, An Application of Fuzzy Set Theory for Evaluation of Human Performance on an Inspection Task, In: W. Karwowski, and A. Mital (Eds.) *Application of Fuzzy Set Theory in Human Factors*, Elsevier, 257-268, 1986.

[7] T. Onisawa, Y. Nishiwaki, Fuzzy Human Reliability Analysis on the Chernobyl Accident, *Fuzzy Sets and Systems*, 28: 115 -127, 1988.

[8] T. Onisawa, Subjective Analysis of System Reliability and its Analyzer, *Fuzzy Sets and Systems*, 83: 249-269, 1996.

[9] J. D. More, R. Tanscheit, M.M. Vellasco, M.A. Pacheco, D.M. Szwarcman, A Fuzzy Approach to the Study of Human Reliability in the Petroleum Industry. In: O. Castillo; P. Melin; O.M. Ross; R.S. Cruz; W. Pedrycz; J. Kacprzyk. (Org.). *Advances in Soft Computing - Theoretical Advances and Applications of Fuzzy Logic and Soft Computing*, Springer, 42: 415-424, 2007.

[10] J. D. More, *A Fuzzy Approach to the Evaluation of Human Reliability in Ultrasonic Nondestructive Examinations*, Ph.D. Thesis COPPE/UFRJ, Brazil, 2004.

[11] H. M. Hsu, C. T. Chen, Aggregation of Fuzzy Opinions under Group Decision Making, *Fuzzy Sets and Systems* 79: 279-285, 1996.

[12] T. J. Ross, *Fuzzy Logic with Engineering Applications*, 2nd Ed., John Wiley & Sons, 2004.

ISBN: 978-989-95079-6-8

Multivariate Segmentation of Time Series with Differential Evolution

Daniel Graves[1] Witold Pedrycz[2]

[1,2]Department of Electrical & Computer Engineering, University of Alberta, Edmoton, Canada
Email: dgraves@ualberta.ca[1], pedrycz@ee.ualberta.ca[2]

Abstract—*A new method of time series segmentation is developed using differential evolution. Traditional methods of time series segmentation focus on single variable segmentation and as such often determine sections of the time series with constant slope (i.e. linear). The problem of segmenting multivariate time series is significantly more involved since several time series have to be jointly segmented. Thus the concept of boundary becomes ill-defined since each time series may not be exactly synchronized and change identically in time. The problem is rectified by minimizing the mean of the variance of the slopes determined in each segment. Performance of the method is measured in terms of the classification rate and the accuracy of determination of boundaries. Experimental evidence shows the effectiveness of the method when applied to synthetic and real-world data compared with multivariate time series clustering approaches.*

Keywords—Multivariate segmentation, differential evolution, time series, fuzzy clustering.

1 Introduction

Many data come in the form of time series that need to be partitioned into contiguous blocks in time. Tools of Computational Intelligence have been applied to solve the time series partitioning problem are commonly referred to as time series segmentation including fuzzy clustering, hierarchical clustering, and genetic algorithms. Time series segmentation of a single variable has been studied extensively in the literature, c.f. [1,2,3,4,5]. However, there is a lack of research in the area of multivariate time series segmentation or the joint segmentation of several time series.

Only a few multivariate segmentation algorithms found in the literature cluster the time series by taking into account temporal information, e.g., [5,6,7,8]. Unfortunately clustering a time series to perform segmentation poses a few problems. An important issue with segmentation of a time series which has to be addressed is that segments must be contiguous in time. Several approaches have been considered to solve this problem. One method uses constrained clustering to find contiguous segments in time, i.e. constrained hierarchical clustering. Another approach uses time information as an auxiliary component exploited directly in the clustering process. A significant problem is to effectively use this auxiliary information in the clustering procedure leading to boundary identification. However, augmenting clustering with time information is not an appropriate vehicle for segmenting a time series since identification of boundaries requires use of time information as auxiliary information in clustering. A common problem produced by augmented clustering algorithms that account for time information is that the resulting partition does not necessarily consist of contiguous segments. Currently, the literature lacks an effective method of using the auxiliary information in clustering.

The focus of this research is in multivariate time series segmentation by a more direct segmentation method that avoids the problems produced by using clustering as a vehicle for segmentation by employing differential evolutionary optimization of an objective function. As well, evolutionary methods such as differential evolution are well suited for structural optimization. The reason for using differential evolution (DE) to segmentation a time series is that the segmented boundaries to be optimized cannot be *explicitly* expressed in the underlying objective function. This makes gradient-based optimization methods not feasible in this problem. The optimization calls for methods of structural optimization and various techniques of Evolutionary Computing are of interest. In this category of methods, DE has shown to be effective as far as quick convergence is concerned.

We focus on detecting slope changes in the time series in order to determine segments. The motivation behind approaching the problem of segmentation by detecting monotonicity is that often a time series can be represented quite effectively using only a handful of monotonic models for each segment. An example is with a set of ECG time series which are often characterized by sharp spikes in the middle. A potential application of segmentation that detects changes in slope is in signal compression, c.f. [2].

There are a number of other applications of multivariate time series segmentation. An example is presented in [9] for process monitoring, diagnosis and control of a medium and high-density polyethylene plant. Another application of multivariate time series segmentation is with music structural segmentation where the objective is to automatically determine a partition of structural elements in music, i.e. verse, chorus, bridge, by detecting significant transition points in the music.

The material is organized in five sections. A literature review is covered in Section 2. Section 3 includes a description of the segmentation algorithm and presents the essentials of differential evolution. Section 4 concentrates on the experimental evaluation of the multivariate segmentation algorithm. The proposed method is evaluated on synthetic data sets and on several real-world time series data sets coming from Time Series Data Library [10]. Section 5 presents the conclusions.

ISBN: 978-989-95079-6-8

2 Literature Review

Time series segmentation is a very important topic with temporal information processing as long time series need to be broken down into relevant segments. The problem of segmentation has not been solved with high boundary detection accuracy in the literature hence most of research conducted in the literature focuses on the segmentation aspect and leaves the labelling of segments to be done later.

Most time series segmentation approaches look at the problem of segmenting a time series of a single variable. There are three main categories of classical segmentation algorithms available in the literature:

- Bottom-up
- Top-down
- Sliding-window

Unfortunately high segmentation accuracy is not always achieved and there is no one method preferred than the others. The three algorithms are described in [1]. All three have an error threshold where they stop segmenting once the error is exceeded. The first bottom-up starts with each sample in its own segment much like agglomerative hierarchical clustering. Likewise top-down is similar to divisive hierarchical clustering in that the entire time series is broken down into smaller segments until the error for each segment is less than the error threshold. Sliding-window is a real-time segmentation algorithm that starts with the first sample and extends the segment with each new sample until an error threshold is exceeded.

Determining segments via linear approximation and evolutionary methods is common in the literature such as methods using genetic algorithms to approximate an ECG time series trough a set of its linear segments, cf. [2]. In [2], the fitness function minimizes the variability between the maximum and minimum slopes for each consecutive sample within the proposed segment. The number of segments is determined beforehand. The experiments are conducted on ECG signals which show fairly good match on the data, i.e. little information appears to be lost during segmentation (which also leads to dimensionality reduction).

Multivariate time series are common in practical problems. Many of the approaches discussed in the literature are not directly applicable to them. An inherent problem with multivariate time series segmentation is that it difficult to determine boundary points where each feature time series in the multivariate time series unanimously agree. For example, a question arises whether a significant change in one feature's value where all other features do not change significantly constitutes a boundary change point for a time series. Also not all features may change at exactly the same point in time; hence, multivariate segmentation approaches require greater boundary point flexibility to account for the latent changes among the feature values. As a result, some limited work has been done in the literature directly applicable to multivariate time series segmentation, c.f. [5,6,7,9].

The multivariate time series segmentation method developed in [9] approaches the problem as a constrained clustering problem with the constraint that a cluster (i.e. segment) must contain patterns at successive time points. The approach also constructs fuzzy sets for specifying the segments detected with fuzzy boundaries. Local Principal Component Analysis (PCA) models are used in clustering of the time series into segments based on Gath-Geva clustering. They apply the segmentation algorithm to process monitoring, diagnosis and control in particular monitoring a medium and high-density polyethylene plant, i.e. production of a versatile plastic. The data is multivariate and is successfully segmented via the proposed clustering algorithm.

A Gath-Geva-based clustering algorithm similar to the approach in [7,9] is introduced for application to multivariate time series segmentation. The approach described in [9] and [7] utilize a modified Gath-Geva clustering based on the probabilistic PCA (PPCA) [11] for each cluster. The method in [9] and [7] constrain the segment (cluster) fuzzy sets (membership matrix) to Gaussian membership functions. The method in [7] uses modified Gath-Geva clustering (a.k.a. fuzzy maximum likelihood clustering) to cluster a multivariate time series. The time information is included as an attribute in the time series. The segments are projected onto principal components based on a localized PPCA models for the rationale that the correlation between variables often changes with multivariate time series. The distance measure in the clustering is chosen to include two parts: Gaussian fuzzy sets determining distance in the time domain and a second term that measures the distance of the data from the PCA hyperplane model. The number of principal components to keep in the PPCA model and the number of segments (clusters) are two important parameters that need to be selected via some other method. The authors include some case studies in the literature for the application of the approach to multivariate time series segmentation although the experimental evidence is limited.

Speaker segmentation is discussed in [5]. The method exploits the BIC (Bayesian Information Criterion) procedure, known also as the minimum description length (MDL). The objective is to determine segments of different speakers. Some improvements are found in their approach over traditional BIC approaches. Multivariate segmentation in [6] is accomplished via constrained clustering for the purpose to determine the important structural components of a musical recording, i.e. intro, verse, chorus, solo, bridge, outro. They report some improvements in the classification rate and boundary f-value with their method over traditional approaches. A multivariate time series segmentation algorithm based on hierarchical Bayesian clustering was developed in [12] specifically designed for astronomical data that can be modelled as a Poisson counting process. Some experiments on synthetic data and real astronomical data are provided however they are limited. Authors in [8] describe multivariate segmentation approach for motion data using singular value decomposition (SVD) in the distance measure between multivariate motion time series. The experimental results are very limited.

ISBN: 978-989-95079-6-8

3 Multivariate Segmentation

3.1 *Problem Formulation*

The problem of multivariate segmentation involves finding the boundaries in a time series that denote changing behaviour in the time series. For instance a time series might be increase steadily and then at a time t_o the time series changes behaviour and starts decreasing steadily. Extending this to the multivariate case, each variable which is a time series in itself, exhibits changes in behaviour after the boundary at t_o. With single-variable time series segmentation, segments are contiguous homogenous partition series that are specified by a set of boundaries that divides segments from their adjacent ones. Multivariate segmentation is the problem of finding a set of boundaries that jointly divides each time series into a set of contiguous partitions of multivariate time series. With multivariate time series, segmentation is more complicated since each feature can be considered a separate time series and hence what might be considered a boundary in one may not be considered a boundary in the other. i.e., the transitions in each feature signal may start and end at different points in time providing conflicting clues as to where the segment starts and ends.

Given a multivariate time series $x_{t,k}$ for t=1…T with M features k=1…M, the time series can be segmented through slope analysis of each feature. The idea is to find segments with similar slopes, i.e. finding segments that are monotonic with relatively constant slopes. Assuming the number of segments c is known a priori, the boundaries for the segments are the free-parameters given by $[b_1, b_2, \dots, b_{c-1}]$ where $b_0 = 1$ and $b_c = T$. An example of a multivariate segmentation is given in figure 1.

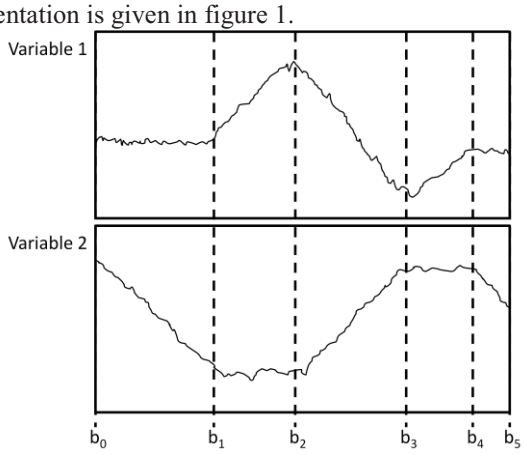


Figure 1: Multivariate segmentation illustration

3.2 *Differential Evolution*

Differential evolution is simple population-based evolutionary optimization technique that minimizes (or maximizes) an objective function much like genetic algorithms, c.f. [13]. Differential evolution uses mutation and crossover operations just as genetic algorithms; however, the mutation operation is based on population vector differences. The population of vectors are given by the set $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_P\}$ where each vector $\mathbf{x}_i$ is real-valued and has dimensionality M. The objective function to be minimized is denoted by f(**x**). The vectors in the population assume values in the unit interval and P is the size of the population. The range of values for the population vectors is on the interval [0,1].

The essence of the differential evolution is expressed via the following pseudo code

1. Initialize Population of size P
2. Loop until stopping criterion met
    1. For each element $\mathbf{x}_i$ in the population
        1. Randomly select three parameter vectors: $\mathbf{x}_{r1}$, $\mathbf{x}_{r2}$, $\mathbf{x}_{r3}$ (not the same)
        2. Generate a mutant vector $\mathbf{v}_i=\mathbf{x}_{r1}+G[\mathbf{x}_{r2}-\mathbf{x}_{r3}]$ *[Mutation]*
        3. Generate trial vector $\mathbf{u}_i=\text{crossover}(\mathbf{v}_i,\mathbf{x}_i)$ by mixing vectors $\mathbf{v}_i$ and $\mathbf{x}_i$ *[Crossover]*
        4. If $f(\mathbf{u}_i) < f(\mathbf{x}_i)$ then replace $\mathbf{x}_i$ with $\mathbf{u}_i$ in population
3. Termination

The difference factor G controls the intensity of mutation (i.e. vector differences). The difference factor G can take on positive real numbers as well as zero, c.f. [4]. There is also a crossover operation. Many different kinds of crossover operations are possible. Here we use the one presented in [4]. This crossover operation combines two vectors $\mathbf{x}_i$ and $\mathbf{y}_i$ according to the pre-specified probability p_{cr}. The resulting crossover vector produced by crossover is give by the expression

$$u_{i,k} = \begin{cases} v_{i,k} & if\ r_k < p_{cr} \\ x_{i,k} & otherwise \end{cases} \quad (1)$$

where r_k is a uniform random sequence of numbers on the interval [0,1] for k=1 … M.

3.3 *Fitness Function*

The boundaries of the segments are determined via differential evolution. Here we elaborate on the pertinent algorithmic details.

A difference (describing slope) sequence is calculated for each variable k=1…M in the multivariate time series

$$D_{tk} = x_{t+1,k} - x_{t,k} \quad (2)$$

for t=1…T-1.

Given a set of boundaries $[b_1, b_2, \dots, b_{c-1}]$ where $b_0 = 1$ and $b_c = T$, the length of each segment is given as

$$L_i = b_i - b_{i-1} \quad (3)$$

The variance of the slopes of each variable are calculated via the expression

$$S_{ik}^2 = \sum_{t=b_{i-1}}^{b_i} \frac{(D_{tk} - \overline{D}_{ik})^2}{L_i - 1} \quad (4)$$

where $\overline{D}_{ik} = 1/L_i \sum_{t=b_{i-1}}^{b_i} D_{tk}$ is the mean slope of each variable k=1…M for each segment i=1...c. The mean of the standard deviations across the variables is calculated for each segment, i.e.

$$S_i = \frac{1}{M} \sum_{k=1}^{M} \sqrt{S_{ik}^2} \quad (5)$$

for i=1…c.

ISBN: 978-989-95079-6-8

Finally multivariate segmentation is accomplished by minimizing the following fitness function

$$F = \sum_{i=1}^{c} S_i L_i \tag{6}$$

where S_i is the standard deviation defined in (4) and L_i is the segment length defined in (2). The resulting segmentation favors segments with slopes that do not vary greatly; hence, the resulting segments are approximately linear.

3.4 *Population Encoding*

Each population vector is encoded in differential evolution as a vector of real-valued numbers denoted by $\mathbf{y} = [y_1, y_2, \ldots y_{c+1}]$. The set of boundaries $\mathbf{b} = [1, b_1, b_2, \ldots, b_{c-1}, T]$ is calculated from the vector $\mathbf{y}$ where the first and last elements ($b_0 = 1$ and $b_c = T$) can be discarded. The boundaries are calculated in the form:

1. $\mathbf{z} = sort - least - to - greatest(\mathbf{y})$
2. $\mathbf{a} = \frac{\mathbf{z} - z_1}{z_{c+1}}$
3. $\mathbf{b} = round(1 + (T-1)\mathbf{a})$

The elements of vector $\mathbf{y}$ is sorted from least to greatest and its sorted version is denoted as a new vector $\mathbf{z}$ so that the boundaries are given in order. The vector $\mathbf{z}$ is normalized to a vector of numbers on the range [0,1] called $\mathbf{a}$ by subtracting the smallest element in $\mathbf{z}$ given by z_1 since the vector is sorted and divided by the largest element z_{c+1}. The vector $\mathbf{a}$ maps directly to boundaries in the time series where 0 is the start of the time series and 1 is the end of the time series. The vector $\mathbf{b}$ is calculated by expanding the range of $\mathbf{a}$ to the set of integer time coordinates $\{1, T\}$. The reason for sorting the vector is that taking vector differences (the mutation operation in differential evolution) can result in negative values. Simply normalizing the vector by its sum as done in [2] would produce erroneous results since vector elements are not necessarily positive. Elements of $\mathbf{y}$ needed to be mapped to the interval [0,1] where the smallest value of $\mathbf{y}$ is mapped to the start of the time series.

4 Experiments

The objective function described in section 3.2 minimizes slope variance for each segment. The optimization procedure is unsupervised, i.e. does not use pre-labelled data during optimization. The following synthetic experiments are evaluated against an *a priori* segmentation provided by a human expert. The classification rate and boundary f-value criteria are used to evaluate the synthetic experiments. The classification rate is the percentage of correctly classified samples according to a pre-segmented time series (human labelled) and is indicative of the precision. The boundary f-value is a measure of the number of correct boundaries in the resulting segmentation. The boundaries discovered by the algorithm are denoted by vector $\mathbf{b}$ and the human labelled boundaries are given by the vector $\mathbf{h}$ where $\mathbf{b}$ and $\mathbf{h}$ have length denoted by $L_{\mathbf{b}}$ and $L_{\mathbf{h}}$ respectively. The boundary f-value is determined according to the following expression

$$f = \frac{2pr}{p + r} \tag{7}$$

where p is the boundary precision and r is the boundary recall calculated as follows

$$n = \sum_{i=1}^{L_{\mathbf{h}}} \begin{cases} 1 & if\ \theta > \left| \mathbf{h} - \min_{\forall j = 1 \ldots L_{\mathbf{b}}} \{|h_i - b_j|\} \right| \\ 0 & otherwise \end{cases} \tag{8}$$

$$p = \frac{n}{L_{\mathbf{b}}} \tag{9}$$

$$r = \frac{n}{L_{\mathbf{h}}} \tag{10}$$

where n is the number of correctly classified boundaries with respect to the provided human labelled boundaries and θ is a threshold value which determines the tolerance or region of acceptable difference between the human labelled boundary and machine labelled boundary. The following figure demonstrates the calculation of the f-value.

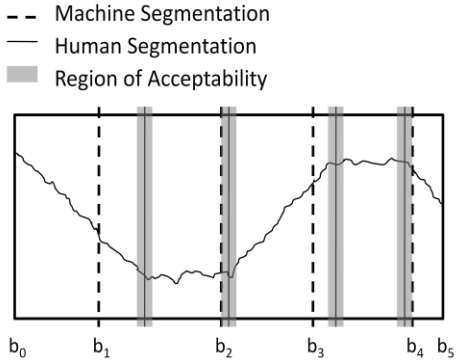


Figure 2: Multivariate segmentation illustration

The reconstruction error is the performance index for the real-world data sets since no human labels are present, i.e.

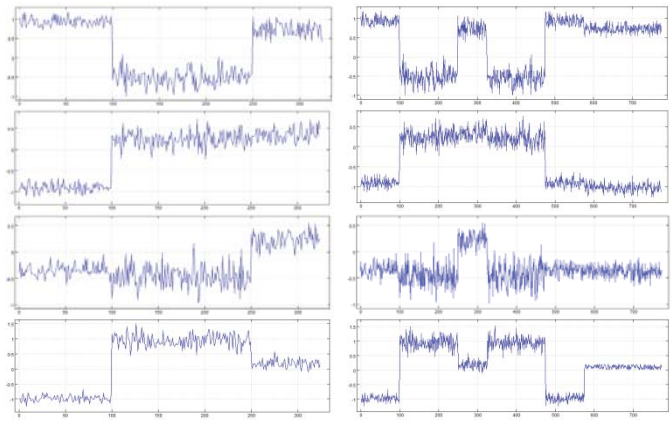
$$E = \frac{1}{MT} \sum_{k=1}^{M} \sum_{t=1}^{T} \left(\hat{x}_{t,k} - x_{t,k} \right)^2 \tag{11}$$

where $\hat{x}_{t,k}$ is the reconstructed time series. The time series $\hat{x}_{t,k}$ is reconstructed by approximating each segment via a linear line of the form $\hat{x}_{t,k} = m_k t + b_k$ from the first point in the segment to the last point in the segment for each variable k=1...M. The values m_k and b_k are the slope and intercept, respectively, of the line for each variable k.

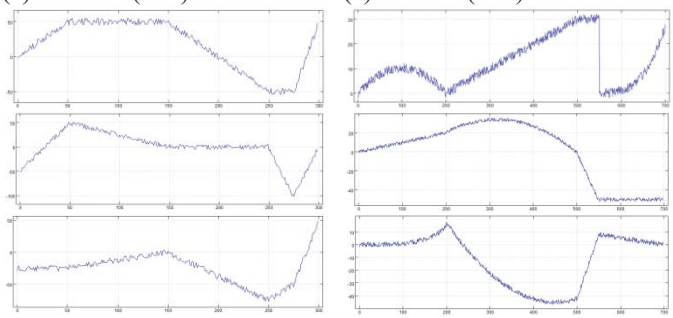
Four synthetic time series were segmented using the described multivariate time series segmentation algorithm. Each variable for time series DATA1, DATA2, DATA3, and DATA4 are given in the figure below where DATA1 and DATA3 have four feature time series and DATA2 and DATA4 have three. The length of each time series is 325, 300, 775, and 700 samples respectively. These time series were generated by creating a collection of linear segments and then adding white noise.

The experimental results on the synthetic data are shown in table 1. They are compared with FCM-DFS (Fuzzy C-Means clustering with Distinct Feature Sets) time series clustering described in [14] for segmentation. The value of theta was chosen to be relatively small, i.e. 3, since the synthetic data sets are relatively simple often with an underlying linear component mixed with white noise. Also, it was noted that generally the segmented boundaries were either very close to the human defined boundary (within 3 samples) or the boundary was not properly discovered by the segmentation algorithm.

ISBN: 978-989-95079-6-8



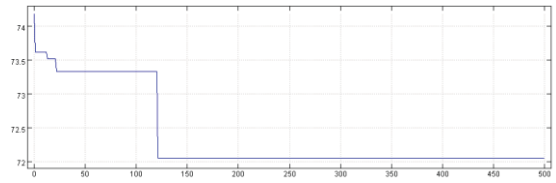
(a) DATA1 (c=3) (c)DATA3 (c=6)



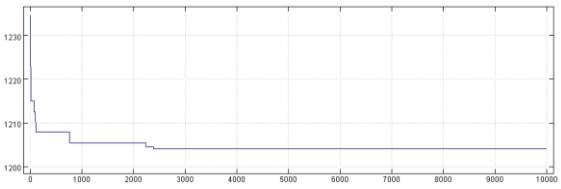
(b) DATA2 (c=5) (d) DATA4 (c=4)

Figure 3: Synthetic Data

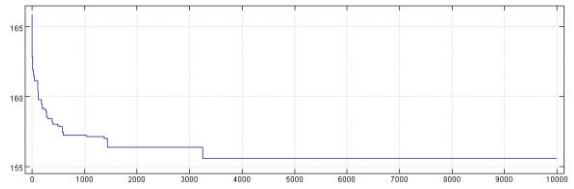
The following figure shows the convergence of differential evolution.



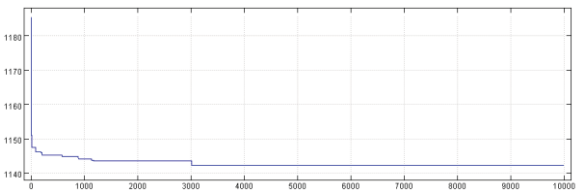
(a) DATA1



(b) DATA2



(c) DATA3



(d) DATA4

Figure 4: Fitness function in successive generations of differential evolution

Table 1: Synthetic experimental results

	P	G	p_{cr}	Num Iter.	F-value ($\theta = 3$)	Classification Rate
DATA1	100	1	0.25	500	100%	99.7%
DATA2	100	0.1	0.35	10000	50%	67.3%
DATA3	100	0.1	0.3	10000	100%	99.6%
DATA4	100	0.1	0.35	10000	66.7%	57.3%

The differential evolution parameters were fairly robust meaning that changes in the crossover rate and differential gain did not change performance dramatically. However, as the time series became more complex, i.e. DATA3, the differential evolution parameters had to be tuned to prevent the method from being trapped in local minima. From the practical point of view, the essential differential evolution parameters are differential gain, and crossover probably. It was found that increasing population size beyond 100 did not greatly improve the final boundaries. Initially the crossover probably was set to 0.5 and differential gain was set to 1. It was found through parameter adjustments that performance improved with slightly smaller crossover probability. A differential gain of 1 was sufficient on simple data sets; however, more complicated data sets such as DATA2, DATA3, and DATA4 required a smaller value for differential gain (i.e. less mutation) in order to obtain better classification rates and boundary f-value.

Table 2: Synthetic experimental results with FCM-DFS

	m	α	F-Value ($\theta = 3$)	Classification Rate
DATA1	2	0.5	66.7%	99.7%
DATA2	2	0.1	25.0%	80.3%
DATA3	2	0.4	71.4%	99.7%
DATA4	1.3	0.5	16.1%	81.2%

The accuracy of the boundaries discovered by FCM-DFS is not as high when compared with differential evolutionary multivariate segmentation on these synthetic data sets. The classification rate is higher for FCM-DFS but since the boundary f-value is much higher for the proposed method, the segmentation with DE is much more accurate with the determined boundary locations. The reason for the smaller classification rate with DE is that in some of the data sets, only one boundary is be determined correctly and its location is way-off the expected value. With FCM_DFS, several boundaries are not determined very well (accuracy is very low) but the locations are a bit closer to their expected values when a boundary is missed. The experiments on FCM-DFS were run for several different values of m {1.3, 1.5, 1.7, 2, 2.3, 2.5, 3} and several values of α {0.05, 0.1, 0.2, 0.4, 0.5, 0.6, 0.8, 0.9, 0.95}.

The segmentation approach was applied to several real valued time series [10]. The performance of the segmentation is evaluated against reconstruction error (11) as shown in Table 3.

The number of clusters was based on a visual estimate of the number of segments from the plot and then running several different values of c in close proximity to the estimated number of clusters. The results show that the segmentation algorithm does fairly well on some of the data sets with small reconstruction errors especially on power station and hog data sets.

ISBN: 978-989-95079-6-8

Table 3: Real-world time series experimental results

	P	G	p_{cr}	Num Iter.	c	Reconstruction Error
Lynx Pelts	100	0.1	0.9	4000	13	0.1253
Power Station	100	0.5	0.3	4000	12	0.0917
Hog	100	0.1	0.9	4000	16	0.0940
Housing	100	0.5	0.3	4000	17	0.1386
Housing Starts	100	2	0.3	4000	23	0.2185

5 Conclusions

The results on synthetic data show that multivariate time series segmentation show promise as a multivariate time series segmentation tool. The results are quite good on noisy synthetic data with boundary f-values reaching upto 100% and classification rates around 99%. Parameter tuning was important especially as the time series become more complicated and required smaller differential gain values, smaller crossover probability values and more iterations to achieve best results.

Future directions in research involve applying the multivariate segmentation algorithm to the problem of musical segmentation – a particularly difficult problem that automatically finds a partition in music that corresponds to the structural labels of a song, i.e. intro, verse, chorus, bridge, outro. Replacing the linear segment models with more flexible autoregressive models will be developed in future segmentation algorithms.

Acknowledgment

Support from the NSERC and Canada Research Chair (CRC) Program (W. Pedrycz) is gratefully acknowledged. This research is supported by a post-graduate PhD iCore fellowship (D. Graves).

References

[1] E. Keogh, S. Kasetty, On the need for time series data mining benchmark: a survey and empirical demonstration, *Proc. of the 8th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, Edmonton, Alberta 102-111, 2002.

[2] A. Gacek, W. Pedrycz, A genetic segmentation of ECG signals, *IEEE Trans. On Biomedical Engineering*, 50(10): 1203-1208, 2003.

[3] F.L. Chung, T.C. Fu, V. Ng, R.W.P. Luk, An evolutionary approach to pattern-based time series segmentation, *IEEE Trans on Evolutionary Computation*, 8(5): 471-489, 2004.

[4] J.A. Crossman, H. Guo, Y.L. Murphey, J. Cardillo, Automotive signal fault diagnostics – part I: signal fault analysis, signal segmentation, feature extraction and quasi-optimal feature selection, *IEEE Trans on Vehicular Technology*, 52(4): 1063-1075, 2003.

[5] P. Delacourt, C.J. Wellekens, DISTBIC: a speaker-based segmentation for audio data indexing, *Speech communications*, 32: 111-126, 2000

[6] M. Levy, M. Sandler, Structural segmentation of music audio by constrained clustering. *IEEE Trans. on Audio, Speech, and Language Processing*, 16(2): 318-326, 2008.

[7] J. Abonyi, B. Feil, S. Nemeth, P. Arva, Modified Gath-Geva clustering for fuzzy segmentation of multivariate time-series, *Fuzzy Sets and Systems*, 149(1): 39-56, 2005.

[8] C. Li, Z. Peng, S.Q. Zheng, B. Prabhakaran, Segmentation and recognition of multi-attribute motion sequences, *Proc ACM Int. Conf. Multimedia*, New York, 836-843, Oct 2004.

[9] J. Abonyi, B. Feil, S. Nemeth, P. Arva, Fuzzy clustering based segmentation of time-series, *Proc. Intelligent Data Analysis*, 2810: 275-285, 2003.

[10] R.J. Hyndman, *Time Series Data Library*, http://www.robhyndman.info/TSDL (accessed on Jan 29, 2009).

[11] M.E. Tipping, M. Bishop, Mixtures of probabilistic principal component analyzers, *Neural Computation*, 11(2): 443-482, 1999.

[12] N. Dobigeon, J.Y. Tourneret, J.D. Scargle, Joint segmentation of multivariate astronomical time series Bayesian sampling with a hierarchical model, *IEEE Transactions on Signal Processing*, 55(2): 414-423, 2007.

[13] R. Storn, K. Price, Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces, *Journal of Global Optimization*, 11: 341-359, 1997.

[14] D. Graves, W. Pedrycz, Structural segmentation of music with fuzzy clustering, *In Proc. Acoustics Week in Canada*, Vancouver, BC, Canadian Acoustics, Oct 2008.

ISBN: 978-989-95079-6-8

Coevolutionary Genetic Fuzzy System to Assess Multiagent Bidding Strategies in Electricity Markets

Igor Walter[1] Fernando Gomide[2]
1.Brazilian Electricity Regulatory Agency – ANEEL
Brasília - DF, Brazil
2.Faculty of Electrical and Computer Engineering – FEEC
University of Campinas – Unicamp
Campinas - SP, Brazil
Email: igorwalter@aneel.gov.br, gomide@dca.fee.unicamp.br

Abstract— *In this paper we study a genetic fuzzy system approach to assess suitable bidding strategies for agents in online auction environments. Assessing efficient bidding strategies allows evaluation of auction models and verification whether the mechanism design achieves its goals. Day-ahead electricity auctions are particularly explored to give an experimental instance of the approach developed in this paper. Previous works have reported successful fuzzy bidding strategies developed by genetic fuzzy systems and coevolutionary algorithms. Here we review the coevolutionary algorithm and present recent results of the bidding strategies behavior. We analyze how the evolutionary strategies perform against each other in dynamic environments. Coevolutionary approaches in which coevolutionary agents interact through their fuzzy bidding strategies permit realistic and transparent representations of the behavior of the agents in auction-based markets. They also improve market representation and evaluation mechanisms. Experimental results show that coevolutionary agents can enhance their profits at the cost of increasing system hourly price paid by demand, an undesirable outcome from the perspective of the buyers.*

Keywords— Genetic fuzzy systems, electricity markets, auctions, multiagent systems, computational economics.

1 Introduction

The restructuring process of electricity markets has raised new challenges and opportunities, because currently there is no consensual market architecture. The progress of the power industry has mirrored the lack of insight about the mechanism design issues to be addressed. An electricity market is inherently a complex system populated by self interested, interacting economic agents. Tools rooted in the neoclassical economic theory have shown to be of limited value to study the behavior of economic agents in electricity markets. The approach considered in this paper is derived from agent-based computational economics (ACE) [15] and computational intelligence tools. The purpose is to find bidding strategies for a set of agents playing in a market using genetic fuzzy systems principles. The approach can be extended to other negotiation environments and contexts.

Evolutionary bidding strategies support decisions of market players and may uncover unknown and unexpected agents behaviors that help designers to simulate and analyze negotiation mechanisms. When coevolutionary agents extract benefits of strategic behavior, the mechanism designer can review the decision making process to correct eventual unfavorable decisions and fix mechanism flaws.

Recently a coevolutionary approach was suggested to assess bidding strategies for multiagent systems. The focus was on strategies encoded in fuzzy rule-based systems [18]. The aim was to learn models represented by coevolving knowledge bases and improve the performance of the agents when acting in competitive environments. In competitive environments data for learning and tuning of knowledge bases are rare and rule bases must evolve jointly with databases. An evolutionary algorithm whose operators use a variable length chromosome, a hierarchical relationship among individuals through fitness, and a scheme that successively explores and exploits the search space along generations has also been developed [17]. In this paper we first review the coevolutionary algorithm of [18], summarize recent results of coevolutionary bidding strategies, and analyze how evolutionary agent strategies react against each other in dynamic environments.

Electricity auctions are the main application example addressed in this paper due to its practical relevance and the experimental instance they provide. The example illustrates the design of bidding strategies to exploit the negotiation space and to take advantage of market power aiming at improved negotiation mechanisms.

This paper is organized as follows. After this introduction, a brief overview of strategic bidding and survey of related work are presented in section 2. Next, section 3 summarizes the approach addressed in this paper. Section 4 presents and discusses the experimental results. Section 5 concludes the paper and suggests issues for further research.

2 Evolutionary Techniques for Electricity Market Bidding Strategies

Currently, many power industries worldwide use auctions as a mechanism of resource allocation and system coordination. The design of auction mechanisms can give agents the ability to explore market imperfections for gaming.

A energy supplier (generator) competing in an electricity market has to decide upon how much energy to offer and at which price. In a perfectly competitive market, risk averse agents have an incentive to offer energy at a price equivalent to their marginal costs [7]. Electricity markets are, however, much more an oligopoly than a laissez-faire, with low or no demand elasticity in the short term, barriers to entry, and physical constraints. Thus, a energy supplier may have an incentive to offer energy at a price other than its marginal costs and to extract some surplus from such an imperfect market.

ISBN: 978-989-95079-6-8

The behavior of a generator bidding other than marginal costs in an effort to exploit market imperfections is defined as strategic bidding [6]. The most common strategy is to maximize the expected profit but other strategies may interest a supplier exercising its market power such as competing for being a base-load generator, increasing market share, increasing profit margin, etc.

2.1 Related Work

Evolutionary approaches for strategic bidding in electricity markets have been addressed in the literature in the realm of genetic algorithms, evolutionary strategies, and classifier systems. Multiagent system approaches have been proposed to model the electricity market as well, among them [8, 14].

The evolutionary strategy approach suggested in [19] evolves a single bid value, assumed to be valid for all hours of a day ahead market. In this sense, it does not give a bidding strategy, but a bid value that is narrowly valid for a specific situation.

The approach proposed in [1] is a classifier system capable of dealing with the grid constraints of the UK market and electrical restrictions that constraint generators to be *on* or *off* by the dispatch.

In [3] a coevolutionary approach based on a coevolutionary cooperative genetic algorithm (CCGA) is devised to analyze electricity market equilibrium. Results for a 3-firms market is presented. They show that the approach finds Cournot-Nash equilibrium and Pareto solutions. In a related work [4] the same coevolutionary CCGA scheme is used together with a Cournot model and a supply function equilibrium (SFE) model of a 5-firms experiment. In this case, the coevolutionary scheme converges to a Nash equilibria.

The coevolutionary approach to model the electricity market can be based on different coevolutionary methods such as the one addressed in [2] for duopolies. In this method, discretization of the state space imposes limitations.

The coevolutionary process proposed in [11] is more complex[1]. The intent is to coevolve agents strategies and market mechanisms. The authors use a genetic programming (GP) method to evolve a function the auctioneer uses to set the price between the ask and the bid price in a double auction framework. GP is used to represent the auctioneer and traders (buyers and sellers). Simulation results based on the Nicolaisen's experiment report a function with several terms after 10,000 generations. The authors claim that the function evolved was approximately equivalent to a discriminatory price k-double auction with $k = 0.5$.

3 Coevolutionary Fuzzy Bidding Strategies

The approach employed in this paper is to coevolve fuzzy bidding strategies. Previous work [17] has devised evolutionary fuzzy bidding strategies. In [18] a coevolutionary algorithm was introduced to study how the evolving strategies react against each other in dynamic bidding environment[2]. Contrary to alternative evolutionary approaches, the aim was to model bidding information and strategies within the framework of fuzzy set theory. By enabling a fuzzy system to learn trough an evolutionary algorithm, one expects to find effective and transparent bidding strategies. We assume that bidding strategies are encoded by fuzzy rule-based systems (FRBS). As opposed to trial and error methods, genetic algorithms are especially attractive to develop bidding strategies because they can optimize the knowledge base (dimension, membership functions and rules) for a given market configuration. Genetic algorithms are also more appropriate than conventional optimization techniques because rule base tuning may involve operators and high dimensionality search spaces.

Here a GFRBS (Genetic Fuzzy Rule-Based System[3]) scheme is adopted to simultaneously evolve the data base (granularity and membership functions) and the rule base of a FRBS with the aim to find the most profitable bidding strategy. Each population of the GFRBS is a coevolving specie that represents a market agent. The result is a set of fuzzy rule-based systems able to handle imprecise data typically found in auction environments.

The evolutionary approach detailed in [17] is used to evolve the data base employing a variable length chromosome to represent the rule base. In the rule base, both, the number of rules and size of each rule may change during the evolutionary process. A particularly suitable crossover operation detailed in [17] was developed to enhance system performance.

Many coevolutionary approaches emphasize explicit competition such as the host-parasite scheme to minimize sorting networks in [10]. The coevolutionary approach focused in this paper is based on the CCGA of Potter and De Jong [13]. As detailed below in Algorithm 1, our approach selects a representative agent for each coevolutionary specie. The species model bidding agents. The representatives are the ones who compete in the auctions. When representatives bid, their strategy may improve the profit of the remaining bidders. Thus, the remaining coevolutionary species may act in a tacit cooperation in this competitive game. Therefore, although CCGA was originally developed to coevolve subcomponents of a problem in an explicit cooperative framework, we use its main idea in a competitive environment: auction in electricity markets[4].

3.1 Genetic Algorithm

The genetic algorithm employed in this paper uses a variable length chromosome representation for both: data base and rule bases. The encoding schemes are similar to the ones adopted in [17]. Crossover operations in data base and rule base are synchronous, that is, crossover points of database and rule base induce exchange of chromosome portions to keep the rule structure and fuzzy terms meaningful in their offspring.

The first step to design a GFRBS must decide which part of the FRBS will be optimized by the genetic algorithm [5]. This decision usually means a trade off between granularity and search efficiency. In the approach adopted here, the following FRBS components are evolved: database (granularity and membership functions parameters) and rule base (number of "active" rules[5] and rule structure).

[1] A more complete report of this work can be found in [12].

[2] This and all previous papers are a result of the PhD thesis of the first author and has no links with the Regulatory Agency.

[3] For a taxonomy, a comprehensive survey of past and recent developments, as well as future trends of genetic fuzzy systems see [9].

[4] A similar approach was adopted in [3].

[5] An "active" rule is a rule that is processed during fuzzy inference while an "inactive" rule remains in the rule base genotype but it is not used during fuzzy inference.

ISBN: 978-989-95079-6-8

The GA also assumes the simultaneous evolution of the FRBS in the sense that each individual represents a complete fuzzy rule-based system as in the Pittsburgh approach. The FRBS is encoded in a chromosome Cr composed by three major components (Cr_1, Cr_2, and Cr_3) to represent the data and rule base.

More specifically, the chromosome component Cr_1 encodes granularity in a variable length chain of integers, and Cr_2 encodes membership functions parameters in a variable length chain of real numbers. The number of rules, rules size and the rules themselves define chromosome component Cr_3. Component Cr_3 has, in addition to two integers (number of rules and rules size), several variable chains of alleles of '0' or '1', where a '1' means that a linguistic term is used in the rule, otherwise it is '0'.

Algorithm 1 Coevolutionary GFRBS Algorithm

```
t ← 0
for each specie s do
   create random initial population Ω_0^s
end for
repeat
   for each specie s do
      for each specie r such that r ≠ s do
         choose a representative in Ω_t^r
      end for
      evaluate each individual in Ω_t^s through fitness function
      with the chosen representatives
      select parents in Ω_t^s based on relative fitness in Ω_t^s
      apply crossover and mutation on parents to produce
      offspring Ω_{t+1}^s
      t ← t + 1
   end for
until (stop criteria is satisfied)
```

As stated before, the coevolutionary algorithm detailed in Algorithm 1 is based on CCGA [13]. The specifics of the genetic operators are described next.

3.2 Genetic Operators

Due to the particular chromosome encoding structure used to represent the knowledge base, the nature of the genetic operators becomes an important issue to effectively evolve the FRBS. Since there are strong relationships among chromosome components, we need operators that work cooperatively on the chromosome.

Selection is performed using *roulette wheel*. We adopt an elitist model, but the best individual is not guaranteed to be selected for crossover, albeit kept intact in the offspring.

Crossover of the database follows the approach detailed in [17]. Summing up, two different crossover operators are used depending on whether the selected individuals have the same granularity or not. When the granularity is the same, a promising zone in the search space is found and must be appropriately exploited. In this case, the granularity of the offspring database (Cr_1) is kept the same and the membership functions parameters (Cr_2) combined following the max-min-arithmetic algorithm of [17]. When the selected pair has different granularity, a random crossover position p is chosen. Both, granularity (Cr_1) and the corresponding parameters of the membership functions (Cr_2) are recombined. Crossover operations of database and rule base are synchronous: crossover points of database and rule base induce exchange of chromosome portions that keep the rule structure and terms meaningful in their offspring.

Different operators can be used to mutate chromosomes, similarly as reported in [17]: a local variation is introduced to the granularity by adding or subtracting one with equal probability, mutation of membership function parameters uses the Michalewicz non-uniform mutation operator, and rule bases are mutated via the standard, bitwise reversing operation.

3.3 Electricity Market

The application example focused in this paper is electricity markets. The approach can, however, be extended to distinct negotiation applications. The coevolutionary GFRBS agents are thermal power suppliers that have to decide how to bid in auctions. The coevolutionary GFRBS species were evolved using a day-ahead electricity auction. We assume that the remaining agents are non evolutionary and conservative competitors. Conservative agents offer energy bidding all their capacity at their marginal costs [7]. The negotiation protocol is an uniform price sealed bid auction: the price paid for the energy is the same for all accepted bids and is equal to the last bid accepted. The auctioneer decides the hourly dispatch choosing the electricity generator agents outputs and the corresponding price to minimize overall energy cost. Cost minimization results in a merit order dispatch procedure, i.e., bids at lower prices come first. The performance of the best FRBS obtained were verified using test demand data for two weeks day-ahead auctions against conservative and evolutionary competitors. The results are presented and discussed in section 4.

The experiments use actual public data made available from the Brazilian Independent System Operator (ISO)[6]. Demand is supposed to be known to all the participant agents. The data used to evolve the GFRBS is the load of the week beginning on May 19th, 2002, for a portion of the national grid: the South sub-market. The fitness of the coevolutionary agents are the profit on the electricity auction of the decoded GFRBS for the training week. The best individual found by training is tested using the next two remaining weeks period that begins on May 26th 2002 [7].

3.4 Cost Function

A pool of power generator agents was setup based on actual, ISO's publicly available data. Electrical constraints and geoelectrical location are neglected in this work: all the plants are supposed to be in the same sub-market. There is some excess of supply and the running costs of producing electricity in coal, gas and oil plants have been modeled as quadratic functions of the power P_s supplied by the agent, expression (1). Nuclear plants are assumed to have linear cost functions.

$$F(P_s) = a + bP_s + cP_s^2 \quad \text{[GJ/h]} \tag{1}$$

The supplier cost function $C^j(.)$ is given by $F(P_s)$ multiplied by the fuel cost in \$/GJ. Hence costs are quadratic func-

[6] ONS: Operador Nacional do Sistema.

[7] These same data set was used in [17, 18]. Data set contains 504 samples; 168 used for training and 336 for testing.

ISBN: 978-989-95079-6-8

Table 1: Thermal power generator agents characteristics and cost functions.

Plant	Type	Capacity (MW)	Marginal cost ($/MWh)	$C^j(.)$
Ibirité	Gas	766.5	39.065	$3,632.08 + 31.966g + 0.00463g^2$
TermoRio	Gas	824.7	39.109	$3,904.05 + 31.912g + 0.00436g^2$
Argentina I	Gas	1,018	41.045	$4,459.61 + 32.775g + 0.00406g^2$
Argentina II	Gas	1,000	41.046	$4,379.82 + 32.774g + 0.00414g^2$

tions given by (2), where g_h^j is the amount supplied by agent j at hour h.

$$C^j(g_h^j) = \alpha + \beta g_h^j + \gamma {g_h^j}^2 \quad [\$/h] \tag{2}$$

3.5 Competitive Environment

We assume power demand D_h inelastic with price. Therefore, the auctioneer must assure, as commonly found in a single buyer auction, that for each hour h: $\sum_{j=1}^{T_h} g_h^j = D_h$, where g_h^j is the power supplied by agent j and T_h is the number of suppliers. Thus, allocation would cost the market $D_h \pi_h(D_h)$. The supplier agents profit is given by(3):

$$P_h^j = \pi_h g_h^j - C^j(g_h^j) \quad [\$/h] \tag{3}$$

Supplier agents (generators) must internalize all costs to a simple bid, a pair (q_h^j, p_h^j) of the amount offered (in MW) and its price ($/MWh), where the amount q_h^j is less than or equal to the agent capacity, G^j. Conservative agents bid a pair $(G^j, MC^j(G^j))$ where $MC^j(G^j)$ is the marginal cost at capacity G^j.

Table 1 shows the thermal generator agents characteristics for the agents we have chosen to coevolve. The complete data set of the thermal generation park can be found in [17]. The fourth column, the marginal cost ($/MWh), is the marginal cost at full capacity corresponding to conservative bids. The last column gives the cost functions of the thermal suppliers[8].

3.6 Fitness

Algorithm 2 Evaluation of Population Individuals

for each individual i such as $i \in \Omega_t^s$ **do**
  decode database and rule base of i as agent A_i
  add the agent A_i to the market
  simulate market (run auction) for the training period
  keep fitness F_{A_i} as the profit for the period
  remove A_i from the market
end for

The fitness of a GFRBS individual is defined as the profit of the corresponding agent during the training week. Algorithm 2 summarizes the evaluation of the fitness of each individual of the population.

4 Results

The experiments reported in this paper use a population of 40 individuals. Each individual encodes a single input-single output (SISO) FRBS: the input is the hourly load and the output the corresponding the bid price. The genetic operators described in section 3.2 were employed in all the experiments[9] with probability of crossover 0.5 and probability of mutation 0.01. We let the evolutionary process run for 500 or 1,000 generations[10]. While several experiments were performed with 2 to 5 species, some of them did not achieve stable behavior. For instance, for 5 coevolutionary species stability was not attained after 1,000 generations. Section 4.2 discusses briefly why stable behavior may not be necessarily expected, as the 5 coevolutionary species case shows.

Table 2: Comparison between coevolutionary and conservative strategies.

conservative strategy			
agent	hours	energy (MWh)	profit ($)
Ibirité	332	232,954	849,960
TermoRio	277	199,390	959,133
Argentina I	202	179,595	1,065,262
Argentina II	149	116,381	1,049,804
coevolutionary strategy: testing			
agent	hours	energy (MWh)	profit ($)
Ibirité	334	255,856	2,082,681
TermoRio	336	248,367	2,237,294
Argentina I	150	89,969	1,825,135
Argentina II	335	301,710	2,304,976
variation: coevolutionary / conservative			
agent	hours	energy (MWh)	profit ($)
Ibirité	+0.60%	+9.83%	+145.03%
TermoRio	+21.30%	+24.56%	+133.26%
Argentina I	-25.74%	-49.90%	+71.33%
Argentina II	+124.83%	+159.24%	+119.56%

4.1 Experiment: Four Coevolutionary Strategies

In [18] two thermal plants coevolved. They are the same used as evolutionary agents in [17]: Argentina I and Argentina II, respectively.

Here we report an experiment in which Argentina I and Argentina II are coevolved together with other two natural gas plants: TermoRio and Ibirité. The fitness of the best individual of each specie during 1,000 generations of the evolutionary process (training period) is depicted in Fig. 1. Notice that, after a short unstable period lasting about 100 genera-

[8] g_h^j is denoted by g for short.

[9] The choice of the evolutionary parameters is not subject to any optimization process.

[10] The training process can take about 2 hours of processing time for 2 species, corresponding to 500 generations, and above 6 hours for 4 species and 1,000 generations. Experiments were done using a Pentium 4,2 GHz 256 Mb RAM PC running GNU/Linux Fedora.

ISBN: 978-989-95079-6-8

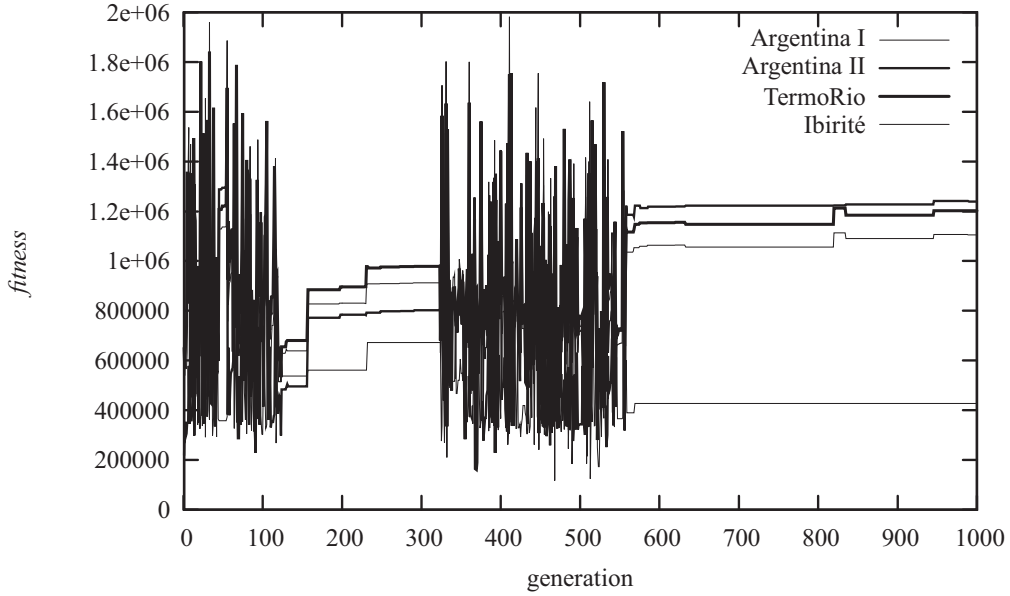


Figure 1: Fitness for coevolutionary species (4 thermal plants: Argentina I, Argentina II, TermoRio and Ibirité).

tions, the fitness of the best individual of each specie increases and stabilizes after about 200 generations and suddenly fitness behavior becomes unstable again. After approximately another 250 generations, the fitness becomes stable again, beginning at 550^{th} generation, and remains stable until the end of the evolutionary process window (1,000 generations). Notice also that the resulting outcome for the specie representing Argentina I becomes worse than in the one achieved during the first stable interval. This behavior suggests that the other three thermal plants may have learned how to exploit Argentina I agent weakness. All the remaining agents increase their profits. A close look at the fitness values shows that the species interact with each other during the whole evolutionary process.

After the end of the evolutionary process, the best individual of each specie is decoded into the corresponding bidding strategies to be played by the four agents. Table 2 summarizes the energy produced and profits obtained when the agents compete in the electricity market using the coevolutionary strategies. Table 2 shows how evolutionary agents outperform the corresponding conservative strategy of bidding their marginal costs at full capacity. The last four rows show the variation when using conservative and evolutionary strategies. We notice in Table 2 that Argentina I decreases energy production by half and still, with the hourly prices that result from the coevolutionary strategic bidding, increases its profit. This is a typical example of an undesirable behavior (from the perspective of the buyers) that mechanism designers wants to avoid before putting any market to work.

Clearly, the coevolutionary strategies affected the electric energy prices. The maximum hourly price of energy reached in the testing period was the same for both, purely conservative strategies or coevolutionary strategies, but coevolution made the average price 10.6% higher. The average price increase for the experiment reported in [18] was smaller, about 7.1%. Similarly as in [18], the individual hourly prices can be up to 53.2% higher for the coevolutionary strategies than purely conservative ones.

4.2 Multiagent Learning Equilibrium

It is important clarify that, while the results presented here and in [18] converge to a stable outcome, some of the experiments performed did not achieve stable outcome at all, remaining unstable for 500 or even 1,000 generations. Unstable outcomes typically occurred when there are many evolutionary agents (5 thermal generators in our experiments). While the dynamics of unstable behavior still is to be investigated, it does not seem to be a novelty, as it has been pointed out by Vidal in [16], who asserts that most multiagent learning systems do not necessarily converge to an equilibrium or stable behavior.

One reason for designers to use learning agents is because they do not know, at design time, the specific circumstances that the agents will face at run time. We will often see a multiagent system with learning agents when the designer can neither predict that an equilibrium solution will be found, nor which equilibria might emerge. The result is a form of closed loop feedback evolving system in which evolution and learning play a complementary role.

As stated in [16], the main reasons behind the difficulty to predict equilibrium solutions of a system include the existence of unpredictable environmental changes that affect the payoffs of the agents, and the fact that, in many systems, an agent only has access to its own set of payoffs. These reasons partially justify the difficulty a designer finds to predict equilibria, if any, a system could reach. However, the agents in a system may still play a game for which an equilibrium exists, even though the designer cannot predict it at design-time. Since in general the payoffs change, it is often the case that the agents are constantly modifying their strategies as an attempt to get better payoffs. This may result in unstable outcomes.

5 Conclusion

A coevolutionary genetic fuzzy systems approach to develop fuzzy bidding strategies was suggested in this paper. Coevo-

ISBN: 978-989-95079-6-8

lutionary approaches provide more realistic representation of agents in auction-based environments because they allow the bidding strategies to interact during the evolutionary process.

The results reported here for an electricity market example show that coevolution can improve agents profits at the expense of increasing the average electricity hourly price paid by demand, an outcome that is undesirable from the perspective of the buyers, and for auction based markets in general. Further research is needed to obtain design methods for bidding strategies in multiagent systems framework populated by intelligent agents. In particular, stability and equilibrium analysis of evolutionary learning in multiagent systems environments still remains a challenge.

Acknowledgment

The first author acknowledges Fábio Stacke for his helpful comments. The second author acknowledges the CNPq, the Brazilian National Research Council, for grant #304 857/2006-8. The authors are also grateful to the anonymous referees for the comments that helped to improve the paper.

Disclaimer

The results, interpretations and conclusions expressed in this work are of exclusive responsibility of its authors and should not be, in any hypothesis, attributed to ANEEL, the Brazilian Electricity Regulatory Agency, neither to its Board of Directors, nor to any Commission the author is affiliated to. The experiments presented in this paper have been accomplished based on general public available data. ANEEL is neither responsible for this work nor any consequence from its use.

References

[1] A. J. Bagnall. A multi-adaptive agent model of generator bidding in the UK market in electricity. In *Genetic and Evolutionary Computation Conference GECCO 2000*, pages 605–612. Morgan Kaufmann, 2000.

[2] T. D. H. Cau and E. Anderson. A co-evolutionary approach to modelling the behaviour of participants in competitive electricity markets. In *2002 IEEE Power Engineering Society Summer Meeting*, volume 3, pages 1534–1540, Chicago, IL, USA, July 2002.

[3] H. Chen, K. Wong, D. Nguyen, and C. Chung. Analyzing oligopolistic electricity market using coevolutionary computation. *IEEE Transactions on Power Systems*, 21(1):143–152, Feb. 2006.

[4] H. Chen, K. Wong, X. Wang, and C. Chung. A coevolutionary approach to modeling oligopolistic electricity markets. In *IEEE 2005 Power Engineering Society General Meeting*, volume 1, pages 230–236, June 2005.

[5] O. Cordón, F. Herrera, F. Hoffman, and L. Magdalena. *Genetic Fuzzy Systems: Evolutionary tuning and learning of fuzzy knowledge bases*, volume 19 of *Advances in Fuzzy Systems: Applications and Theory*. World Scientific, Singapore, 2001.

[6] A. K. David and F. S. Wen. Strategic bidding in competitive electricity markets: A literature survey. In *IEEE PES 2000 Summer Power Meeting*, volume 4, pages 2168–2173, Seattle, USA, July 2000. IEEE Power Engineering Society, IEEE.

[7] R. Green. Competition in generation: The economic foundations. *Proceedings of the IEEE*, 88(2):128–139, Feb. 2000.

[8] S. A. Harp, S. Brignone, B. F. Wollenberg, and T. Samad. SEPIA: A simulator for electric power industry agents. *IEEE Control Systems Magazine*, 20(4):53–69, Aug. 2000.

[9] F. Herrera. Genetic fuzzy systems: Taxonomy, current research trends and prospects. *Evolutionary Intelligence*, 1:27–46, 2008.

[10] W. D. Hillis. Co-evolving parasites improve simulated evolution as an optimization procedure. *Physica D*, 42:228–234, 1990.

[11] S. Phelps, S. Parsons, P. McBurney, and E. Sklar. Co-evolution of auction mechanims and trading strategies: Towards a novel approach to microeconomic desing. In *Proceedings of the 2nd Workshop on Evolutionary Computation and Multi-Agent Systems*, New York, NY, 2002.

[12] S. G. Phelps. *Evolutionary Mechanism Design*. PhD thesis, Univeristy of Liverpool, UK, July 2007.

[13] M. A. Potter and K. A. De Jong. Cooperative coevolution: An architecture for evolving coadapted subcomponents. *Evolutionary Computation*, 8(1):1–29, 2000.

[14] I. Praça, C. Ramos, Z. Vale, and M. Cordeiro. A new agent-based framework for the simulation of electricity markets. In *Proceedings of the IEEE/WIC International Conference on Intelligent Agents (IAT'03)*, pages 1931–1938, Halifax, Canada, 2003. IEEE.

[15] L. Tesfatsion and K. L. Judd, editors. *Handbook of Computational Economics: Agent-Based Computational Economics*, volume 2 of *Handbooks in Economics*. North Holland, 2006.

[16] J. M. Vidal. Learning in multiagent systems: An introduction from a game-theroretic perspective. *Lecture Notes on Artifical Intelligence*, 2636:202–215, 2003.

[17] I. Walter and F. Gomide. Design of coordination strategies in multiagent systems via genetic fuzzy systems. *Soft Computing*, 10(10):903–915, Aug. 2006. Special Issue: New Trends in the Design of Fuzzy Systems.

[18] I. Walter and F. Gomide. Coevolutionary fuzzy multiagent bidding strategies in competitive electricity markets. In *3^{rd} International Workshop on Genetic and Evolving Fuzzy Systems (GEFS 08)*, pages 53–58, Witten-Bommerholz, Germany, Mar. 2008. IEEE.

[19] G. Xiong, T. Hashiyama, and S. Okuma. An evolutionary computation for supplier bidding strategy in electricity auction market. In *IEEE Power Engineering Society Transmission and Distribution Conference*, volume 1, pages 83–88, Yokohama, Japan, 2002. IEEE.

ISBN: 978-989-95079-6-8

A Type-2 Fuzzy Portfolio Selection Problem Considering Possibility Measure and Crisp Possibilistic Mean Value

Takashi Hasuike, Hiroaki Ishii

Graduate School of Information Science and Technology, Osaka University, Japan
2-1 Yamadaoka, Suita, Osaka 565-0871, Japan
Email: {thasuike, ishii}@ist.osaka-u.ac.jp

Abstract—*This paper considers a portfolio selection problem with type-2 fuzzy future returns involving ambiguous and subjectivity. Since this proposed problem is not well-defined due to fuzziness, introducing the fuzzy goal for the total future return and the degree of possibility, the main problem is transformed into the standard fuzzy programming problem including the secondary fuzzy numbers. Furthermore, using the hybrid solution approaches based on the linearity of the deterministic equivalent problem and the crisp possibilistic mean value, the efficient solution is constructed.*

Keywords— Portfolio selection, Type-2 fuzzy, Possibility measure, Crisp possibilistic mean value, Efficient solution method

1 Introduction

Portfolio selection problems are standard and most important problems in investment and financial research fields, and various studies have been so far performed. Furthermore, in recent investment fields, not only big companies and institutional investors but also individual investors called Day-Traders invest in stock, currency, land and property. Therefore, the role of investment theory called portfolio theory becomes more and more important. As for the research history on mathematical approach, Markowitz (Markowitz [18]) has proposed the mean-variance analysis model. It has been central to research activity in the real financial field and numerous researchers have contributed to the development of modern portfolio theory (cf. Elton and Gruber [4], Luenberger [17]). On the other hand, many researchers have proposed models of portfolio selection problems which extended Markowitz model; mean-absolute deviation model (Konno [13], Konno, et al. [14]), safety-first model [4], Value at Risk and conditional Value at Risk model (Rockafellar and Uryasev [19]), etc.. As a result, nowadays it is common practice to extend these classical economic models of financial investment to various types of portfolio models. In practice, after Markowitz's work, many researchers have been trying different mathematical approaches to develop the theory of portfolio selection.

Particularly, the prediction of future returns is one of the most important factors in theoretical and practical investment, and they have been treated as only random variables in many previous studies. Then, the expected returns and variances also have been assumed to be fixed values. However, investors receive effective or ineffective information from the real world and ambiguous factors usually exist in it. Furthermore, investors often have the subjective prediction for future returns which are not derived from the statistical analysis of historical data. Then, even if investors hold a lot of information from the investment field, it is difficult that the present or future random distribution of each asset is strictly set. Consequently, we need to consider not only random conditions but also ambiguous and subjective conditions for portfolio selection problems.

Recently, in the sense of mathematical programming, some researchers have proposed various types of portfolio models under randomness and fuzziness. These problems with probabilities and possibilities are generally called stochastic programming problems and fuzzy programming problems, respectively, and there are some basic studies using a stochastic programming approach, goal programming approach, etc.., and fuzzy programming approach to treat ambiguous factors as fuzzy sets (Inuiguchi and Ramik [8], Leon, et al. [15], Tanaka and Guo [20], Tanaka, et al. [21], Watada [23]). Furthermore, some researchers have proposed the mathematical programming problems with both randomness and fuzziness as fuzzy random variables (for instance, Katagiri et al. [11, 12]). In the studies [11, 12], fuzzy random variables were related with the ambiguity of the realization of a random variable and dealt with a fuzzy number that the center value occurs according to a random variable. Then, Yazenin considered some models for portfolio selection problems in the probabilistic-possibilistic environment, that profitabilities of financial assets are fuzzy random variables (Yazenin [25, 26]). On the other hand, future returns may be dealt with random variables derived form the statistical analysis, whose parameters are assumed to be fuzzy numbers due to the decision maker's subjectivity, i.e., random fuzzy variables which Liu (Liu [16]) defined. There are a few studies of random fuzzy programming problem (Katagiri et al. [9, 10], Huang [7]). Most recently, Hasuike et al. [6] proposed several portfolio selection models including random fuzzy variables and developed the analytical solution method.

However, in [6], the random distribution of each asset is assumed to be a normal distribution. From some practical studies with respect to the present practical market, it is not clear that price movements of assets occur according to normal distributions. In fact, considering the existence of various types of investors in the practical market and the subjectivity of investors, it is important that we need to develop a new portfolio selection problem to deal with a lot of subjectivity. In this paper, we assume future returns to be Type-2 fuzzy numbers which can be dealt with various types of membership functions, and propose a new Type-2 portfolio selection problem.

ISBN: 978-989-95079-6-8

In the sense of mathematical programming, since the proposed model is not formulated as a well-defined problem due to fuzziness, we need to set some certain optimization criterion so as to transform into well-defined problems. In this paper, introducing a fuzzy goal and the degree of possibility, we transform the main problem into the possibility maximization problem. However, this problem includes the fuzzy number and it is also not a well-defined. Therefore, in order to solve possibility maximization problem analytically, we introduce the crisp possibilistic mean value proposed by Carlsson and R. Fullér [1] and develop an efficient solution method to find a global optimal solution of deterministic equivalent problem.

This paper is organized in the following way. In Section 2, we introduce mathematical concepts of type-2 fuzzy set and crisp possibilistic mean value. In Section 3, we propose a type-2 fuzzy portfolio selection problem maximizing the total future return. Then, introducing the degree of possibility and the crisp possibilistic mean value, we transform the proposed model into the deterministic equivalent problem. In Section 4, in order to compare our proposed models with other models for portfolio selection problems, we provide a numerical example derived from current practical market data. Finally, in Section 5, we conclude this paper.

2 Mathematical concept

2.1 Type-2 fuzzy set

A type-2 fuzzy set is a set in which we also have uncertainty about the membership function, i.e., a type-2 fuzzy set is characterized by a fuzzy membership function whose grade for each element is a fuzzy set [0,1] (Castillo and Melin [3]). In the real world, there are many problems where the decision maker cannot determine the exact form of the membership function such as in time series prediction because of noise in the data. Therefore, it is important that the type-2 fuzzy set is introduced into real world problems.

Example 1 (Castillo and Melin [3])
Consider the case of a fuzzy set characterized by a Gaussian membership function with mean m and a standard deviation that can take values in $[\sigma_1, \sigma_2]$, i.e.,

$$\mu(x) = \exp\left\{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right\},\ \sigma \in [\sigma_1, \sigma_2]$$

In this paper, with respect to future returns we treat the following *L*-shape fuzzy numbers as fuzzy numbers based on the type-2 fuzzy set:

$$\mu_{\tilde{r}_j}(\omega) = \max\left\{0, L\left(\frac{\omega - \bar{r}_j}{\alpha_j}\right)\right\},\ \alpha_j \in [\alpha_j^L, \alpha_j^U] \tag{1}$$

where $L(\omega)$ is a shape function from $\mathbb{R}$ to $\mathbb{R}$ satisfying the following conditions:

1. $L(-\omega) = L(\omega)$ for $\forall \omega \in \mathbb{R}$
2. $L(0) = 1$
3. $L(\cdot)$ is nonincreasing on $[0, \infty)$
4. Let $t_0 = \inf\{t > 0 | L(t) = 0\}$. Then $0 < t_0 < \infty$

The *L*-shape fuzzy number includes the more general membership function than the Gaussian in Example 1. Therefore, by using the *L*-shape fuzzy number, it is possible that we represent more versatile social problems.

2.2 Crisp possibilistic mean value

Carlsson and Fullér [1] introduced the notation of crisp possibilistic mean value of continuous possibility distribution, which are consistent with the extension principle. Let A a fuzzy number. Then, $[A]^\gamma$ denote the γ -level set of A as the following form:

$$[A]^\gamma = [a_L(\gamma),\ a_U(\gamma)]$$

Using this γ -level set of A, the crisp possibilistic mean value of A is introduced as follows:

$$E(A) = \int_0^1 \gamma\left(a_L(\gamma) + a_U(\gamma)\right) d\gamma$$

Example 2
In the case that a fuzzy number A is a trapezoidal fuzzy number with tolerance interval $[a, b]$, left spread α and right spread β. Then, the crisp possibilistic mean value of A is

$$\begin{aligned} E(A) &= \int_0^1 \gamma\left(a - (1-\gamma)\alpha + b + (1-\gamma)\beta\right) d\gamma \\ &= \frac{a+b}{2} + \frac{\beta - \alpha}{6} \end{aligned}$$

This crisp possibilistic mean value has been applied to various types of portfolio selection problems with fuzzy numbers (in detail, see [5]), and it becomes one of the most useful tool in the sense of fuzzy programming problem.

3 Type-2 fuzzy portfolio selection problem

In this paper, we deal with the following portfolio selection problem with future returns based on the type-2 fuzzy set maximizing total future returns.

$$\begin{aligned} &\text{Maximize} && \sum_{j=1}^n \tilde{r}_j x_j \\ &\text{subject to} && \sum_{j=1}^n x_j = 1, \quad 0 \le x_j \le p_j, \quad j = 1, 2, \cdots, n \end{aligned} \tag{2}$$

where $\boldsymbol{x} = (x_1, \cdots, x_n)$ is a composition of the portfolio and p_j is a limited upper rate of *j*th asset. In this problem, the objective function $\tilde{Z} = \sum_{j=1}^n \tilde{r}_j x_j$ is also a fuzzy number characterized by the following membership function:

ISBN: 978-989-95079-6-8

$$\mu_{\tilde{Z}}(\omega)=\max\left\{0,L\left(\frac{\omega-\sum_{j=1}^{n}\overline{r}_j x_j}{\sum_{j=1}^{n}\alpha_j x_j}\right)\right\},\ \alpha_j\in\left[\alpha_j^L,\alpha_j^U\right] \tag{3}$$

Due to the fuzziness of objective function, problem (2) is not a well-defined problem, and so we need to set some certain optimization criterion. Until now, researchers have proposed some solution approaches to solve fuzzy programming problem based on not only strict solution methods such as linear and nonlinear programming but also approximate solution methods such as GA and NN. In this paper, we focus on the strict solution method in the sense of mathematical programming, and so we introduce the possibility measure as follows:

$$\mathrm{Pos}\left\{\sum_{j=1}^{n}\tilde{r}_j x_j\geq f\right\}=\sup_{\omega}\left\{\mu_{\tilde{Z}}(\omega)\middle|\omega\geq f\right\}$$

where f is the target value to total future return. The possibility measure means that the total future return is more than f as much as possible in the aspiration level of each investor.

3.1 Possibility maximization model for the proposed type-2 fuzzy portfolio selection problem

On the other hand, in practical situations, the investor considers increasing the goal of total future return and that of possibility, simultaneously. Furthermore, considering many real decision cases and taking account of the vagueness of human judgment and flexibility for the execution of a plan, the investor often has subjective and ambiguous goals with respect to the target return f such as "Total future return $\sum_{j=1}^{n}\tilde{r}_j x_j$ is approximately larger than f_1 .". In this subsection, we propose the more flexible model considering the aspiration level to the goal for the total future return. We represent the subjective and ambiguous goals with respect to f as a fuzzy goal characterized by the following membership function:

$$\mu_{\tilde{G}}(f)=\begin{cases}1 & f_1\leq f\\ g(f) & f_0\leq p<f_1\\ 0 & f<f_0\end{cases} \tag{4}$$

where $g(f)$ is the strict increasing function. Furthermore, using a concept of possibility measure, we introduce the degree of possibility as follows:

$$\prod_{\tilde{Z}}(\tilde{G})=\sup_{f}\min\left\{\mu_{\tilde{Z}}(f),\mu_{\tilde{G}}(f)\right\} \tag{5}$$

Using this degree of possibility, we formulate a possibility maximization model for the proposed portfolio selection problem as the following form:

$$\begin{aligned}&\text{Maximize}\quad \prod_{\tilde{Z}}(\tilde{G})\\ &\text{subject to}\quad \sum_{j=1}^{n}x_j=1,\quad 0\leq x_j\leq p_j,\quad j=1,2,\cdots,n\end{aligned} \tag{6}$$

Then, this problem is equivalently transformed into the following problem introducing a parameter h:

$$\begin{aligned}&\text{Maximize}\quad h\\ &\text{subject to}\quad \prod_{\tilde{Z}}(\tilde{G})\geq h,\\ &\qquad\sum_{j=1}^{n}x_j=1,\quad 0\leq x_j\leq p_j,\quad j=1,2,\cdots,n\end{aligned} \tag{7}$$

In this problem, the possibility constraint $\prod_{\tilde{Z}}(\tilde{G})\geq h$ is transformed into the following form:

$$\begin{aligned}&\prod_{\tilde{Z}}(\tilde{G})\geq h\\ \Leftrightarrow\ &\sup_{f}\min\left\{\mu_{\tilde{Z}}(f),\mu_{\tilde{G}}(f)\right\}\geq h\\ \Leftrightarrow\ &\mu_{\tilde{Z}}(f)\geq h,\ \mu_{\tilde{G}}(f)\geq h\\ \Leftrightarrow\ &L\left(\frac{f-\sum_{j=1}^{n}\overline{r}_j x_j}{\sum_{j=1}^{n}\alpha_j x_j}\right)\geq h,\ f\geq g^{-1}(h)\\ \Leftrightarrow\ &f\leq\sum_{j=1}^{n}\overline{r}_j x_j+L^*(h)\sum_{j=1}^{n}\alpha_j x_j,\ f\geq g^{-1}(h)\\ \Leftrightarrow\ &\sum_{j=1}^{n}\overline{r}_j x_j+L^*(h)\sum_{j=1}^{n}\alpha_j x_j\geq g^{-1}(h)\end{aligned} \tag{8}$$

Using these inequalities, the main problem (6) is equivalently transformed into the following problem:

$$\begin{aligned}&\text{Maximize}\quad h\\ &\text{subject to}\quad \sum_{j=1}^{n}\overline{r}_j x_j+L^*(h)\sum_{j=1}^{n}\alpha_j x_j\geq g^{-1}(h),\\ &\qquad\sum_{j=1}^{n}x_j=1,\quad 0\leq x_j\leq p_j,\quad j=1,2,\cdots,n\end{aligned} \tag{9}$$

In this problem, if each spread of fuzzy number α_j is fixed, this problem is analytically solved by using the hybrid solution method of bisection algorithm on parameter h and the linear programming problem. However, in the case that α_j is not fixed but a fuzzy number, it is difficult that this problem is analytically solved by standard linear programming approaches due to including secondary fuzzy numbers.

3.2 Solution method based on the crisp possibilistic mean value

In order to solve problem (9) in the sense of mathematical programming, we introduce the crisp possibilistic mean value. Subsequently, we assume that each spread $\tilde{\alpha}_j$ is the following trapezoidal fuzzy number:

ISBN: 978-989-95079-6-8

$$\mu_{\tilde{\alpha}_j}(\omega)=\begin{cases}\dfrac{\omega-\alpha_j^L}{\underline{\alpha}_j-\alpha_j^L} & \left(\alpha_j^L\le\omega<\underline{\alpha}_j\right)\\ 1 & \left(\underline{\alpha}_j\le\omega\le\overline{\alpha}_j\right)\\ \dfrac{\alpha_j^U-\omega}{\alpha_j^U-\overline{\alpha}_j} & \left(\overline{\alpha}_j<\omega\le\alpha_j^U\right)\\ 0 & \text{otherwise}\end{cases},\ j=1,2,...,n \qquad (10)$$

Using these trapezoidal fuzzy numbers and the fuzzy extension principle, the membership function of fuzzy number $\tilde{\alpha}(\boldsymbol{x})=\sum_{j=1}^{n}\tilde{\alpha}_j x_j$ is also the following trapezoidal fuzzy number:

$$\mu_{\tilde{\alpha}(\boldsymbol{x})}(\omega)=\begin{cases}\dfrac{\omega-\sum_{j=1}^{n}\alpha_j^L x_j}{\sum_{j=1}^{n}\underline{\alpha}_j x_j-\sum_{j=1}^{n}\alpha_j^L x_j} & \left(\sum_{j=1}^{n}\alpha_j^L x_j\le\omega<\sum_{j=1}^{n}\underline{\alpha}_j x_j\right)\\ 1 & \left(\sum_{j=1}^{n}\underline{\alpha}_j x_j\le\omega\le\sum_{j=1}^{n}\overline{\alpha}_j x_j\right)\\ \dfrac{\sum_{j=1}^{n}\alpha_j^U x_j-\omega}{\sum_{j=1}^{n}\alpha_j^U x_j-\sum_{j=1}^{n}\overline{\alpha}_j x_j} & \left(\sum_{j=1}^{n}\overline{\alpha}_j x_j<\omega\le\sum_{j=1}^{n}\alpha_j^U x_j\right)\\ 0 & \text{otherwise}\end{cases},\ j=1,2,...,n \qquad (11)$$

Then, the γ -level set of $\tilde{\alpha}(\boldsymbol{x})$ is as follows:

$$\left[\tilde{\alpha}(\boldsymbol{x})\right]^{\gamma}=\left[\gamma\sum_{j=1}^{n}\underline{\alpha}_j x_j+(1-\gamma)\sum_{j=1}^{n}\alpha_j^L x_j,\ \gamma\sum_{j=1}^{n}\overline{\alpha}_j x_j+(1-\gamma)\sum_{j=1}^{n}\alpha_j^U x_j\right] \qquad (12)$$

Therefore, from the mathematical concept in Subsection 2.2, the crisp possibilistic mean value of $\tilde{\alpha}(\boldsymbol{x})$ is as follows:

$$\begin{aligned}E\left(\tilde{\alpha}(\boldsymbol{x})\right)&=\int_0^1\gamma\left(\gamma\sum_{j=1}^{n}\left(\underline{\alpha}_j+\overline{\alpha}_j-\alpha_j^L-\alpha_j^U\right)x_j+\sum_{j=1}^{n}\left(\alpha_j^L+\alpha_j^U\right)x_j\right)d\gamma\\&=\frac{1}{3}\sum_{j=1}^{n}\left(\underline{\alpha}_j+\overline{\alpha}_j\right)x_j+\frac{1}{6}\sum_{j=1}^{n}\left(\alpha_j^L+\alpha_j^U\right)x_j\end{aligned} \qquad (13)$$

Consequently, problem ( ) is transformed into the following problem from the viewpoint of crisp possibilistic mean value:

$$\begin{aligned}&\text{Maximize}\quad h\\&\text{subject to}\quad\sum_{j=1}^{n}\overline{r}_j x_j+\frac{L^*(h)}{3}\sum_{j=1}^{n}\left(\underline{\alpha}_j+\overline{\alpha}_j+\frac{1}{2}\left(\alpha_j^L+\alpha_j^U\right)\right)x_j\ge g^{-1}(h),\\&\qquad\sum_{j=1}^{n}x_j=1,\quad 0\le x_j\le p_j,\quad j=1,2,\cdots,n\end{aligned} \qquad (14)$$

In this problem, coefficients $\underline{\alpha}_j+\overline{\alpha}_j+\frac{1}{2}\left(\alpha_j^L+\alpha_j^U\right)$ are initially set as fixed values, and so this problem is similar to problem (9), which is one of the standard possibility programming problem. Then, it should be noted here that problem (14) is a nonconvex programming problem and it is not directly solved by the linear programming techniques or convex programming techniques. However, a decision variable h is involved only in first constraint. Therefore, we introduce the following subproblem involving a parameter q :

$$\begin{aligned}&\text{Maximize}\quad\sum_{j=1}^{n}\overline{r}_j x_j+\frac{L^*(q)}{3}\sum_{j=1}^{n}\left(\underline{\alpha}_j+\overline{\alpha}_j+\frac{1}{2}\left(\alpha_j^L+\alpha_j^U\right)\right)x_j\\&\text{subject to}\quad\sum_{j=1}^{n}x_j=1,\quad 0\le x_j\le p_j,\quad j=1,2,\cdots,n\end{aligned} \qquad (15)$$

In the case that parameter q is fixed, problem (15) is degenerated to a linear programming problem. Therefore, in a way similar to the standard possibility maximization model, we construct the following efficient solution method using the hybrid approach of bisection algorithm and linear programming based on the study [6].

Solution algorithm

STEP1: Elicit the membership function of a fuzzy goal for the total future return with respect to the objective function value.

STEP2: Set $q\leftarrow 1$ and solve problem (14). If the optimal objective value $Z(q)$ of problem (14) satisfies $Z(q)>g^{-1}(q)$, then terminate. In this case, the obtained current solution is an optimal solution of main problem.

STEP 3: Set $q\leftarrow 0$ and solve problem (14). If the optimal objective value $Z(q)$ of problem (14) satisfies $Z(q)\le g^{-1}(q)$, then terminate. In this case, there is no feasible solution and it is necessary to reset a fuzzy goal for the probability or the aspiration level f.

STEP 4: Set $U_q\leftarrow 1$ and $L_q\leftarrow 0$.

STEP 5: Set $\gamma\leftarrow\left(U_q+L_q\right)/2$.

STEP 6: Solve problem (14) and calculate the optimal objective value $Z(q)$ of problem (14). If $Z(q)>g^{-1}(q)$, then set $L_q\leftarrow q$ and return to Step 5. If $Z(q)\le g^{-1}(q)$, then set $U_q\leftarrow q$ and return to Step 5. If $Z(q)=g^{-1}(q)$, then terminate the algorithm. In this case, $\boldsymbol{x}^*(q)$ is equal to a global optimal solution of main problem.

4 Numerical example

In order to compare our proposed models with other models for portfolio selection problems, let us consider a numerical example based on the data of securities on the Tokyo Stock Exchange. In this paper, we compare our proposed model (14) in Section 3 with Carlsson et al. model [2] and Vercher et al. model [22]. These models are type-1 fuzzy portfolio models and include the possibilistic mean value and variance. These problems are formulated as the following form:

(Carlsson et al. model)

$$\begin{aligned}&\text{Maximize}\quad\sum_{j=1}^{n}\frac{1}{2}\left[a_j+b_j+\frac{1}{3}\left(\beta_j-\alpha_j\right)\right]x_j\\&\qquad-\frac{0.0123}{4}\left(\sum_{j=1}^{n}\frac{1}{2}\left[b_j-a_j+\frac{1}{3}\left(\alpha_j+\beta_j\right)\right]x_j\right)^2-\frac{0.0123}{72}\left(\sum_{j=1}^{n}\left(\alpha_j+\beta_j\right)x_j\right)^2\\&\text{subject to}\quad\sum_{j=1}^{n}x_j=1,\quad 0\le x_j\le p_j,\quad j=1,2,\cdots,n\end{aligned}$$

ISBN: 978-989-95079-6-8

(Vercher et al. model)

$$\text{Minimize} \quad \sum_{j=1}^{n}\frac{1}{2}\left[b_j - a_j + \frac{1}{3}(\alpha_j + \beta_j)\right]x_j$$

$$\text{subject to} \quad \sum_{j=1}^{n}\frac{1}{2}\left[a_j + b_j + \frac{1}{3}(\beta_j - \alpha_j)\right]x_j \geq \rho,$$

$$\sum_{j=1}^{n} x_j = 1, \quad 0 \leq x_j \leq p_j, \quad j = 1,2,\cdots,n$$

Let us consider ten securities shown in Table 1, whose mean values and standard deviations are based on historical data in the decade between 1995 and 2004. Then, we introduce the asset allocation rate x_j, $(j = 1,2,...,10)$ to each security, and its upper value is assumed to be 0.2.

Table 1: Sample data from Tokyo Stock Exchange

Returns	Sample mean	SD
R1	0.055	0.445
R2	0.046	0.289
R3	0.015	0.306
R4	0.114	0.208
R5	0.043	0.253
R6	0.034	0.269
R7	0.018	0.230
R8	0.171	0.297
R9	0.087	0.388
R10	0.090	0.318

With respect to our proposed type-2 fuzzy portfolio model, parameter $\overline{r}_j$ of the primary *L*-fuzzy number is assumed to be sample mean in Table 1. Then, in two previous models parameters, parameters of trapezoidal fuzzy number a_j, b_j, α_j, β_j, $(j = 1,2,...,10)$ based on study [22] are shown in Table 2 using the historical data in Table 1.

Table 2. Each parameter value based on the historical data

Returns	a_j	b_j	α_j	β_j
R1	-0.123	0.005	0.239	0.868
R2	-0.069	0.069	0.301	0.467
R3	-0.129	0.025	0.200	0.713
R4	0.005	0.177	0.198	0.235
R5	-0.082	0.114	0.217	0.323
R6	-0.052	0.108	0.290	0.382
R7	-0.056	0.067	0.236	0.369
R8	0.060	0.193	0.310	0.456
R9	-0.093	0.130	0.312	0.626
R10	0.009	0.236	0.563	0.264

The secondary membership functions for $\tilde{\alpha}_j$ of type-2 fuzzy numbers are assumed to be triangle fuzzy numbers, whose $\underline{\alpha}_j = \overline{\alpha}_j$ are variances based on the SD in Table1 and the spreads $\underline{\alpha}_j - \alpha_j^L$ and $\alpha_j^U - \overline{\alpha}_j$ become α_j and β_j in Table 2, respectively. Therefore, using these data and introducing the following fuzzy goals for our proposed model;

$$\mu_G(f) = \begin{cases} 1 & (0.15 \leq f) \\ \dfrac{f - 0.12}{0.03} & (0.12 \leq f < 0.15) \\ 0 & (f < 0.12) \end{cases},$$

we set the parameter ρ in the Vercher et al. model as $\rho = 0.06$, and obtain optimal portfolios for three models shown in Table 3.

Subsequently, we consider the case where an investor purchases securities at the end of 2004 according to each portfolio shown in Table 3. Then, the total return of three models at term ends of 2005, 2006 and 2007 become the following values shown in Table 4, respectively.

Table 3. Each optimal solution with respect to three models

	Proposed model	Carlsson model	Vercher model
R1	0.200	0	0
R2	0	0	0
R3	0	0	0
R4	0.200	0.200	0.200
R5	0	0	0.200
R6	0	0.200	0.200
R7	0	0	0.200
R8	0.200	0.200	0.200
R9	0.200	0.200	0.200
R10	0.200	0.200	0.200

Table 4. Total return to the portfolio of each model

Term end	Proposed	Carlsson	Vercher
2005	0.2524	0.2194	0.2852
2006	0.4050	0.3575	0.3257

From the result in Table 4, we find that our proposed model obtains the larger total return than the Carlsson et al. model dealt with the only crisp possibilistic mean value. Furthermore, from the total return at the term end of 2006, in the case that the investment period is longer, our proposed model obtains much larger total return.

5 Conclusion

In this paper, we have proposed a new portfolio selection problem based on the type-2 fuzzy set. Our proposed model has been initially not a well-defined problem due to primary and secondary fuzzy numbers. Therefore, in order to solve analytically in the sense of mathematical programming, we have introduced the possibility maximization model with fuzzy goal for the object and degree of possibility. Furthermore, by using the concept of crisp possibilistic mean value for the secondary fuzzy numbers, the main problem has been equivalently transformed into the parametric linear programming problem. Consequently, we have developed the efficient solution method to combine the bisection algorithm with the standard linear programming approach. Since this proposed model includes various types of investor's subjectivity in the form of type-2 fuzzy number, it is more versatile than previous models. Our proposed model is dealt with various practical situations in the investment such as the case putting the whole policy of investors including a lot of subjectivity by using the possibilistic mean value and the case

ISBN: 978-989-95079-6-8

indicating the sensible decision making to the investor with shaken investment policy due to uncertain but attractive information.

In this paper, as the criterion for the object, we introduced the possibility measure and the crisp possibilistic mean value. On the other hand, there are various types of criterions with respect to fuzzy object and constraints such as necessity and credibility measure, fuzzy interval approach, etc.. Therefore, as the future works, we need to consider more general and versatile type-2 fuzzy portfolio selection problems with such criterions.

Acknowledgment

The first author would like to thank to Japan Society for the Promotion of Science to support this work.

References

[1] C. Carlsson and R. Fullér, "On possibilistic mean value and variance of fuzzy numbers", *Fuzzy Sets and Systems* 122, pp. 315-326, 2001.

[2] C. Carlsson, R. Fullér and P. Majlender, "A possibilistic approach to selecting portfolios with highest utility score", *Fuzzy Sets and Systems* 131, pp. 13-21, 2002.

[3] O. Castillo and P. Melin, *Type-2 Fuzzy Logic: Theory and Applications*, Studies in Fuzziness and Soft Computing 223, Springer, 2008.

[4] E.J. Elton, M.J. Gruber, *Modern Portfolio Theory and Investment Analysis*, Wiley, New York, 1995.

[5] Y. Fang, K.K. Lai and S. Wang, *Fuzzy Portfolio Optimization; Theory and Method*, LNEMS 609, Springer, 2008.

[6] T. Hasuike, H. Katagiri, H. Ishii, Portfolio selection problems with random fuzzy variable returns, *Fuzzy Sets and Systems*, to appear.

[7] X. Hung, "Two new models for portfolio selection with stochastic returns taking fuzzy information", *European Journal of Operational Research* 180, pp. 396-405, 2007.

[8] M. Inuiguchi, J. Ramik, "Possibilistic linear programming: A brief review of fuzzy mathematical programming and a comparison with stochastic programming in portfolio selection problem", *Fuzzy Sets and Systems* 111, pp. 3-28, 2000.

[9] H. Katagiri, T. Hasuike, H.Ishii, I. Nishizaki, "Random Fuzzy Programming Models based on Possibilistic Programming", *Proceedings of the 2008 IEEE International Conference on Systems, Man and Cybernetics*, pp. 1788-1793.

[10] H. Katagiri, H. Ishii and M. Sakawa, "Linear programming problems with random fuzzy variable coefficients", *Proceedings of 5th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty* 1, pp. 55-58, 2002.

[11] H. Katagiri, H. Ishii, M. Sakawa, "On fuzzy random linear knapsack problems", *Central European Journal of Operations Research* 12, pp. 59-70, 2004.

[12] H. Katagiri, M. Sakawa, H. Ishii, "A study on fuzzy random portfolio selection problems using possibility and necessity measures", *Scientiae Mathematicae Japonocae* 65, pp. 361-369, 2005

[13] H. Konno, "Piecewise linear risk functions and portfolio optimization", *Journal of Operations Research Society of Japan* 33, pp. 139-156, 1990.

[14] H. Konno, H. Shirakawa, H. Yamazaki, "A mean-absolute deviation-skewness portfolio optimization model", *Annals of Operations Research* 45, pp. 205-220, 1993.

[15] R T. Leon, V. Liern, E. Vercher, "Validity of infeasible portfolio selection problems: fuzzy approach", *European Journal of Operational Researches* 139, pp. 178-189, 2002.

[16] B. Liu, Theory and Practice of Uncertain Programming, Physica Verlag, 2002.

[17] D.G. Luenberger, *Investment Science*, Oxford Univ. Press, 1997.

[18] H.M. Markowitz, Portfolio Selection: Efficient Diversification of Investments, Wiley, New York, 1959

[19] R.T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk", *Journal of Risk* 2(3), pp. 1-21, 2000.

[20] H. Tanaka and P. Guo, "Portfolio selection based on upper and lower exponential possibility distributions", *European Journal of Operational Researches* 114, pp. 115-126, 1999.

[21] H. Tanaka, P. Guo and I.B. Turksen, "Portfolio selection based on fuzzy probabilities and possibility distributions", *Fuzzy Sets and Systems* 111, pp. 387-397, 2000.

[22] E. Vercher, J.D. Bermúdez and J.V. Segura, "Fuzzy portfolio optimization under downside risk measures", *Fuzzy Sets and Systems* 158, pp. 769-782, 2007.

[23] J. Watada, "Fuzzy portfolio selection and its applications to decision making", *Tatra Mountains Math. Pub*. 13, pp. 219-248, 1997.

[24] A.V. Yazenin, "Fuzzy and Stochastic Programming", *Fuzzy Sets and Systems*, 22, pp. 171-180, 1987.

[25] A.V. Yazenin, "Possibilistic-Probabilistic Models and Methods of Portfolio Optimization", *Studies in Computational Intelligence* 36, Springer-Verlag Berlin Heidelberg, Germany, pp. 241-259, 2007.

ISBN: 978-989-95079-6-8

Finding the Best Flexibility Strategies by Using an Integrated Method of FAHP and QFD

Müjde Erol Genevois Tuncay Gürbüz

Industrial Engineering Department, Galatasaray University

Ciragan Cad. No:36 Ortakoy, Istanbul, Turkey
Email: merol@gsu.edu.tr, tgurbuz@gsu.edu.tr

Abstract—*In this study, Turkish automotive sector needs for flexibility will be determined, and the relationships between flexibility levers and flexibility performance criteria will be evaluated with respect to the sectoral needs. The choice of the appropriate lever combinations will be made for the sector by using the results of this evaluation. The flexibility strategies will be ranked by using a fuzzy extended AHP approach. At the second phase of this study, the QFD tool will be used to find the best flexibility lever portfolio related to automotive sector strategies.*

Keywords— Flexibility Management, Fuzzy AHP, QFD, Turkish Automotive Sector.

1 Introduction

The markets in which manufacturers and service firms compete are increasingly influenced by intense foreign competition, rapid technological change, shorter product life-cycles and customers increasingly unwilling to settle for mass-produced items or services with limited value. The "new breed of customer" [1], who demands greater responsiveness to a dynamic set of requirements, and a new competitive environment, which exposes local companies to competition with the companies around the globe, form a new scenario that has challenged firms in most industries ([2]; [3]; [4]). In this new scenario, flexibility may be one of the most important capabilities needed for firms to achieve competitive advantage ([5]; [6]; [7]).

We defined flexibility as the capacity of adaptation under the double constraint of uncertainty and the urgency. This uncertainty can come from the providers (rupture, problems of transport), from the customers (variation of the request) or from the company itself (breakdowns of the equipment, problems of provisioning). The possible behaviors of the company face to these problems are called levers of flexibilities. After the mass production era of Ford and lean management era of Toyota, these days are witnessing the era of flexibility. Modern organization should keep as much as possible flexibility capability to be used in case of an incident.

This paper will briefly summarize the flexibility concept and flexibility levers as well as the existing situation of Turkish automotive industry by the flexibility point of view. It will introduce a model, based on Analytic Hierarchy Process (AHP) and Quality Function Deployment (QFD) methodology, which determine the best flexibility levers' portfolio to deal with consumers needs for flexibility in term of delivering a 5R product (right place-RP, right time-RT, right quantity-RQ, right product-RPr, right price-RPc). The Turkish automotive industry including the parts and components producers is considered in this study as a case industry. Automotive industry is considered as a complete industry covering all entities of the value chain.

In the second section of the paper, the flexibility management literature will be reviewed and Turkish automotive sector needs for flexibility will be determined. In the third section, the fuzzy extended AHP model will be introduced and the flexibility strategies will be ranked by using this approach. At the fourth phase of this study, the QFD tool will be used to find the best flexibility lever portfolio related to automotive sector strategies. The relationships between flexibility levers and flexibility performance criteria will be evaluated with respect to the sectoral needs. The choice of the appropriate lever combinations will be made for the sector by using the results of this evaluation. The study will be concluded with the presentation of the selected portfolio.

2 Flexibility Management

2.1 Literature Review

Flexibility becomes a strategic resource from the late 70s due to the uncertainty in the environment faced by most of the enterprises. This uncertainty is due to several factors, such as diversification in product range, reduced product life cycle, fragmentation and globalization of markets, sophistication of customers etc. In a broad view flexibility is defined as the capability of adaptation to change. The flexibility is a total capacity, which will be activated as a function of events, which are not controlled by the company. It has a passive vision and defensive usage of the capability. But flexibility may create some offensive strategies as well. By offering more diverse products than the competitors, by renewing the products more frequently or by customizing associated services, flexible pioneers can set the rules of game of the industry by creating uncertainty for their competitors. Therefore flexibility is defined as a strategic asset for the companies not only for adapting to the changes in the environment but also to change the environment in favor of itself.

ISBN: 978-989-95079-6-8

Some authors distinguish between internal and external flexibilities. For instance, Swamidass [8] distinguishes machine-level flexibility from plant-level flexibility. The former being "predominantly technology based" and the latter being derived from a combination of technology, infrastructure, design and engineering capabilities, and the competitive goals and objectives of a firm. Upton [9] defines internal flexibility as the operations strategy and the set of capabilities a firm nurtures to respond to its environment, and external flexibility as capabilities possessed by the firm and used to accommodate sources of variability to which the firm must respond and which are seen as flexible by the market. This external dimension fits the two major strategies proposed by Hyun and Ahn [10] for using flexibility: reactive and proactive. In the same vein, Gerwin [11] also suggests two major strategies for using flexibility: adaptive and redefinition. The adaptive strategy refers to the defensive or reactive use of flexible competencies to accommodate unknown uncertainty, while the redefinition strategy refers to the proactive use of flexible competencies to raise customer expectations, increase uncertainty for rivals and gain competitive edge.

The reasons companies are motivated to be flexible include the need to make design changes quickly, when competitors introduce new models and customers start switching supply sources; to focus on volume flexibility, when large customers reduce inventories and their demand rates become volatile; more flexible product mixes, when importers or domestic competitors start offering multiple quality and price levels; to respond quickly and supply the new products/services, when the customer tastes change quickly.

Especially, four types of flexibilities are important in point of view of the consumer: the volume (VF), the product mix (PMF), the new product (NPF) and the design flexibilities (DF). Volume flexibility can be defined as the ability to operate profitably at different production volumes [9]. Product mix flexibility is adopted to deal with uncertainty about the products that will be demanded by customers at a particular period ([12]; [13]). It is characterized by the ability to produce several products at one manufacturing facility without incurring a major cost penalty. New product flexibility is the ability of a system to add or substitute new products to the product mix. Finally, the design flexibility is the ability to change the design of a product very economically and quickly

The need for flexibility can be expressed as a satisfaction of the demand with delivery on time (I), at the right place (II), in required quantity (III) with the right product (IV) and the right price (V).

It may be noted that each of the above external elements are the causes beyond the control of any company. Further, none of these needs can be satisfied by the mass production strategy. Consequently, the classical methods as the affair with large production runs and economies of scale are a thing of the past, and are being replaced by a new concept - flexibility management.

To satisfy the need for various types of flexibility, Aggorwal [14] define three groups of levers or mechanisms. The first group is referred to as levers for internal flexibility. Each of these levers remains under the control of management at all times. The salient levers in this group are Planning/Scheduling Flexibility (1), Sequencing Flexibility (2), Routing Flexibility (3), Labor Flexibility (4), Machine/Equipment Flexibility (5), Design/Development Flexibility (6), Process/Technology Flexibility (7), Raw Materials Flexibility (8), Transport/Shipping Flexibility (9), Layout Flexibility (10), Expansion Flexibility (11), Financial Resources Flexibility (12).

The second group consists of levers for soft flexibilities. Each of these levers can enhance internal flexibility. In contrast to hard levers, they are never completely under management's control as each requires heightened participation and support of specialists and other employee groups. The levers constituting this group and their respective code numbers are Organizational Structure Flexibility (13), Decision Making Flexibility (14), Job Design Flexibility (15), Employee's Willingness for Change Flexibility (16), Managerial Perception Change Flexibility (17). These flexibility levers cannot be measured on a hard (quantitative) scale, as was the case with the levers of the first group; therefore they are designated soft levers flexibilities.

The third group of levers involves managerial manipulation of intangibles. Each of these levers can enhance one or more of the flexibilities either externally or internally. Again, their effectiveness cannot be measured on a hard scale, so they are called intangible levers (flexibilities). These are Reputation Building Flexibility (18), Flexibility in Enhancing Knowledge and Experience Base (19), Flexibility in Identifying Undiscovered and Unused Talents of Workforce (20), Flexibility in Development of Standing Agreements with Suppliers and Customers (21).

2.2 Changing Landscape in Turkish Automotive Industry

The present-day automotive industry of Turkey was first developed within the broader context of import substitution in the 1960s. High tariffs, quantitative restrictions and local content regulations were used to create a local vehicle industry. Access to the local market could only be achieved through local production. In Turkey, the government succeeded in attracting a number of assemblers (most notably Ford, Fiat and Renault), and by the early 1970s cars with a high degree of local content were being produced. While the assemblers were joint ventures between foreign companies and local groups by the late 1970s, a thriving components industry was created, much of which was locally owned. The industry continues to grow in the 1970s and 1980s mainly by protectionist policies. The rapid expansion of the automotive industry arose from a coming together of two distinct interests. However, the developing impact of the automotive industry in the 1990s is likely to be very different from that seen in the 1960s and 1970s. This is because the industry has globalized and changed its governance structures

Protectionist policies continued throughout the early 1990s. In 1995, by signing the Customs Union agreement with European Union, Turkey abolished high taxes over European origin vehicles. With customs union Turkish automotive industry changed its nature and it results Turkish

ISBN: 978-989-95079-6-8

car manufacturers to produce competitive products both in price and in quality. Another impact of the customs union over the industry is the shift of focus from local market to global markets. As a direct result of globalization, local producers have begun to manufacture vehicles for global markets. Due to global strategies of multinational vehicle producers, manufacturing facilities in Turkey are dedicated to manufacture specific models. This shift in the assemblers had an immediate impact over the suppliers that are faced to the obligations of the globalized environment. It become important in this uncertain environment to be armed with performer flexibility levers in order to deal with the rapid changes in the sector.

3 Proposed Methodology

3.1 Ranking Flexibility Strategies with Fuzzy Extended AHP Approach

In this paper the fuzzy extension of one of the most widely used MCDM methods namely FAHP is used to find the relative importance of the problems alternatives and criteria. The choice of the method used in this stage is arbitrary but the wide range and the ease of use for FAHP was determinant for us to choose it. And with fuzzy logic, we wanted to handle as precisely as possible the unprecised qualitative data collected from the experts.

The FAHP approach is introduced, with the use of Triangular Fuzzy Numbers-TFNs for pairwise comparison scale of FAHP according to the method of Chang's [15] fuzzy extent analysis and the correct normalization formula given later by Wang et al. [16].

In the conventional AHP, the pairwise comparisons for each level with respect to the goal are conducted using a nine-point scale proposed by Saaty [17]. According to Zadeh [18], it is very difficult for conventional quantification to define the complex situations, so the notion of a linguistic variable, whose values are words or sentences, is necessary. To assess the relative importance of the criteria and to evaluate the alternatives with respect to the problems criteria an assumed weighting set has been developed. The decision makers can use this linguistic rating set. The triangular fuzzy conversion scale of the linguistic values in the weighting set is shown in Table 1.

Table1. The triangular fuzzy conversion scale

Linguistic Values	***Triangular Fuzzy Numbers***
Very Low (VL) Very Poor (VP)	$\left(\frac{1}{5},\frac{1}{5},\frac{1}{3}\right)$
Low (L) Poor (P)	$\left(\frac{1}{5},\frac{1}{3},1\right)$
Exactly Equal	$(1,1,1)$
Medium (M) Fair (F)	$\left(\frac{1}{2},\frac{3}{2},\frac{5}{2}\right)$
High (H) Good (G)	$(1,3,5)$
Very High (VH) Very Good (VG)	$(3,5,5)$

Assume that $X=\{x_1,x_2,...,x_n\}$ is an object set, and $U=\{u_1,u_2,...,u_m\}$ is a goal set. According to the method of Chang's [24] fuzzy extent analysis, each object is taken and extent analysis is performed for each goal respectively. Therefore, *m* extent analysis values for each object can be obtained, with the following representation:

$$M_{g_i}^1, M_{g_i}^2,..., M_{g_i}^m \text{ for } i=1,2,...,n \quad (1)$$

Note that all the $M_{g_i}^j$, $j=1,2,\ldots,m$, are TFNs representing the performance of the object x_i with regard to each goal u_j. The steps of Chang's [19] extent analysis can be given as in the following:

- First, by fuzzy arithmetic operations, take the sum of each row of the fuzzy comparison matrix.

$$RS_i=\sum_{j=1}^{n}\tilde{a}_{ij}=\left(\sum_{j=1}^{n}l_{ij},\ \sum_{j=1}^{n}m_{ij},\ \sum_{j=1}^{n}u_{ij}\right),\quad i=1,\ldots,n \quad (2)$$

- Using fuzzy synthetic extent analysis, the value of fuzzy synthetic extent with respect to the i^{th} object x_i, $i=1,2,\ldots,n$ that represents the overall performance of the object across all goals can be determined by the following normalization formula given by Wang et al. [25]:

$$\tilde{S}_i=\frac{RS_i}{\sum_{j=1}^{n}RS_j}=\left(\frac{\sum_{j=1}^{n}l_{ij}}{\sum_{j=1}^{n}l_{ij}+\sum_{k=1,k\neq i}^{n}\sum_{j=1}^{n}u_{kj}},\ \frac{\sum_{j=1}^{n}m_{ij}}{\sum_{k=1}^{n}\sum_{j=1}^{n}m_{ij}},\ \frac{\sum_{j=1}^{n}u_{ij}}{\sum_{j=1}^{n}u_{ij}+\sum_{k=1,k\neq i}^{n}\sum_{j=1}^{n}l_{kj}}\right),\quad i=1,\ldots,n \quad (3)$$

- The degree of possibility of $M_1\geq M_2$ is defined as :

$$V(M_1\geq M_2)=\sup_{x\geq y}\left[\min\left(\mu_{M_1}(x),\mu_{M_2}(y)\right)\right] \quad (4)$$

- Compute the degree of possibility of $\tilde{S}_i\geq\tilde{S}_j$ by the following equation where $\tilde{S}_i=(l_i,m_i,u_i)$ and $\tilde{S}_j=(l_j,m_j,u_j)$:

$$V(\tilde{S}_i\geq\tilde{S}_j)=\begin{cases}1, & \text{if } m_i\geq m_j,\\ \dfrac{u_i-l_j}{(u_i-m_i)+(m_j-l_j)}, & \text{if } l_j\leq u_i,\quad i,j=1,\ldots,n;i\neq j\\ 0, & \text{otherwise,}\end{cases} \quad (5)$$

- To compare $\tilde{S}_i=(l_i,m_i,u_i)$ and $\tilde{S}_j=(l_j,m_j,u_j)$, we need both the values of $V(\tilde{S}_i\geq\tilde{S}_j)$ and $V(\tilde{S}_j\geq\tilde{S}_i)$. The degree of possibility for a convex fuzzy number $\tilde{S}$ to be greater than *k* convex fuzzy numbers $\tilde{S}_i$, *i*=1,…,k can be defined by:

$$V(M\geq M_1,\ldots,M_k)=V[(M\geq M_1)\wedge\ldots\wedge(M\geq M_k)]=\min V(M\geq M_i)$$
$$\text{for } i=1,2,...,k$$

- Assume that for the alternative A_i,

$$d'(A_i)=\min V(S_i\geq S_j),\qquad \text{for } j=1,2,...,n;\ j\neq i \quad (6)$$

- Then the weight vector is given by,

$$W'=(d'(A_1),d'(A_2),...,d'(A_n))^T \quad (7)$$

- Via normalization, the normalized weight vectors are,

$$W=(d(A_1),d(A_2),...,d(A_n))^T \quad (8)$$

ISBN: 978-989-95079-6-8

To illustrate our model, a decision-making group is formed that consists of the 5 experts from the Turkish automotive industry. After a detailed discussion on every criterion, four main criteria of flexibility types have been identified for the hierarchical structure. The AHP hierarchy scheme is constructed and shown in Figure 1.

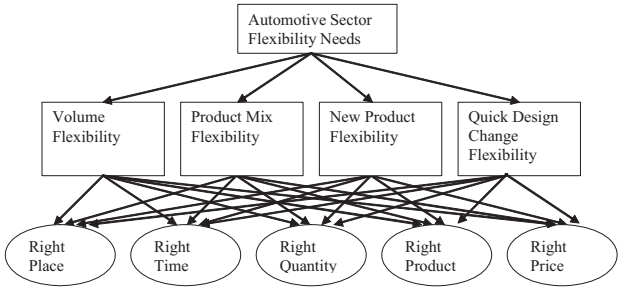


Figure 1. Hierarchical Structure of Flexibility Needs

The aim of the evaluation is to rank the importance of consumers needs for flexibility in the sector. The normalized weight vectors of alternatives with respect to criteria are calculated and shown in Table 2.

Table 2. The normalized weight vectors of the alternatives

	VOLUME	MİX	NEW PRODUCT	QUICK DESIGN CHANGE	
	0,752	0,1	0,045	0,103	
RP	0,108	0,129	0,131	0,083	0,109
RT	0,148	0,144	0,154	0,203	0,154
RQ	0,221	0,094	0,096	0,144	0,195
RPr	0,276	0,228	0,489	0,252	0,278
RPc	0,247	0,404	0,131	0,318	0,265

The need for volume flexibility is highly important in the automotive sector. The consumer demand in this sector is significantly sensible to the global economic indicators. On the other hand, the right price is very important when we are faced to a mix product flexibility problem. In the same way, placing the right product in the market is extremely important when we have to satisfy a new product flexibility need. Finally, the 5R strategies are ranked for their relative importance as follows: $RPr \succ RPc \succ RQ \succ RT \succ RP$.

3.2 Selection of the best flexibility lever portfolio using QFD

Quality function deployment (QFD) is ‘‘an overall concept that provides a means of translating customer requirements into the appropriate technical requirements for each stage of product development and production (i.e., marketing strategies, planning, product design and engineering, prototype evaluation, production process development, production, sales)’’ [19]. QFD belongs to the sphere of quality management methods, offering us a linear and structured guideline for converting the customer’s needs into specifications for, and characteristics of new products and services. The method involves developing four matrixes, or ‘houses’, that we enter by degrees as a project for a given product or production process is developed on increasingly specific levels [20]. In the present article, our attention focuses on the Planning Matrix, or (HOQ) [21] (Fig. 2).

The HOQ provides the specifications for product design (or engineering characteristics) in terms of their relative importance and of target values that have to be reached in design and production. In a sense, the HOQ is the hub of the whole QFD method: its construction enables us to proceed from the customer’s requirements to the design specifications ([22]; [23]). This paper describes the HOQ and its process following the approaches suggested by Brown [24], and Griffin and Hauser [25]. Step 1: Identify the WHATs. The expected benefits in a product or service in the customer’s own words are customer needs and are usually called customer attributes (CA) or ‘‘WHATs’’, area (A) in Fig. 2. In assigning priorities to WHATs, it is necessary to balance efforts in order to accomplish those needs that add value to the customer. The priorities are usually indicated in the area designated as (B) in Fig. 2. Step 2: Determination of HOWs. Engineering characteristics are specified as the ‘‘HOWs’’ of the HOQ and also called measurable requirements. HOWs are identified by a multidisciplinary team [26] and positioned on the area marked as (C) on the matrix diagram, Fig. 2. Step 3: Preparation of the relationship matrix (D). A team judges which WHATs impact which HOWs and to what degree. Step 4: Elaboration of the correlation matrix. The physical relationships among the technical requirements are specified on an array known as ‘‘the roof matrix’’ and identified as (E) in Fig. 2. Step 5: Action plan. The weights of the HOWs, identified as area (F), are placed at the base of the quality matrix. These weights are one of the main outputs of the HOQ, and are determined by

$$\text{Weight (HOW)}_i = V(\text{HOW})_{i1} * \text{imp}(\text{WHAT}_1) + \ldots + V(\text{HOW})_{in} * \text{imp}(\text{WHAT}_n)$$

where $V(\text{HOW})_{in}$ is the correlation value of HOW_i with WHAT_n, and imp (WHAT_n) represents the importance or priority of WHAT_n.

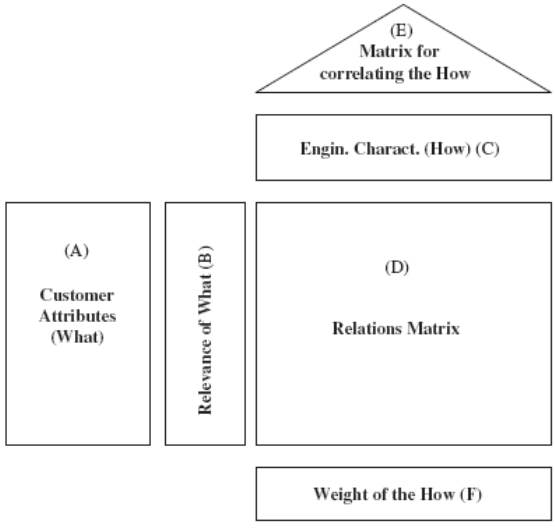


Figure 2. House of quality.

Various quantitative methods have been suggested to use in QFD to improve its reliability and objectiveness, noticeably the methods of management science/operational research (MS/OR), marketing research, and fuzzy logic.

ISBN: 978-989-95079-6-8

			Column Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
			Max Relationship Value in Column	3	3	3	9	3	9	9	3	9	3	9	9	3	3	3	3	3	3	3	3	9
			Requirement Weight	214,47	93,371	98,853	471,84	183,76	276,98	388,36	109,98	312,18	157,28	258,78	296,75	194,62	194,62	155,67	102,97	109,71	155,67	247,58	162,94	467
			Relative Weight	4,61	2,01	2,12	10,14	3,95	5,95	8,35	2,36	6,71	3,38	5,56	6,38	4,18	4,18	3,35	2,21	2,36	3,35	5,32	3,50	10,04
			Difficulty (0=Easy to Accomplish, 10=Extremely Difficult)	4	5	7	6	5	7	9	6	5	8	8	6	6	6	7	5	5	7	6	4	6
			Minimize (▼), Maximize (▲), or Target (x)	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
			Target or Limit Value																					
Row Number	Max Relationship Value in Row	Relative Weight	Flexibility Levers / Flexibility Requirements	Planning/Scheduling Flexibility	Sequencing Flexibility	Routing Flexibility	Labor Flexibility	Machine/Equipment Flexibility	Design/Development Flexibility	Process/Technology Flexibility	Raw Materials Flexibility	Transport/Shipping Flexibility	Layout Flexibility	Expansion Flexibility	Financial Resources Flexibility	Organizational Structure Flexibility	Decision Making Flexibility	Job Design Flexibility	Employee's Willingness for Change Flexibility	Managerial Perception Change Flexibility	Reputation Building Flexibility	Flexibility in Enhancing Knowledge and Experience Base	Flexibility in Identifying Undiscovered and Unused Talents of Workforce	Standing Agreements with Suppliers and Customers
1	9	10,86	Right Place									9				1	1	1		1	1	1		3
2	3	15,35	Right Time	3	3	1	3	1				3	1			1	1	1		1	1	1		3
3	9	19,48	Right Quantity	3	1		9	3		3		3	3	9	3	3	3	1	1		1	3		3
4	9	27,83	Right Product	3	1	3	9	3	9	9	3	3	3	3		3	3	3	3	3	3	3	3	9
5	9	26,48	Right Price	1				1	1	3	1	1			9	1	1	1			1	3	3	3

Figure 3. The HOQ applied to the flexibility levers portfolio choice

Works applying MS/OR methods to QFD include analytic hierarchy process (AHP) to prioritize customer needs (whats) ([27]; [28]; [29]), AHP and benchmarking integrated to rate ‘‘whats’’ [30], AHP and QFD for combining customers’ requirements and preferences [31], AHP and two multi-attribute decision-making methods for rating ‘‘whats’’ [32], AHP, multiattribute utility theory, and linear programming methods for rating ‘‘whats’’ [33], value functions to capture ‘‘whats’’ [34], AHP to study the ranking sensitivity of ‘‘whats’’ in QFD [35].

In this paper, we are using the F-extended AHP method which is described in the previous section for ranking the WHAT column of the HOQ (fig. 3). These are the flexibility levers in the HOW column. The experts have judged which flexibility levers impact which flexibility needs and to what degree. The physical relationships among the technical requirements are also specified but not given here. Finally, the table 3 indicates the importance weights of the flexibility levers.

4 Conclusions

In this paper, a decision making model, based on Analytic Hierarchy Process (AHP) and Quality Function Deployment (QFD) methodology, which determine the best flexibility levers’ portfolio to deal with consumers needs for flexibility in term of delivering a 5R product is developed. Thus, the need for flexibility is calculated with the F-extended AHP method. The experts have judged which flexibility levers impact which flexibility needs and to what degree.

The need for volume flexibility is highly important in the automotive sector. On the other hand, the right price is very important when we are faced to a mix product flexibility problem. In the same way, placing the right product in the market is extremely important when we have to satisfy a new product flexibility need.

Facing continuous change in the automotive industry, using the flexible human resources factors and being able to develop the flexible agreements with suppliers and customers are very important. On the other hand, an automotive firm must have a flexible technology to adapt any change and be highly flexible in his transport and shipping design.

For the future works, the study will develop a Fuzzy-QFD for better represent the vagueness of the experts’ opinions on the flexibility levers. In a next time, the three phases of QFD will be included for a global design strategy.

Table 3. The importance weight of the flexibility levers

Row Number	Flexibility Levers	Relative Weight (Relative Importance)
4	Labor Flexibility	10,14%
21	Flexibility in Development of Standing Agreements with Suppliers and Customers	10,04%
7	Process/Technology Flexibility	8,35%
9	Transport/Shipping Flexibility	6,71%
12	Financial Resources Flexibility	6,38%
6	Design/Development Flexibility	5,95%
11	Expansion Flexibility	5,56%
19	Flexibility in Enhancing Knowledge and Experience Base	5,32%
1	Planning/Scheduling Flexibility	4,61%
13	Organizational Structure Flexibility	4,18%
14	Decision Making Flexibility	4,18%
5	Machine/Equipment Flexibility	3,95%
20	Flexibility in Identifying Undiscovered and Unused Talents of Workforce	3,50%
10	Layout Flexibility	3,38%
15	Job Design Flexibility	3,35%
18	Reputation Building Flexibility	3,35%
8	Raw Materials Flexibility	2,36%
17	Managerial Perception Change Flexibility	2,36%
16	Employee's Willingness for Change Flexibility	2,21%
3	Routing Flexibility	2,12%
2	Sequencing Flexibility	2,01%

ISBN: 978-989-95079-6-8

Acknowledgment

This research has been financially supported by Galatasaray University Research Fund.

References

[1] Handfield, R.B. and Nichols, E.L. *Supply Chain Redesign: Transforming Supply Chains into Integrated Value Systems*, Prentice-Hall, Upper Saddle River, NJ. 2002.

[2] Mehta, S.N., Cisco fractures its own fairy tale, *Fortune*, May, p. 104, 2001.

[3] Monckza, R.M. and Morgan, J., Competitive supply strategies for the 21st century, *Purchasing*, 128(1): 48-59, 2000.

[4] Pagell, M., Understanding the factors that enable and inhibit the integration of operations, purchasing and logistics, *Journal of Operations Management*, 22(5): 459-87, 2004.

[5] Holweg, M., The three dimensions of responsiveness, *International Journal of Operations & Production Management*, 25(7/8): 603-22, 2005.

[6] Storey, J., Emberson, C. and Reade, D., The barriers to customer responsive supply chain management, *International Journal of Operations & Production Management*, 25(3): 242-60, 2005.

[7] Reichhart, A. and Holweg, M., Creating the customer-responsive supply chain: a reconciliation of concepts, *International Journal of Production Planning & Control*, 27(11): 1144-72, 2007.

[8] Swamidass, P., Manufacturing flexibility: operations management association, *Monograph no. 2, National Association of Manufacturers*, Washington, DC., 1988.

[9] Upton, D.M., The management of manufacturing flexibility, *California Management Review*, 36(2): 72-89, 1994.

[10] Hyun, J. and Ahn, B.H., A unifying framework for manufacturing flexibility, *Manufacturing Review*, 5(4): 251-60, 1992.

[11] Gerwin, D., Manufacturing flexibility: a strategic perspective, *Management Science*, 39(4): 395-410, 1993.

[12] Gerwin D., Manufacturing flexibility: a strategic perspective, *Management Science,* 39(4):395–410, 1993.

[13] Bengtsson J, Olhager J., The impact of the product mix on the value of flexibility, *Omega*, 30:265–73, 2002.

[14] Aggorval, S., Flexibility Management: the Ultimate Strategy, *Industrial Management,* Jan.-Fev., 1997.

[15] Chang, D. Y., Applications of the Extent Analysis Method on Fuzzy AHP, *European Journal of Operational Research,* 95: 649–655, 1996.

[16] Wang, Y.-M., Luo, Y., Hua, Z., On the Extent Analysis Method for Fuzzy AHP and its Applications*, European Journal of Operational Research,* 186: 735–747, 2008.

[17] Saaty, T. L., *The Analytic Hierarchy Process*, Mc. Graw-Hill, New York, 1980.

[18] Zadeh, L. A., Fuzzy sets, *Information and Control*, 8: 338-353, 1965.

[19] Sullivan, L.P., Quality function deployment, *Quality Progress*, 19(6): 39–50, 1986.

[20] Akao, Y., *Quality Function Deployment: Integrating Customer Requirements into Product Design*, Productivity Press 1990.

[21] Hauser, J.R., Clausing, D. The house of quality, *Harvard Business Review*, 63–73, 1988.

[22] Schmidt, R. The implementation of simultaneous engineering in the stage of product concept development: A process orientated improvement of quality function deployment, *European Journal of Operational Research* 100(2): 293–314, 1997.

[23] Fariborz, Y.P., Rafael, A.C., Quality function deployment for the good of soccer, *European Journal of Operational Research*, 137: 642–656, 2002.

[24] Brown, P.G., QFD: echoing the voice of the customer, *AT&T Technical Journal*, 18–32, 1991.

[25] Griffin A., Hauser, J., The voice of the customer, Technical report, Working paper 92-106, *Marketing Science Institute*, Cambridge, MA. 1992.

[26] Hauser, J.R., How Puritan7Bennett used the house of quality, *Sloan Management Review*, 34(3): 61–70, 1993.

[27] Armacost, R.T., Componation, P.J., Mullens, M.A. and Swart, W.W., An AHP framework for prioritizing custom requirements in QFD: an industrialized housing application, *IIE Transactions*, 26(4): 72–79, 1994.

[28] Chuang, P.T., Combining the analytic hierarchy process and quality function deployment for a location decision from a requirement perspective, *International Journal of Advanced Manufacturing Technology,* 18(11): 842–849, 2001.

[29] Madu, C.N., Kuei, C.H., Aheto, J., Winokur, D., Integrating total quality management in the adoption of new technologies, *Benchmarking: An International Journal*, 1(3): 52–66, 1994.

[30] Lu, M.H., Madu, C.N., Kuei, C. and Winokur, D., Integrating QFD, AHP and benchmarking in strategic marketing, *Journal of Business & Industrial Marketing*, 9(1): 41–50, 1994.

[31] Lin, B., Fite, D., Managing a sea of quality information at Ark-La-Tex Aquatics, *National Productivity Review* 15(1): 79–85, 1995.

[32] Chan, L.K., Wu, M.L., Prioritizing the technical measures in quality function deployment, *Quality Engineering*, 10(3): 467–479, 1998.

[33] Ho, E.S.S.A., Lai, Y.J. and Chang, S.I. An integrated group decision-making approach to quality function deployment, *IIE Transactions*, 31: 553–567, 1999.

[34] Askin, R.G., Dawson, D.W., Maximizing customer satisfaction by optimal specification of engineering characteristics, *IIE Transactions*, 32(1): 9–20, 2000.

[35] Xie, M., Goh, T.N., Wang, H., A study of the sensitivity of ‘‘customer voice’’ in QFD analysis, *International Journal of Industrial Engineering*, 5(4): 301–307, 1998.

ISBN: 978-989-95079-6-8

Continuous OWA operator and its calculation

Yasuo Narukawa[1] Vicenç Torra[2]

1.Toho Gakuen, 3-1-10, Naka, Kunitachi, Tokyo, 186-0004, Japan
1.The Institute for Research in Business Administration, Waseda University,
1-6-1, Nishi-Waseda, Shinjuku-ku, Tokyo,169-8050, Japan
2.IIIA - CSIC,
Campus UAB s/n, 08193 Bellaterra, Catalonia, Spain
Email: narukawa@d4.dion.ne.jp, vtorra@iiia.csic.es

Abstract*— *In this paper, we study the OWA operator on the real line, which corresponds to the Continuous OWA operator (COWA). After defining it, we introduce some properties and fundamental formulas for its computation. Among them, we give as an example, a differential equation for the COWA operator. **Keywords:** Fuzzy measures, Order weighted averaging operator, Choquet integral, Continuous OWA operator

1 Introduction

Aggregation operators [1, 3, 8] are used to combine information to obtain a datum of better quality. In recent years there is an increasing interest in these topics for their application in decision problems and artificial intelligence.

Among aggregation operators, one of the most well known and useful one is the Ordered Weighted Averaging operator (OWA) introduced by Yager [12, 13, 14]. The OWA is regarded as a Choquet integral with respect to a fuzzy measure[6, 9].

In classical statistics, when the amount of data is large, it is usual to approximate a discrete distribution, such as a binomial distribution, by a continuous distribution. A continuous distribution, such as the normal distribution, is a Lebesgue integral on the real line and is based on lots of results of classical integral theory. At present, there are a very few theoretical results about the Choquet integral on the real line. This paper is the first step for a theory of Choquet integral on the real line.

The structure of this paper is as follows. In Section 2 we review fuzzy measures and the OWA operator, and introduce a few results related to the OWA operator.

In Section 3, we define the OWA operator on the real line, called the Continuous OWA operator (COWA) operator and introduce some fundamental formulas and expressions for their computation.

In Section 4, we show an example of differential equation, which will be a hint to find a weighting function for the COWA operator. The paper finishes with some conclusions.

2 Preliminaries

In this section, we define fuzzy measures, the Choquet integral and the OWA operator, and show their basic properties.

To introduce both a discrete space and a non-discrete space in a unified way, we use the terms in general topology in this section. Let X be a locally compact Hausdorff space and $\mathcal{B}$ be a class of Borel sets, that is, the smallest $\sigma-$algebra which includes the class of all closed sets. We say that $(X, \mathcal{B})$ is a measurable space.

Example 1 *We consider two examples of Hausdorff spaces:*

(1) *The set of all real numbers R is a locally compact Hausdorff space. If $X = R$, $\mathcal{B}$ is the smallest $\sigma-$ algebra which includes the class of all closed intervals.*

(2) *Let $X := \{1, 2, \ldots, N\}$. X is a compact Hausdorff space with a discrete topology. Then we have $\mathcal{B} = 2^X$.*

Definition 1 [7] Let $(X, \mathcal{B})$ be a measurable space. A fuzzy measure (or a non-additive measure) μ is a real valued set function, $\mu : \mathcal{B} \longrightarrow [0, 1]$ with the following properties;

(1) $\mu(\emptyset) = 0$

(2) $\mu(A) \leq \mu(B)$ whenever $A \subset B$, $A, B \in \mathcal{B}$.

We say that the triplet $(X, \mathcal{B}, \mu)$ is a fuzzy measure space if μ is a fuzzy measure.

A fuzzy measure is said to be continuous if $A_n \uparrow A$ implies $\mu(A_n) \uparrow \mu(A)$ and $A_n \downarrow A$ implies $\mu(A_n) \downarrow \mu(A)$.

Definition 2 Let $(X, \mathcal{B})$ be a measurable space. A function $f : X \to R$ is said to be measurable if $\{x | f(x) \geq \alpha\} \in \mathcal{B}$ for all $\alpha \in R$.

Example 2 *Let f be a continuous function. Then, for all $\alpha \in R$, $\{f \geq \alpha\}$ is a closed set. Therefore, f is measurable.*

$\mathcal{F}(X)$ denotes the class of non-negative measurable functions, that is,

$$\mathcal{F}(X) = \{f | f : X \to R^+, f : \text{measurable}\}$$

Definition 3 [2, 5] Let $(X, \mathcal{B}, \mu)$ be a fuzzy measure space. The Choquet integral of $f \in \mathcal{F}(X)$ with respect to μ is defined by

$$(C) \int f d\mu = \int_0^\infty \mu_f(r) dr,$$

where $\mu_f(r) = \mu(\{x | f(x) \geq r\})$.

Let $A \subset X$. The Choquet integral restricted on A is defined by

$$(C) \int_A f d\mu := (C) \int f \cdot 1_A d\mu.$$

ISBN: 978-989-95079-6-8

Definition 4 *Let $D \subset R^N$. An aggregation operator Ag is a function $Ag : D \to R$ with the following properties;*

(1) *(Unanimity or idempotency)*

$Ag(a, \ldots, a) = a$ *if* $(a, \ldots, a) \in D$

(2) *(Monotonicity)*

If $a_i \leq b_i$ *for all* $i = 1, \ldots, n$, $\mathbf{a} = (a_1, \ldots, a_N), \mathbf{b} = (b_1, \ldots, b_N)$ $\mathbf{a}, \mathbf{b} \in D$, *then* $Ag(\mathbf{a}) \leq Ag(\mathbf{b})$.

Yager introduced the Ordered Weighted Averaging operator in [12].

Definition 5 *[12] Given a weighting vector* $\mathbf{w}$ *with weights* $(w_1, \ldots, w_N)$, *the Ordered Weighted Averaging operator is defined as follows:*

$$OWA_{\mathbf{w}}(\mathbf{a}) = \sum_{i=1}^{N} w_i a_{\sigma(i)}$$

where σ *defines a permutation of* $\{1, \ldots, N\}$ *such that* $a_{\sigma(i)} \geq a_{\sigma(i+1)}$, $\mathbf{a} = (a_1, \ldots, a_n)$.

A fuzzy measure μ on $\mathcal{B}$ is said to be symmetric [4] if $\mu(A) = \mu(B)$ for $|A| = |B|$, $A, B \in \mathcal{B}$. Symmetric fuzzy measures on $\{1, \ldots, N\}$ can be represented in terms of N weights w_i for $i = 1, \ldots, N$ so that $\mu(A) = \sum_{i=1}^{|A|} w_i$. Using a symmetric fuzzy measure, we can represent any OWA operator as a Choquet integral.

Proposition 6 *Let* $X := \{1, 2, \ldots, N\}$; *then, for every* $OWA_{\mathbf{w}}$, *there exists a symmetric fuzzy measure satisfying* $\mu(\{1\}) := w_1$ *and* $\mu(\{1, \ldots, i\}) := w_1 + \cdots + w_i$ *for* $i = 1, 2, \ldots, N$, *such that*

$$OWA_{\mathbf{w}}(\mathbf{a}) = (C) \int \mathbf{a} d\mu$$

for $\mathbf{a} \in R_+^N$.

3 Continuous OWA operator

In the following we consider aggregation operators on the real line. Let λ be a Lebesgue measure on $[0, 1]$, that is, $\lambda([a, b]) = b - a$ for $[a, b] \subset [0, 1]$.

Definition 7 *Let* $\mathcal{F}_b([0, 1])$ *be a class of bounded measurable function on* $[0, 1]$. *A continuous aggregation operator Ag on the real line is a functional* $Ag : \mathcal{F}_b([0, 1]) \to R$ *with the following properties;*

(1) *(Unanimity or idempotency)*

$Ag(a) = a$ *if* $a(x) = a$ *for all* $x \in [0, 1]$

(2) *(Monotonicity)*

If $a(x) \leq b(x)$ *for all* $x \in [0, 1]$, *then* $Ag(a) \leq Ag(b)$.

(3) *(Continuity) Let* a_n, $a \in \mathcal{F}_b([0, 1])$ *for* $n = 1, 2, 3, \ldots$ *and* $\lim_{n\to\infty} a_n = a$. *Then,* $\lim_{n\to\infty} Ag(a_n) = Ag(a)$.

Let $D \subset R^N$. For every $a := (a_1, \ldots, a_n) \in D$, we can define a function $f \in \mathcal{F}_b([0, 1])$ by $f(x) := a_k$ if $(k-1)/n \leq x < k/n$. Therefore the definition above is one of the generalization of aggregation operators on D.

Since the Choquet integral with respect to a continuous fuzzy measure satisfies all the conditions above, the Choquet integral with respect to a continuous fuzzy measure is a continuous aggregation operator.

Using the Choquet integral we can define the continuous OWA (COWA) operator.

Definition 8 *Let* μ *be a fuzzy measure on* $([0, 1], \mathcal{B})$. μ *is said to be symmetric, if* $\lambda(A) = \lambda(B)$ *implies* $\mu(A) = \mu(B)$.

Definition 9 *Let* $a \in \mathcal{F}_{\lfloor}([0, 1])$. *The continuous OWA operator is defined by*

$$COWA_{\mu}(\mathbf{a}) = (C) \int \mathbf{a} d\mu$$

where μ *is a symmetric fuzzy measure.*

Let μ be a symmetric fuzzy measure on $([0, 1], \mathcal{B})$. Suppose that $\lambda(A) < \lambda(B)$. Then there exists $B' \in \mathcal{B}$ such that $\lambda(B) = \lambda(B')$ and $A \subset B'$. Then we have

$$\mu(A) < \mu(B') = \mu(B).$$

Therefore we have the next proposition.

Proposition 10 *Let* μ *be a symmetric fuzzy measure on* $([0, 1], \mathcal{B})$. *Then there exists a monotone function* $\varphi : [0, 1] \to [0, 1]$ *such that* $\mu = \varphi \circ \lambda$.

It follows from the proposition above that we can consider the Choquet integral with respect to $\varphi \circ \lambda$ as the COWA operator. We will write $COWA_{\varphi}$ instead of $COWA_{\varphi\circ\lambda}$ and we will say that φ is the weight for the COWA operator.

Let $f : [0, 1] \to R$ be monotone increasing with $f(0) = 0$ and differentiable. We define the sequence of functions $\{f_k\}$ by $f_1 = f$, $f_{k+1} = \int_0^x f_k d\lambda$ for $x \in [0, 1]$, $k = 1, 2, \ldots$. Then we have

$$\begin{aligned}(C) \int_{[0,1]} f d\lambda^n &= \int_0^{\infty} \lambda^n (f \cdot 1_{[0,x] \geq \alpha}) d\alpha \\ &= \int_0^{f(x)} (x - f^{-1}(\alpha))^n d\alpha\end{aligned}$$

Let $t := x - f^{-1}(\alpha)$, Since we have $d\alpha = -f'(x - t)dt$, $t = x$ if $\alpha = 0$ and $t = 0$ if $\alpha = f(x)$, then

$$\begin{aligned}(C) \int_{[0,1]} f d\lambda^n &= \int_x^0 t^n \cdot (-f'(x - t)) dt \\ &= \int_0^x t^n f'(x - t) dt\end{aligned}$$

Next let $s := x - t$, we have

$$\begin{aligned}(C) \int_{[0,1]} f d\lambda^n &= \int_0^x (x - s)^n f'(s) ds \\ &= [(x - s)^n f(s)]_0^x + n \int_0^x (x - s)^{n-1} f_1(s) ds\end{aligned}$$

ISBN: 978-989-95079-6-8

Since $f(x) = 0$, we have

$$(C)\int_{[0,1]} f d\lambda^n = n\int_0^x (x-s)^{n-1} f_1(s) ds.$$

It follows from integration by part again that

$$\int_0^x (x-s)^{n-1} f_1(s) ds = (n-1)\int_0^x (x-s)^{n-2} f_2(s) ds.$$

Repeating above calculation, we have the next lemma.

Lemma 11 *Let $f : [0,1] \to R$ be monotone increasing with $f(0) = 0$ and differentiable. We define the sequence of functions $\{f_k\}$ by $f_1 = f$, $f_{k+1} = \int_0^x f_k d\lambda$ for $x \in [0,1]$, $k = 1, 2, \ldots$. Then we have*

$$(C)\int_{[0,x]} f d\lambda^n = n! f_n(x)$$

for $x \in [0,1]$.

Example 3 *Let $f(t) = t$, we have $f_1 = \frac{1}{2}x^2, \cdots, f_n = \frac{1}{(n+1)!}x^{n+1}$.*

$$(C)\int_{[0,x]} t d\lambda^n(t) = \frac{1}{n+1}x^{n+1}$$

for $x \in [0,1]$.

Let the weight w be ∞-order differentiable. Then we can express w by

$$w(x) := \sum_{k=1}^{\infty} a_k x^k.$$

Since the Choquet integral is linear with respect to the fuzzy measures. We have

$$(C)\int_{[0,x]} f dw \circ \lambda = \sum_{k=1}^{\infty} k! a_k f_k(x)$$

for $x \in [0,1]$. Therefore we have the next theorem.

Theorem 12 *Let $f : [0,1] \to R$ be monotone increasing with $f(0) = 0$ and differentiable. We define the sequence of functions $\{f_k\}$ by $f_1 = f$, $f_{k+1} = \int_0^x f_k d\lambda$ for $x \in [0,1]$, $k = 1, 2, \ldots$.*

$$COWA_{\mathbf{w}}(f) = \sum_{k=1}^{\infty} k! a_k f_k(1)$$

for $x \in [0,1]$.

4 Differential equation for COWA operator

In this section we consider the definition of a weight for the COWA. As we will see below, a differential equation helps in this definition.

Let us define a weight $w(x) := \dfrac{e^x - 1}{e - 1}$. Then we have

$$w(x) = \frac{1}{e-1}\sum_{k=1}^{\infty} \frac{1}{k!} x^k.$$

Since $a_k = \dfrac{1}{(e-1)k!}$, we have

$$(C)\int_{[0,x]} f dw \circ \lambda = \frac{1}{e-1}\sum_{k=1}^{\infty} k! f_k(x)$$

for $x \in [0,1]$.

Then we have the next differential equation for a weighting function.

$$\frac{d}{dx}(C)\int_{[0,x]} f dw \circ \lambda = f(x) + \frac{1}{e-1}\sum_{k=1}^{\infty} k! f_k(x)$$

for $x \in [0,1]$. Therefore we have the next differential equation.

$$\frac{d}{dx}(C)\int_{[0,x]} f dw \circ \lambda = f(x) + (C)\int_{[0,x]} f dw \circ \lambda$$

for $x \in [0,1]$.

Proposition 13 *Let f be a monotone increasing function with $f(0) = 0$ and let f be differentiable. If the weighting function satisfies the next equation:*

$$\frac{d}{dx}(C)\int_{[0,x]} f dw \circ \lambda = f(x) + (C)\int_{[0,x]} f dw \circ \lambda$$

for $x \in [0,1]$, then we have

$$w(x) = \frac{e^x - 1}{e - 1}$$

5 Conclusion

We defined the OWA operator on the real line (COWA operator), and show some fundamental formulas for its calculation. We give an example of differential equation to find a weighting function for COWA operator. This methods will be applicable to the Weighted OWA (WOWA) operator introduced by Torra [10, 11], which is one of the generalization of OWA operator. We expect to have some new results related to the WOWA operator in the near future.

Acknowledgements

Partial support by the Spanish MEC (projects ARES – CONSOLIDER INGENIO 2010 CSD2007-00004 – and eAEGIS – TSI2007-65406-C03-02) is acknowledged.

References

[1] Calvo, T., Mayor, G., Mesiar, R. (eds.) (2002) Aggregation Operators, Physica-Verlag.

[2] Choquet, G. (1955) Theory of capacities. *Ann. Inst. Fourier, Grenoble.* 5,131-295.

[3] Grabisch, M., Murofushi, T., Sugeno, M. (eds.) (2000) Fuzzy Measures and Integrals: Theory and Applications, Physica-Verlag.

[4] Miranda, P., Grabisch, M. (2002) p-symmetric fuzzy measures, Proc. of the IPMU 2002 Conference, 545-552, Annecy, France.

ISBN: 978-989-95079-6-8

[5] Murofushi, T., Sugeno, M. (1989) An interpretation of fuzzy measures and the Choquet integral as an integral with respect to a fuzzy measure, Fuzzy Sets and Systems 29 201-227.

[6] Ralescu, Anca L. , Ralescu, Dan A. (1997) Extensions of fuzzy aggregation, Fuzzy Sets and Systems, 86, 321-330.

[7] Sugeno, M. (1974) *Theory of fuzzy integrals and its applications,* Doctoral Thesis, Tokyo Institute of Technology.

[8] Torra, V., Narukawa, Y. (2007) Modeling decisions: information fusion and aggregation operators, Springer.

[9] Torra, V. Narukawa, Y. (2008) Choquet Stieltjes Integral, Losonczi's Means and OWA Operators. In V. Torra, Y. Narukawa (Eds.): 5th International Conference, MDAI 2008, Sabadell, Spain, October 30-31, 2008. Proceedings. Lecture Notes in Computer Science 5285 Springer, 62-73.

[10] Torra, V. (1996) Weighted OWA operators for synthesis of information, Fifth IEEE International Conference on Fuzzy Systems (IEEE-FUZZ'96) (ISBN 0-7803-3645-3), 966-971, New Orleans, USA.

[11] Torra, V. (1997) The weighted OWA operator, Int. J. of Intel. Syst. 12 153-166.

[12] Yager, R. R. (1988) On ordered weighted averaging aggregation operators in multi-criteria decision making, IEEE Trans. on Systems, Man and Cybernetics, 18 183-190.

[13] Yager, R. R., Filev, D. P. (1994) Parameterized and-like and or-like OWA operators, Int. J. of General Systems 22 297-316.

[14] Yager, R.R. (1993) Families of OWA operators, Fuzzy Sets and Systems, 59 125-148.

ISBN: 978-989-95079-6-8

A Process Algebra Approach to Fuzzy Reasoning

Liliana D'Errico Michele Loreti

Dipartimento di Sistemi e Informatica
Università di Firenze, Italy
Email: {liliana.derrico, michele.loreti}@unifi.it

***Abstract*— *Fuzzy systems address the imprecision of the input and output variables, which formally describe notions like "rather warm" or "pretty cold", while provide a behaviour that depends on fuzzy data. This class of systems are classically represented by means of Fuzzy Inference Systems (FIS), a computing framework based on the concepts of fuzzy if-then rules and fuzzy reasoning. Even if FIS are largely used, these lack in compositionality. Moreover, the analysis of modeled behaviuors needs complex analytic tools. In this paper we propose a process algebraic approach to specification and analysis of fuzzy behaviours. Indeed, we introduce a Fuzzy variant of CCS (*Calculus of Communicating Processes*), that permits compositionally describing fuzzy behaviours. Moreover, we also show how standard process algebra formal tools, like modal logics and behavioural equivalences, can be used for supporting fuzzy reasoning.*

***Keywords*—** Fuzzy Systems, Process Algebras, Compositional Fuzzy Reasoning

1 Introduction

Human perception of real world abounds with concepts without strictly defined constraints (examples are fat, very, more, slowly, old, etc). Such concepts can be described by means of Fuzzy Sets [1, 2]: classes of objects in which transit from membership to not membership gradually takes place.

Fuzzy Sets are widely used in control systems where the system behaviour can depend on data without precise values. In a *Fuzzy System* input values are described by means of variables that model data (temperature, speed) representing them as fuzzy sets, each of which identifies a different range of values (e.g. *cold*, *warm*, *hot*,...).

These systems are classically represented by means of Fuzzy Inference Systems (FIS) [3], a computing framework based on the concepts of fuzzy if-then rules, fuzzy set theory and fuzzy reasoning. A Fuzzy If-Then-Rule is a rule of the form *If* x *is* A *then* y *is* B where A and B are linguistic values defined by fuzzy sets on universes of discourse X and Y, respectively. Fuzzy reasoning is an inference procedure used to derive conclusions from a set of fuzzy If-Then-Rules and one or more conditions.

Defuzzification is the essential "process" to translate the fuzzy result in a crisp result. The most frequently used defuzzification strategy is the centroid of area defined as:

$$z_{coa} = \frac{\int_z \mu_{C'}(z)\, z\, dz}{\int_z \mu_{C'}(z)\, dz}$$

where $\mu_{C'}(z)$ is the aggregated output membership function. Other defuzzification strategies arise for specific applications: maximum membership, mean of maximum, largest of maximum, smallest of maximum, weighted average and so on. Generally speaking, these defuzzification methods are computation intensive Theoretical results are available [4, 5]. Even if Fuzzy Inference Systems are largely used, these lack in compositionality, in the sense that the rules interactions as well as how a rule interferes with the others is not completely clear. Moreover, it is difficult to compare different implementations of a system as well as verifying the properties satisfied by a given specification.

Process algebras are a set of mathematically rigorous languages with well defined semantics that permit modelling behaviour of concurrent and communicating systems. Verification of concurrent systems within the process algebraic approach can be performed by checking that processes enjoy properties described by some temporal logic's formulae.

In this paper we propose a process algebra approach to specification and analysis of fuzzy behaviours. Indeed, we introduce a Fuzzy variant of CCS (*Calculus of Communicating Processes*), that permits compositionally describing fuzzy behaviours. Operational semantics of Fuzzy CCS (FCCS) is described by means of *Fuzzy Labelled Transition Systems* [6] (FLTS). These are an extension of Labelled Transition Systems where fuzziness is used for modeling imprecision in concurrent systems.

Moreover, we also show how standard process algebra formal tools can be used for supporting fuzzy reasoning. Indeed, we define a fuzzy behavioural equivalence and a fuzzy modal logic that can be used for specifying and verifying properties of fuzzy systems.

The rest of the paper is organised as follows. In Section 2 we introduce Process Algebras conceptually. In Section 3 we recall the basic notions related to L-Fuzzy Sets while in Section 4 we present the Fuzzy CCS. In Section 5 we show how Fuzzy Hennessy-Milner Logic can be used for specifying and verifying properties of Fuzzy Systems. Finally, Section 6 concludes the paper.

2 Process Algebras

Process algebras are a set of mathematically rigorous languages with well defined semantics that permits describing and verifying properties of concurrent communicating systems. They can be seen as mathematical models of processes, regarded as agents that act and interact continuously with other similar agents and with their common environment. The agents may be real-word objects (even people), or they may be artefacts, embodied perhaps in computer hardware or software systems.

Process algebras provide a number of constructors for system description and are equipped with an operational semantics that describes systems evolution. Moreover, they often

ISBN: 978-989-95079-6-8

come equipped with observational mechanisms that permit identifying (through behavioural equivalences) those systems that cannot be taken apart by external observations. In some cases, process algebras have also complete axiomatizations, that capture the relevant identifications.

There has been a huge amount of research work on process algebras carried out during the last 25 years that started with the introduction of CCS [7], CSP [8] and ACP [9].

The main ingredients of a specific process algebra are:

- a minimal set of well thought operators capturing the relevant aspect of systems behaviour and the way systems are composed;
- a transition system associated with the algebra via structural *operational semantics* to describe the evolution of all systems that can be built from the operators;
- an equivalence notion that permits abstracting from irrelevant details of systems descriptions.

Verification of concurrent systems within the process algebraic approach is performed either by resorting to behavioural equivalences or by checking that processes enjoy properties described by some temporal logic's formulae. Equivalences are used for proving conformance of a process to specifications, expressed within the same process algebra notation. Verification with logical formulae is implemented by "model checking", an automatic method to prove properties verification.

In the former case two descriptions of a given system, one very detailed and close to the actual concurrent implementation, the other more abstract describing the abstract tree of relevant actions the system has to perform, are provided and tested for equivalence.

In the latter case, concurrent systems are specified as terms of a process description language while properties are specified as temporal logic formulae. Labelled Transition Systems are associated with terms via a set of structural operational semantics rules and model checking is used to determine whether the transition system associated with those terms enjoys the property specified by the given formulae.

Process algebras and modal logics have been largely used as tools for specifying and verifying properties of concurrent systems. This, also thanks to model checking algorithms, permits verifying whether a given specification satisfies the expected properties.

3 L-Fuzzy Sets

Human perception of real world abounds with concepts without strictly defined constraints (*fat*, *very*, *more*, *slowly*, *old*, etc...). Such concepts can be described by means of Fuzzy Sets: classes of objects in which transit from membership to not membership gradually takes place. A fuzzy set is a simple and intuitive generalization of the classical *crisp*[1] one.

Fuzzy sets are denoted by means of a generalised *membership function* that gives the *membership degree* of each element of the *universe*. This degree usually takes values in $[0, 1]$, the interval of real numbers from 0 to 1 inclusive. Although above range is the one most commonly used for representing membership degrees, any arbitrary set with some natural full or partial ordering can be used. Elements of this set are not required to be numbers as long as the ordering among them can be interpreted as representing various strengths of membership degree [10].

Let U be a universal set and L be a *complete lattice*, a *L-Fuzzy Set* [2, 11] A is denoted by a membership function $\mu_A : U \to L$.

Standard operations on sets and lattices L, like complement, intersection and union, can be generalised to L-Fuzzy Sets. These operations rely on the use of three function $c(\cdot)$, $i(\cdot,\cdot)$ and $u(\cdot,\cdot)$ that, respectively, give the measure of complement, intersection and union of fuzzy degrees.

The complement of a L-Fuzzy Set A, denoted by $\overline{A}$, is specified by a function $c : L \to L$ which assigns a value $\mu_{\overline{A}}(x) = c(\mu_A(x))$ to each membership degree $\mu_A(x)$. This assigned value is interpreted as the membership degree of the element x in the L-Fuzzy Set representing the negation of the concept represented by A.

Intersection and union of two L-Fuzzy Sets A and B are defined using functions $i : L \times L \to L$ and $u : L \times L \to L$. For each element a in the universal set U, these functions take as argument the pair consisting of the membership degrees of a in A and in B, respectively. Function i, also named *t-norm*, yields the membership degree of a in $A \cap B$, while u, also named *t-conorm*, returns the membership degree of a in $A \cup B$. Thus, $\mu_{A\cap B}(a) = i(\mu_A(a), \mu_B(a))$ while $\mu_{A\cup B}(a) = u(\mu_A(a), \mu_B(a))$.

Function c, i and u operating on L-Fuzzy Sets must be *continuous* on L and satisfy all axioms in Table 1, where 0 and 1 denote respectively the least and the greatest element in L while $\leq$ denotes the partial ordering on L. In the rest of this paper we will use $\mathcal{L}$ to denote a tuple $\langle L, c, i, u\rangle$ containing a complete lattice L together with its complement, intersection and union functions used for defining a family of L-Fuzzy Sets.

Table 1: Axioms of fuzzy operations

Axioms for $c(\cdot)$

$$c(1) = 0 \qquad c(0) = 1 \qquad \frac{x \leq y}{c(y) \leq c(x)} \qquad c(c(x)) = x$$

Axioms for $i(\cdot,\cdot)$

$$i(1,1) = 1 \qquad i(0,x) = 0 \qquad i(1,x) = x \qquad i(x,y) = i(y,x)$$

$$\frac{x_1 \leq x_2 \quad y_1 \leq y_2}{i(x_1,y_1) \leq i(x_2,y_2)} \qquad i(i(x,y),z) = i(x,i(y,z))$$

Axioms for $u(\cdot,\cdot)$

$$u(0,0) = 0 \qquad u(0,x) = x \qquad u(1,x) = 1 \qquad u(x,y) = u(y,x)$$

$$\frac{x_1 \leq x_2 \quad y_1 \leq y_2}{u(x_1,y_1) \leq u(x_2,y_2)} \qquad u(u(x,y),z) = u(x,u(y,z))$$

Fuzzy set theory was initially formulated by considering the

[1] Crisp set is defined to split individuals belonging to a certain universe into two groups: members (who surely belong to the set) and not members (who surely do not belong).

ISBN: 978-989-95079-6-8

complete lattice $[0,1]$, denoting the interval of real numbers from 0 to 1 (inclusive), and the following complement, intersection and union functions:

$$c(x) = 1 - x \quad i(x,y) = \min[x,y] \quad u(x,y) = \max[x,y]$$

It is easy to prove that these functions are continuous on $[0,1]$ and satisfy the axioms of Table 1.

4 A Fuzzy Process Algebra

Fuzzy systems are used for handling behaviours that depend on data without precise values that are described by means of variables taking values between 0 and 1. Variables, like temperature or speed, are represented by means of fuzzy sets each of which identifies a different range of values (e.g. *cold*, *warm*, *hot*,...).

As an example, we can consider a *Temperature Control System* (TCS). This is composed of an air conditioner (AC) system and a temperature sensor installed in the room. The TCS regulates the AC power, according to the values read from the sensor, to guarantee a suitable room temperature.

Following the fuzzy approach, room temperature can be modeled by considering different states identifying different range of values. For instance, *cold*, *warm* and *hot* like in figure 1.

Behaviour of TCS is described by a set of rules like: *if the temperature is warm then slightly increase the AC power*, *if the temperature is hot then increase the AC power*.

These systems are classically represented by means of Fuzzy Inference Systems (FIS).

In this section we present a fuzzy variant of CCS (*Calculus of Communicating Processes* [12]), named *Fuzzy CCS* (FCCS). In the new calculus the standard CCS actions are enriched with a fuzzy value modeling the enabling-degree. We aim at defining a formal language that permits compositionally describing fuzzy systems and that can be used for supporting fuzzy reasoning. This, also thanks to the use of standard formal tools like modal logics.

Like other process algebrae, FCCS provides a set of operators that permit describing the complete system starting from the specification of its subcomponents. Following the CCS approach, components interact with each other by means of *actions*, atomic and not interruptible steps, which represent input/output operations on communication ports or internal computations of the system. Let Λ be an infinite numerable set of labels or ports, not containing τ. A FCCS action can be: $a \in \Lambda$, the action to receive a signal on port a; $\bar{a}$ with $a \in \Lambda$, the action to deliver a signal on port a; τ, an internal computation step. We assume $\bar{\bar{\alpha}} \triangleq \alpha$, where $\alpha \in \Lambda \cup \{\bar{a} \mid a \in \Lambda\} \cup \{\tau\}$. Actions $\bar{\alpha}$ and α are said complementary, they represent input and output actions on the same channel.

The fundamental difference from CCS is the introduction of an attribute "x_i", that we define *action execution degree*. It is a fuzzy value able to represent, on a quality level, action behaviour. By this way CCS actions become fuzzy and therefore more representative of real world.

We define FCCS sintax by means of the following grammar:

$$\begin{aligned} Q &::= nil \mid X \mid \sum_{i \in I} (act_i,\, x_i).Q_i \\ P &::= Q \mid P_1 \mid P_2 \mid P\backslash A \mid P\,[f] \\ act &::= \bar{a} \mid a \mid \tau \end{aligned}$$

Now a brief description of operators:

- nil is the *inactive* process.
- X is the *process constant*, if $X \triangleq P$ then X denotes the invocation of process P. It is useful in defining recursive processes.
- $\sum_{i \in I}(act_i,\, x_i).Q_i$ is the *choice or sum* operator and denotes a choice among i possible behaviours that evolve with action act_i and degree x_i.
- $P_1 \mid P_2$ is the *parallel composition* operator and represents the concurrent execution of processes P_1 e P_2. If during the composition two complementary actions match, the resulting composed action is the internal one τ.
- $P\backslash A$ is the *restriction* operator. $A \in \Lambda$ and $P\backslash A$ behaves like P exception made for the impossibility to interact using actions in A.
- $P\,[f]$ is the *relabelling* operator. $f : \Lambda \to \Lambda$ allows relabelling process actions, to ease the description of complex processes.

4.1 Fuzzy Operational Semantics

Operational semantics of FCCS processes is defined in term of Fuzzy Labelled Transition Systems [6]. These generalize Labelled Transition Systems by defining the transition relation in term of a L-Fuzzy Set. This approach permits modeling situations like: "the transition takes place *rarely*" or "the transition occurs *frequently*" which may be distinguished and treated as a consequence.

Definition 4.1 ($\mathcal{L}$-FLTS) *Let $\mathcal{L} = \langle L, c, i, u\rangle$, a Fuzzy Labelled Transition System $\mathcal{F}$ for $\mathcal{L}$ ($\mathcal{L}$-FLTS) is a tuple $\langle Q, A, \chi_{\to}\rangle$ where:*

- *Q is a set whose elements are called states*
- *A is a finite set whose elements represent actions*
- *$\chi_{\to} : (Q \times A \times Q) \to L$ is the total membership function.*

We will write $q_0 \xrightarrow{\alpha}_{\varepsilon} q_1$ to denote that a transition from state q_0 to state q_1 by action α has a membership degree ε to the automaton.

In FLTS next states are selected nondeterministically. The membership degree associated to each transition is used to give a measure to computations. This measure is not *exact* as the probability one induced by PLTS [13], but can be used as the base for *approximate* reasoning in the spirit of Fuzzy Logic. The membership degree associated to a transition can be thought of as a value describing how much this transition is *enabled* in the FLTS.

Semantics for FCCS processes is defined by considering function $\mathcal{N}$ associating to each process P and transition α the fuzzy set $\mathcal{P}$ of processes reachable from P with α; we identify $\mathcal{P}$ with the set

$$\{Q : \varepsilon \mid Q \text{ is reachable from } P \text{ with action } \alpha \text{ and degree } \varepsilon\}$$

ISBN: 978-989-95079-6-8

Function $\mathcal{N}$ is formally defined in Table 2 where we use $\mathcal{P}|Q$ (resp. $Q|\mathcal{P}$) denotes the fuzzy set obtained by composing each element of $\mathcal{P}$ in parallel with Q. Similarly, $\mathcal{P}|\mathcal{Q}$ is the fuzzy set containing the parallel composition of each P in $\mathcal{P}$ with each Q in $\mathcal{Q}$, where the membership degree of $P|Q$ in $\mathcal{P}|\mathcal{Q}$ is defined as $i\left(\mu_{\mathcal{P}}(P), \mu_{\mathcal{Q}}(Q)\right)$.

Table 2: $\mathcal{N}$ext Function

$$\mathcal{N}\left(\langle\alpha,\varepsilon\rangle.P,\beta\right) = \begin{cases} \{P:\varepsilon\} & if\ \alpha=\beta \\ \emptyset & else \end{cases}$$

$$\mathcal{N}\left(P+Q,\alpha\right) = \mathcal{N}\left(P,\alpha\right) \cup \mathcal{N}\left(Q,\alpha\right)$$

$$\mathcal{N}\left(P\backslash L,\alpha\right) = \begin{cases} \mathcal{N}\left(P,\alpha\right)\backslash L & if \alpha \notin L \\ \emptyset & else \end{cases}$$

$$\mathcal{N}\left(P\left[f\right],\alpha\right) = \bigcup_{\beta:f(\beta)=\alpha} \mathcal{N}\left(P,\beta\right)\left[f\right]$$

$$\mathcal{N}\left(P|Q,\alpha\right) = \begin{cases} \left[\mathcal{N}\left(P,\alpha\right)|Q\right] \cup \left[P|\mathcal{N}\left(Q,\alpha\right)\right] & if \alpha \neq \tau \\ \left[\mathcal{N}\left(P,\tau\right)|Q\right] \cup \left[P|\mathcal{N}\left(Q,\tau\right)\right] \cup \left[\bigcup_{\alpha\in\Lambda}\left(\mathcal{N}\left(P,\alpha\right)|\mathcal{N}\left(Q,\bar{\alpha}\right)\right)\right] \cup \left[\bigcup_{\alpha\in\Lambda}\left(\mathcal{N}\left(P,\bar{\alpha}\right)|\mathcal{N}\left(Q,\alpha\right)\right)\right] & if \alpha = \tau \end{cases}$$

$$\mathcal{N}\left(A,\alpha\right) = \mathcal{N}\left(P,\alpha\right) \qquad if\ A \triangleq P$$

We say that P can evolve to Q (written $P \succ Q$) if and only if there exists an action α such that if $\mathcal{P} = \mathcal{N}(P,\alpha)$, $\mathcal{P}(Q) \neq 0$. We use $\succ^*$ to denote the reflexive and transitive closure of $\succ$.

Let P be a FCCS process, we denote with $FLTS(P) = \langle S, \Lambda, \chi_{\rightarrow}\rangle$ the FLTS such that:

- $S = \{Q \mid P \succ^* Q\}$;
- $\chi_{\rightarrow}(P,\alpha,Q) = \mathcal{P}(Q)$, where $\mathcal{P} = \mathcal{N}(P,\alpha)$.

Standard behavioural equivalences, like for instance bisimulation, can be easily generalized in order to take into account fuzziness. This kind of equivalences are useful when one aims at comparing different specifications. Definition of Fuzzy Bisimulation is straightforward and it is somehow reminiscent of Stochastic Bisimulation [14].

Definition 4.2 (Fuzzy Bisimulation) *Let $\langle S, A, \chi_{\rightarrow}\rangle$ a Fuzzy LTS. An equivalence relation $R \subseteq S \times S$ is a* fuzzy bisimulation *if and only if for each p and q in S, for each equivalent class C of R in S, and for each transition label α:*

$$\chi_{\rightarrow}(p,\alpha,C) = \chi_{\rightarrow}(q,\alpha,C)$$

where:

$$\chi_{\rightarrow}(p,\alpha,C) = \bigvee_{p'\in C} \chi_{\rightarrow}(p,\alpha,p')$$

Definition 4.3 (Fuzzy Bisimilarity) *Let $\langle S, A, \chi_{\rightarrow}\rangle$ be a Fuzzy LTS. We say that $p, q \in Q$ are bisimilar ($p \sim_F q$), if there exists a fuzzy bisimulation R such that pRq.*

Relation $\sim_F$ is a *fuzzy bisimilarity* and can be defined as the largest fuzzy bisimulation, namely:

$$\sim_F \triangleq \bigcup \{R \mid R \quad \text{is a fuzzy bisimulation}\}$$

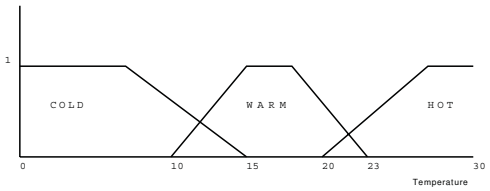


Figure 1: Temperature Fuzzy sets: COLD, NORMAL, HOT

Modeling Fuzzy Systems with FCCS We can now use FCCS for modeling the TCS described in section 4, where we consider three fuzzy sets for describing possible values of the room temperature. These sets are represented in Figure 1. Systems TCS is modeled in FCCS by splitting it into four subcomponents representing behaviours of the system. They "mime" TCS by interacting among each of them. Interactions are the result of synchronization between an input and an output action with the same name. The fuzzy value of the resulting action is calculated following semantics in Tab. 2. Process SYS_{st} is defined as follows:

$$SYS_{st} \triangleq (SENS_{st}|H|N|C)\backslash\{hot, warm, cold, inc, dec, noop\}$$

where st is the starting temperature while process $SENS_t$, which models the behaviour of temperature sensor, is defined as follows:

$$\begin{aligned} SENS_t \triangleq\ & \left\langle \overline{hot}, \mu_{hot}(t)\right\rangle.AC_t \\ & + \left\langle \overline{warm}, \mu_{warm}(t)\right\rangle.AC_t \\ & + \left\langle \overline{cold}, \mu_{cold}(t)\right\rangle.AC_t \\ & + \left\langle temp_t, 1\right\rangle.SENS_t \end{aligned}$$

$$\begin{aligned} AC_t \triangleq\ & \left\langle inc, 1\right\rangle.SENS_{t+1} \\ & + \left\langle dec, 1\right\rangle.SENS_{t-1} \\ & + \left\langle noop, 1\right\rangle.SENS_t \end{aligned}$$

Notice that in the processes above, non-determism is used for modeling the unpredictable changes in room temperature.

Controller behaviour is rendered by means of processes H, N and C that modify the AC power. These processes are defined as follows:

$$\begin{aligned} H &\triangleq \left\langle hot, 1\right\rangle.\left\langle \overline{inc}, 1\right\rangle.H \\ N &\triangleq \left\langle warm, 1\right\rangle.\left\langle \overline{noop}, 1\right\rangle.N \\ C &\triangleq \left\langle cold, 1\right\rangle.\left\langle \overline{dec}, 1\right\rangle.C \end{aligned}$$

For this system one could be interested on verifying that for each fixed starting temperature, the system always is able to reach a state where the room temperature is in a given range. In the next section, we will introduce a modal logic that will permit specifying and verifying properties of fuzzy systems.

An alternative description of this system could be obtained by considering process SYS^2_{st} defined as follows:

$$SYS^2_{st} \triangleq (SENS_{st}|CONT)\backslash\{hot, warm, cold\ inc, dec, noop\}$$

ISBN: 978-989-95079-6-8

where, differently from the previous implementation, the controller is obtained as a single process $CONT$ that nondeterministically can behave like H, N and C defined above:

$$CONT \triangleq (H + N + C).CONT$$

It easy to prove that SYS_{st} and SYS^2_{st} provide the same behaviour. Namely, that $SYS_{st} \sim_F SYS^2_{st}$.

5 Fuzzy Hennessy-Milner Logic

Fuzzy Hennessy-Milner Logic (FHML) [6] is an extension of HML, which aims at specifying properties of concurrent systems whose behavior is detailed by means of FLTS. Let $\mathcal{L} = \langle L, c, i, u\rangle$, $\Phi_{\mathcal{L}}$ be the set of formulas φ defined by the following syntax:

$$\varphi ::= tt \mid \neg\varphi \mid \varphi \bowtie \varepsilon \mid \varphi_1 \wedge \varphi_2 \mid \langle\alpha\rangle\varphi \mid X \mid \nu X.\varphi$$

where $\varepsilon \in L$.

Table 3: Formulae semantics

$$\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![tt]\!]\delta(p) = 1$$
$$\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\neg\varphi]\!]\delta(p) = \mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![\varphi]\!]\delta(p)$$
$$\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\varphi_1 \wedge \varphi_2]\!]\delta(p) = i\left(\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\varphi_1]\!]\delta(p), \mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\varphi_2]\!]\delta(p)\right)$$
$$\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\varphi \bowtie \varepsilon]\!]\delta(p) = \begin{cases} 1 & if\ \mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\varphi]\!]\delta(p) \bowtie \varepsilon \\ 0 & else \end{cases}$$
$$\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\langle\alpha\rangle\varphi]\!]\delta(p) = u_{q\in Q}\left(i\left(\chi_{\rightarrow}(p,\alpha,q), \mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\varphi]\!]\delta(q)\right)\right)$$
$$\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![X]\!]\delta = \delta(X)$$
$$\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\nu X.\varphi]\!]\delta = \cup\left\{\chi \mid \chi \leq \mathfrak{F}^{\delta,\varphi}_X(\chi)\right\}$$
$$\mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![tt]\!]\delta(p) = 0$$
$$\mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![\neg\varphi]\!]\delta(p) = \mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\varphi]\!]\delta(p)$$
$$\mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![\varphi_1 \wedge \varphi_2]\!]\delta(p) = u\left(\mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![\varphi_1]\!]\delta(p), \mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![\varphi_2]\!]\delta(p)\right)$$
$$\mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![\varphi \bowtie \varepsilon]\!]\delta(p) = \begin{cases} 1 & if\ \mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![\varphi]\!]\delta(p) \not\bowtie \varepsilon \\ 0 & else \end{cases}$$
$$\mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![\langle\alpha\rangle\varphi]\!]\delta(p) = $$
$$= i_{q\in Q}\left(u\left(i\left(\chi_{\rightarrow}(p,\alpha,q), \mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![\varphi]\!]\delta(q)\right), c\left(\chi_{\rightarrow}(p,\alpha,q)\right)\right)\right)$$
$$\mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![X]\!]\delta = \delta(X)$$
$$\mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![\nu X.\varphi]\!]\delta = \cap\left\{\chi \mid \chi \geq^{\frown} \mathfrak{F}^{\delta,\varphi}_X(\chi)\right\}$$

where

$$\mathfrak{F}^{\delta,\varphi}_X = \mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\varphi]\!]\delta[\chi/X] \qquad {}^{\frown}\mathfrak{F}^{\delta,\varphi}_X = \mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![\varphi]\!]\delta[\chi/X]$$

Operators are usual logical tt (true), $\neg$ (not), $\wedge$ (and). Moreover $\nu X.\varphi$ is Tarski's fixed point and $\langle\alpha\rangle\varphi$ the modal operator representing the property of evolving with action α to a state that behaves like φ. FHML extends HML by considering new operator $\varphi \bowtie \varepsilon$ ($\bowtie \in \{<, >\}$) that states about the satisfaction degree of a given formula in a given state. Such operator makes possible to describe the satisfaction degree of a formula in terms of an upper or a lower bound[2].

[2]This is somehow reminiscent of operator $[\varphi]_p$ proposed by Parma and Segala [15]

Other operators can be defined as macros in FHML. In the rest of this paper we let: $ff = \neg tt$, $\varphi_1 \vee \varphi_2 = \neg(\neg\varphi_1 \wedge \neg\varphi_2)$, $[\alpha]\varphi = \neg\langle\alpha\rangle\neg\varphi$ and $\mu X.\varphi = \neg\nu X\neg\varphi[\neg X/X]$.

Semantics of FHML is defined in term of functions $\mathcal{M}_{\mathcal{L},\mathcal{F}}$ and $\mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}$ that for each formula φ yield the L-Fuzzy Set that gives the measure, respectively, of satisfaction and unsatisfaction of φ. This approach permits defining semantics of FHML in a general way without considering any special constraint on the underlying L-Fuzzy Sets. Indeed, in general, standard properties of sets do not hold in the case of L-Fuzzy Sets. For instance, the intersection between a L-fuzzy set and its complement could be not empty. Notice that, in general, $\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\neg\varphi]\!] \neq c\left(\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\varphi]\!]\right)$.

To give a semantics to recursive formulae, interpretation functions $\mathcal{M}_{\mathcal{L},\mathcal{F}}$ and $\mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}$ take also, as a parameter, a function $\delta : Q \rightarrow L$, that associates to each logical variable X a L-fuzzy set.

We assume the usual definitions on positive and negative variables in [6] for guaranteeing well-definedness of interpretation formulae.

Definition 5.1 (Well formed formula) *A formula is well formed if in each subformula of the form $\nu X.\varphi$, variable X occurs positive in φ.*

Interpretation functions are parameterized with respect to $\mathcal{L}$, used for defining the underlying L-Fuzzy Set and the relative operations, and with respect to the $\mathcal{L}$-FLTS $\mathcal{F}$ used for interpreting formulae. Note that, interpretation of FHML coincides with the standard interpretation of HML when considering standard Boolean lattices.

Definition 5.2 (Formulae Semantics) *Let $\mathcal{L} = \langle L, c, i, u\rangle$ and $\mathcal{F}_{\mathcal{L}} = \langle Q, A, \chi_{\rightarrow}\rangle$. Functions $\mathcal{M}_{\mathcal{L},\mathcal{F}} : \Phi_{\mathcal{L}} \rightarrow Q \rightarrow L$ and $\mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}} : \Phi_{\mathcal{L}} \rightarrow Q \rightarrow L$ are inductively defined in Table 3.*

$\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\varphi]\!]\delta(q)$ denotes the satisfaction degree of formula φ by q. Formulae tt and ff are satisfied by every state with degree 1 and 0 respectively. A state p satisfies $\neg\varphi$ with degree ε if and only if p *does not satisfy* φ with degree ε. Fuzzy set $\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\varphi_1 \wedge \varphi_2]\!]$ is defined as the intersection between $\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\varphi_1]\!]$ and $\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\varphi_2]\!]$. If a state p satisfies φ with a degree that is $<$ (resp. $>$) of ε then p satisfies $\varphi < \varepsilon$ (resp. $\varphi > \varepsilon$) with degree 1 (resp. 0). $\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\langle\alpha\rangle\varphi]\!]\delta(p)$, which gives a measure of how p can reach with a transition labelled α a state satisfying φ, is defined as the disjunction, for each q, of $i(\chi_{\rightarrow}(p,\alpha,q), \mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\varphi]\!](q))$. Where $\chi_{\rightarrow}(p,\alpha,q)$ returns the membership degree of the transition from p to q with label α to the behaviour of the system. Finally, interpretation of $\nu X.\varphi$ is defined as greates fixed point of the interpretation of φ ($\mathfrak{F}^{\delta,\varphi}_X$), which is monotone in the complete lattice of Q L-Fuzzy Sets. This, thanks to the Tarski's fixed point theorem [16], guarantees the well-definedness of functions $\mathcal{M}_{\mathcal{L},\mathcal{F}}$ and $\mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}$.

The definition of function $\mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}$ is similar and straightforward. However, more attention has to be paid for $\mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![\langle\alpha\rangle\varphi]\!]$. Each state q, contributes to the unsatisfaction of $\langle\alpha\rangle\varphi$ by p as a factor that depends on the degree of the transition α from p to q and on the unsatisfaction degree of formula φ by q. This value is obtained as a disjunction of: $c\left(\chi_{\rightarrow}(p,\alpha,q)\right)$ and $i\left(\chi_{\rightarrow}(p,\alpha,q), \mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![\varphi]\!](q)\right)$. The former indicates how much the transition from p to q with label

ISBN: 978-989-95079-6-8

α *does not belong* to the behaviour of the system. The latter, gives the measure of the unsatisfaction of φ when the action is executed.

Example 5.1 *FHML can be used for specifying that system process SYS_{st} can always reach a configuration where the room temperature is between* 18 *and* 20 *degrees. The following formula states that a configuration where temperature is between* 18 *and* 20 *degrees is eventually reached:*

$$\varphi = \mu X.\langle\overline{temp}_{18}\rangle tt \vee \langle\overline{temp}_{19}\rangle tt \vee \langle\overline{temp}_{20}\rangle tt \vee \langle\tau\rangle X$$

While formula:

$$\nu Y.\varphi \wedge [\tau]Y$$

states that φ is always satisfied.

Logical characterization of Fuzzy Bisimulation Formulae satisfaction induces an equivalence on the interpretation model and two states in a Fuzzy LTS are equivalent if (and only if) they satisfy the same set of formulae. A classical result relating modal logic and behavioural equivalence is the one in [12] showing that the equivalence induced by HML coincides with the bisimulation equivalence. A similar result can be proved when one considers FHML and Fuzzy Bisimulation. However, to prove this correspondence, one has to guarantee that the considered Fuzzy LTS is *finite-branching*.

Definition 5.3 (Finite-branching) *A Fuzzy LTS $\mathcal{F} = \langle S, A, \chi_{\rightarrow}\rangle$ is finite-branching if and only if $\forall p \in S$, $\forall \alpha \in A$ and $\forall \varepsilon \in \chi_{\rightarrow}$, the set $\left\{q' \in Q | q \xrightarrow{\alpha}_{\varepsilon} q'\right\}$ is finite.*

Theorem 1 *Let $\mathcal{F} = \langle S, A, \chi_{\rightarrow}\rangle$ be finite-branching. For each $p, q \in S$,*

$$p \sim_F q \quad \Leftrightarrow \quad \forall \varphi.\, \mathcal{M}[\![\varphi]\!](p) = \varepsilon \Leftrightarrow \mathcal{M}[\![\varphi]\!](q) = \varepsilon \text{ and } \mathcal{M}^{\smile}[\![\varphi]\!](p) = \varepsilon \Leftrightarrow \mathcal{M}^{\smile}[\![\varphi]\!](p) = \varepsilon$$

Due to lack of space, the proof is not reported.

6 Conclusions and Future Works

In this paper we have presented FCCS, a Fuzzy variant of CCS (*Calculus of Communicating Processes*), that aims at compositionally describing fuzzy behaviours. Operational semantics of FCCS has been defined by means of *Fuzzy Labelled Transition Systems* [6] (FLTS). These are an extension of Labelled Transition Systems where fuzziness is used for modeling uncertainty in concurrent systems. In the paper we have also shown how standard process algebras formal tools, like modal logics and behavioural equivalences, can be used for supporting fuzzy reasoning.

The idea to combine process algebrae and fuzzy sets is not new. For instance, in [17] is introduced a model based on interval values, to solve the problem of nondeterministic choices arising in concurrency and communication systems. However, differently from previous approaches, the present work introduces both a behavioural equivalence and a modal logic for reasoning about fuzzy specifications.

As a future work we plan to use the proposed approach for modeling the examples proposed in literature [18, 19, 20] where Fuzzy Theory is used for modeling the behaviour of "systems", like for instance those involving human interactions, where imprecision is a central feature.

References

[1] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[2] G.J. Klir and T.A. Folger. *Fuzzy sets, uncertainty, and information*. Prentice-Hall International, Inc, 1988.

[3] Jyh-Shing Roger Jang and Chuen-Tsai Sun. Neuro-fuzzy modeling and control. *Proceedings of the IEEE*, 83(3), 1995.

[4] Qiang Song and Robert P. Leland. Adaptive learning defuzzification techniques and applications. *Fuzzy Sets and Systems*, 81(3):321 – 329, 1996.

[5] Werner Van Leekwijck and Etienne E. Kerre. Defuzzification: criteria and classification. *Fuzzy Sets and Systems*, 108(2):159 – 178, 1999.

[6] Liliana D'Errico and Michele Loreti. Modeling fuzzy behaviours in concurrent systems. In L. Laura G. Italiano, E. Moggi, editor, *Proceedings of the 10th Italian Conference on Theoretical Computer Science ICTCS'07*, pages 94–105, 2007.

[7] R. Milner. *Communication and concurrency*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1989.

[8] S. D. Brookes, C. A. R. Hoare, and A. W. Roscoe. A theory of communicating sequential processes. *J. ACM*, 31(3):560–599, 1984.

[9] J.A. Bergstra and J.W. Klop. Algebra of communicating processes with abstraction. *Theoretical Computer Science*, (37):77–121, 1985.

[10] L. A. Zadeh. Fuzzy logic = computing with words. *IEEE Transactions on Fuzzy Systems*, 4(2):103–111, 1996.

[11] J. A. Goguen. L-fuzzy sets. *J. Math. Anal. Appl.*, 18:145–174, 1967.

[12] Matthew Hennessy and Robin Milner. Algebraic laws for nondeterminism and concurrency. *J. ACM*, 32(1):137–161, 1985.

[13] Roberto Segala and Nancy Lynch. Probabilistic simulations for probabilistic processes. *Nordic J. of Computing*, 2(2):250–273, 1995.

[14] Ed Brinksma and Holger Hermanns. Process algebra and markov chains. In Ed Brinksma, Holger Hermanns, and Joost-Pieter Katoen, editors, *Euro Summer School on Trends in Computer Science*, volume 2090 of *Lecture Notes in Computer Science*, pages 183–231. Springer, 2001.

[15] A. Parma and R. Segala. Logical characterizations of bisimulations for discrete probabilistic systems. In Springer-Verlag, editor, *FOSSACS 2007, LNCS*, volume 4423, pages 287–301, 2007.

[16] A. Tarski. A lattice-teoretichal fixpoint theorem and its applications. *Pacific J. Math.*, 5:285–309, 1955.

[17] Yixiang Chen and Jie Zhou. Fuzzy interval-valued processes algebra. In S.Barry Cooper, Benedict Lowe, and Leen Torenvliet, editors, *CiE 2005 Abstracts*, ILLC Publications, pages 44–57, 2005.

[18] L. A. Zadeh. Commonsense reasoning based on fuzzy logic. In *WSC '86: Proceedings of the 18th conference on Winter simulation*, pages 445–447, New York, NY, USA, 1986. ACM Press.

[19] Eviatar Tron and Michael Margaliot. Mathematical modeling of observed natural behavior: a fuzzy logic approach. *Fuzzy Sets and Systems*, 2004.

[20] Bernd Möller, Wolfgang Graf, and Song Ha Nguyen. Modeling the life cycle of a structure using fuzzy processes. *Computer-Aided Civil and Infrastructure Engineering*, 2004.

ISBN: 978-989-95079-6-8

A Generalized Numerical Solution for Fuzzy Relation Equations

Iman Mohammadi Ardehali[1] Milad Avazbeigi[2]

1 Faculty of Management, University of Tehran
Tehran, Iran
2 Department of Industrial Engineering, Amirkabir University of Technology
Tehran, Iran
Email: iman@hamgerayesh.com, miladavazbeigi@aut.ac.ir

Abstract— *In this paper, line search based on Sequential Quadratic Programming is implemented in order to find a solution to Fuzzy Relation Equations. Sequential Quadratic Programming is a gradient-based method that uses a quadratic estimation of the objective function in each iteration's neighborhood. Unlike analytical approaches, the method can handle equations with any combinations of t-norms and t-conorms and at any dimensions. It is assumed that the FRE problem has at least one solution.*

Keywords— Fuzzy Relation Equation (FRE), Fuzzy Triangular Norms, Line Search, Numerical Solution, Sequential Quadratic Programming (SQP).

1 Introduction

Fuzzy Relation Equations were first introduced by Sanchez [1] in 1976. Most of the works done on fuzzy relation equations are focused on specific non-parametric t-norms (min and product) and t-conorms (max and probabilistic sum). Markovskii [2], Shieh [3-5], Molai and Khorram [6], Hana et al. [7], and Perfilieva and Nosková [8] have presented some algorithms for FREs with such compositions. In special cases, an exact solution can be obtained through direct analytical methods. However, for parametric operators like *Schweizer-Sklar*, *Frank*, *Yager*, *Sugeno-Weber*, *Dombi*, and *Dubois-Prade* there is no direct solution available. This is mostly because of the complex non-linear nature of these operators.

An FRE can be formed as:

$$P \circ Q = R \tag{1}$$

where $\circ$ denotes the S-T composition. The problem involves finding P from Q and R.

Stamou and Tzafestas [9] developed a criterion for the existence of minimal solutions, and interpreted a *Fuzzy Inference System* (FIS) as an FRE, indicating that FREs are appropriate for FIS implementation. Shieh [3, 4] recently extended the discussion to a general condition using continuous t-norms. In such cases, when the solution set for (1) is not empty, it can be completely determined by a unique maximum solution and a finite number of minimal solutions.

However, the current models mostly discuss analytical approaches to solve the FRE problem in case-specific conditions—such as max-min or max-product compositions and sizes of $1\times m$ and $1\times n$ for binary fuzzy relations P and R.

In this paper, we propose a new numerical method based on *Sequential Quadratic Programming* (SQP) line search. The proposed model's advantages are solving FREs for any set of parametric operators and capability of handling sizes of $p\times m$ and $p\times n$ for P and R.

It should be noted that the fundamental assumption of this effort is that there exists at least one solution to the FRE problem.

The rest of this paper is organized as follows. In section 2, basic concepts of Fuzzy Relation Equations and Sequential Quadratic Programming are presented. Section 3 describes the method in detail. In section 4, some examples are provided for demonstration of the model's performance. Finally, section 5 gives the conclusions and suggests ideas for future research.

2 Basic Concepts

2.1 Parametric Fuzzy Operators

The most familiar fuzzy operators are max and algebraic sum (as t-conorms or s-norms) and min and product (as t-norms). But they are not limited to these special cases, and can be parametrically constructed.

The advantages of using parametric operators are:

- Most of the parametric operators have the capability to reduce to the aforementioned non-parametric operators, with specific values for the parameters. For example, the Frank t-norm reduces to min when $p = 0$.
- The parameters can be tuned according to the specific requirements of the system in question.

On the other hand, the disadvantage is that the increasing complexity leads to longer running times and more sensitivity to the tuning mechanism.

2.2 Fuzzy Relation Equations

An FRE is as presented in (1), in which $\circ$ acts as the S-T composition, meaning:

$$\bigvee_{j=1}^{m}\left[T\left(p_{ij}, q_{jk}\right)\right] = r_{ik} \tag{2}$$

where P, Q, and R are fuzzy binary relations and p, q, and r are their elements respectively [10]. The Yager series of t-norms and t-conorms are well-known fuzzy operators:

$$T_w(a,b) = 1 - \min\left\{1, \left((1-a)^w + (1-b)^w\right)^{1/w}\right\} \tag{3}$$

$$S_w(a,b) = \min\left\{1, \left(a^w + b^w\right)^{1/w}\right\} \tag{4}$$

ISBN: 978-989-95079-6-8

Equation (3) represents the t-norm and (4) is the t-conorm of Yager operators.

2.3 Sequential Quadratic Programming

SQP methods represent the state of the art in nonlinear programming methods. Schittkowski [11], for example, has implemented and tested a version that outperforms every other tested method in terms of efficiency, accuracy, and percentage of successful solutions, over a large number of test problems.

The method closely mimics Newton's method for constrained optimization just as is done for unconstrained optimization. At each major iteration, an approximation is made of the Hessian of the Lagrangian function using a quasi-Newton updating method. This is then used to generate a Quadratic Programming sub-problem whose solution is used to form a search direction for a line search procedure.

Consider the following general non-linear minimization problem (*P*):

$$Minimize\ f(x)$$

$$subject\ to \begin{cases} h_i(x) = 0; & i = 1,...,l \\ h_j(x) \le 0; & j = l+1,...,m \end{cases}$$

where $x \in R^n, f: R^n \to R, h: R^n \to R^m$.

At *k*th iteration, the SQP algorithm generates a search direction (*d*) solving the definite quadratic sub-problem (*Q*) below:

$$Minimize\ \nabla f_k^T d + \tfrac{1}{2} d^T B_k d$$

$$subject\ to \begin{cases} h_i(x_k) + \nabla h_i^T(x_k) d = 0; & i = 1,...,l \\ h_i(x_k) + \nabla h_i^T(x_k) d \le 0; & j = l+1,...,m \end{cases}$$

where B_k is a positive definite approximation to the Hessian matrix of the Lagrangian function of the original problem (*P*):

$$L(x,\mu) = f(x) + \sum_{i=1}^{m} \mu_i h_i(x) \tag{5}$$

At each major iteration a positive definite quasi-Newton approximation of the Hessian of the Lagrangian function, *L*, is calculated using Broyden–Fletcher–Goldfarb–Shanno (BFGS) method [12].

A detailed description of the SQP method for nonlinear optimization can be found in Fletcher [13], Gill et al. [14], Powell [15], Hock and Schittkowski [16], Nocedal and Wright [17], and Bartholomew-Biggs [18].

3 The Method

SQP, as any other line search method, requires an initial solution to start from. A fairly large random search of the solution space can provide us with this initial solution. Random solutions are generated as matrices of *p*×*m*—dimensions of *P* in (1)—with elements uniformly distributed between 0 and 1.

There are various line search methods for solving nonlinear problems, the most prominent of which are gradient descent, the Newton method, quasi-Newton methods, and Sequential Quadratic Programming. These four were thoroughly tested for our problem and SQP turned out to be the most superior in terms of running time and error measure. So SQP is the line search algorithm of choice for our method.

Since our goal is to minimize the final error, the objective function would rationally be *RMSE*, that is, the root-mean-square of the errors obtained by the difference between the actual *R* and the one calculated from the composition below.

$$R' = P' \circ Q \tag{6}$$

where *P′* is the solution in search of *P*. Thus, the objective function will be as follows.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{p}\sum_{k=1}^{n}(r_{ik} - r'_{ik})^2}{p \times n}} \tag{7}$$

Using (2), we arrive at this formula:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{p}\sum_{k=1}^{n}\left\{\bigcirc\!\!\!\!\!\!\sum_{j=1}^{m}\left[T(p_{ij}, q_{jk})\right] - \bigcirc\!\!\!\!\!\!\sum_{j=1}^{m}\left[T(p'_{ij}, q_{jk})\right]\right\}^2}{p \times n}} \tag{8}$$

Constraints for *P′* elements are obviously the lower band of 0 and the upper band of 1.

This approach guarantees that regardless of the operators chosen, the difference between *R* and *R′* will be minimized. Moreover, continuity and differentiability are necessary conditions for the line search. Specifically, operators containing such terms as $\max\{c, f(a,b)\}$ render SQP unusable. In situations like this, the following substitutions will be used:

$$\max\{c, f(a,b)\} = \frac{c + f(a,b)}{2} + \frac{|c - f(a,b)|}{2} \tag{9}$$

$$\min\{c, f(a,b)\} = \frac{c + f(a,b)}{2} - \frac{|c - f(a,b)|}{2} \tag{10}$$

It's worth mentioning that unlike the "minimum" and "maximum" functions, the "absolute value" function is differentiable using symbolic math—across all its domain except where $u(x) = 0$.

$$\frac{d}{dx}|u(x)| = \frac{|u(x)|}{u(x)} u'_x(x) \tag{11}$$

4 Sample Results

The MATLAB 2008a environment was chosen to implement the proposed method, because of its efficiency in matrix operations and also the built-in functions and methods. The sample M code for Yager's operators can be found in the appendix.

For each run, *P* and *Q* are generated randomly and *R* is obtained through $P \circ Q$ with predetermined t-norm and t-conorm parameters to guarantee the existence of at least one solution for *P*. This initial value of *P* is then discarded so as not to interfere with the search process.

P, *Q*, and *R* are all 5×5 in the samples, since dimensions do not cause any significant change in the method's behavior. The initial random search produces 2000 solutions and chooses the one with the least *RMSE* for the SQP initialization. The stopping criterion for SQP is the number of *RMSE* evaluations, which was determined to be 1500.

ISBN: 978-989-95079-6-8

Operators of the same type (e.g. Dombi) were chosen in each run. Sample results are summarized in Table 1 for the following triangular norms and conorms:

- Frank
- Dombi
- Schweizer & Sklar 1
- Schweizer & Sklar 2
- Schweizer & Sklar 3
- Schweizer & Sklar 4
- Yager

More information on parametric fuzzy operators can be found in [10].

The method was repeated 10 times for each operator, with the best, worst, and average results for *RMSE* reported in the Table 1. Also included is the average running time.

Table 1 shows the performance of the proposed method with various parametric operators. The results imply that the Frank operators provide the most suitable search space for this method.

Table 1: test results for various fuzzy operators

Fuzzy Operator	Worst RMSE	Best RMSE	Average RMSE	Average Running Time
Frank	0.0008	0.00005	0.000313	222.48
Dombi	0.064	0.0009	0.026438	88.11
Schweizer & Sklar 1	0.014	0.0003	0.003795	300.98
Schweizer & Sklar 2	0.064	0.0050	0.018533	188.55
Schweizer & Sklar 3	0.046	0.0006	0.016644	166.08
Schweizer & Sklar 4	0.036	0.0003	0.008164	363.53
Yager	0.072	0.0001	0.020254	1006.42

For the sake of clarity, a numerical example will be fully featured here .Given the following values for Q and R, we want to estimate P such that

$$P \circ Q = R$$

using the Frank norm.

$$Q = \begin{bmatrix} 0.3500 & 0.3500 & 0.2900 & 0.0800 & 0.1300 \\ 0.2000 & 0.8300 & 0.7600 & 0.0500 & 0.5700 \\ 0.2500 & 0.5900 & 0.7500 & 0.5300 & 0.4700 \\ 0.6200 & 0.5500 & 0.3800 & 0.7800 & 0.0100 \\ 0.4700 & 0.9200 & 0.5700 & 0.9300 & 0.3400 \end{bmatrix}$$

$$R = \begin{bmatrix} 0.8551 & 0.9785 & 0.9356 & 0.9769 & 0.6889 \\ 0.6664 & 0.9419 & 0.8942 & 0.6672 & 0.6712 \\ 0.6423 & 0.9227 & 0.8260 & 0.8592 & 0.5874 \\ 0.4230 & 0.8127 & 0.8054 & 0.5878 & 0.5892 \\ 0.6713 & 0.9630 & 0.9023 & 0.9533 & 0.6681 \end{bmatrix}$$

After 57 iterations (1500 evaluations), P' is the estimated value for P.

$$P' = \begin{bmatrix} 0.2921 & 0.4785 & 0.7671 & 0.9538 & 0.8433 \\ 0.6804 & 0.9593 & 0.2627 & 0.5497 & 0.2491 \\ 0.2617 & 0.5455 & 0.3835 & 0.4141 & 0.7563 \\ 0.1811 & 0.5821 & 0.7026 & 0.1342 & 0.2487 \\ 0.0139 & 0.2564 & 0.8796 & 0.3088 & 0.9254 \end{bmatrix}$$

Then, R' is calculated according to (6):

$$R' = P' \circ Q$$

$$R' = \begin{bmatrix} 0.8549 & 0.9785 & 0.9360 & 0.9772 & 0.6888 \\ 0.6664 & 0.9418 & 0.8942 & 0.6672 & 0.6713 \\ 0.6417 & 0.9236 & 0.8279 & 0.8590 & 0.5859 \\ 0.4231 & 0.8125 & 0.8052 & 0.5878 & 0.5894 \\ 0.6715 & 0.9629 & 0.9021 & 0.9536 & 0.6682 \end{bmatrix}$$

And here's the difference between the actual R and the one obtained by the estimation of P (P'), which will lead to our error measure:

$$R' - R = \begin{bmatrix} -0.0002 & 0.0000 & 0.0005 & 0.0003 & -0.0001 \\ 0.0000 & -0.0001 & 0.0000 & 0.0000 & 0.0001 \\ -0.0005 & 0.0009 & 0.0019 & -0.0002 & -0.0015 \\ 0.0001 & -0.0002 & -0.0002 & 0.0000 & 0.0002 \\ 0.0002 & -0.0001 & -0.0002 & 0.0003 & 0.0000 \end{bmatrix}$$

Finally, by applying (7), we arrive at the value of *RMSE* as a universal error measure.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{5}\sum_{k=1}^{5}(r_{ik} - r'_{ik})^2}{5 \times 5}} = 5.5246\text{E-}04$$

5 Conclusions and Future Research

In this paper, we have proposed a new numerical method based on Sequential Quadratic Programming for solving fuzzy relation equations. This work is distinguished from previous methods in that:

- it views the FRE problem in the most general sense, being capable of handling sizes of $p \times m$ and $p \times n$ for P and R; and
- unlike analytical approaches, it can handle equations with any combinations of t-norms and t-conorms and at any dimensions.

However, the proposed method is not capable of tuning the parameters of t-norms and t-conorms. Also, it can be extended to be able to handle systems of Fuzzy Relation Equations. Such improvements can be considered for future research.

Appendix

The sample M file only for Yager's fuzzy operators is available through the following URL:
http://h1.ripway.com/commonlove1985/SQP.txt

References

[1] E. Sanchez, Resolution of Composite Fuzzy Relation Equations, *Information and Control*, 30: 38-48, 1976.

[2] A. V. Markovskii, On the Relation between Equations with Max-Product Composition and the Covering Problem, *Fuzzy Sets and Systems*, 153: 261-273, 2005.

[3] B. Shieh, Solutions of Fuzzy Relation Equations Based on Continuous t-norms, *Information Sciences*, 177: 4208-4215, 2007.

ISBN: 978-989-95079-6-8

[4] B. Shieh, Infinite Fuzzy Relation Equations with Continuous t-norms, *Information Sciences*, 178: 1961-1967, 2008.

[5] B. Shieh, Deriving Minimal Solutions for Fuzzy Relation Equations with Max-Product Composition, *Information Sciences*, 178: 3766-3774, 2008.

[6] A. A. Molai and E. Khorram, An Algorithm for Solving Fuzzy Relation Equations with Max-T Composition Operator, *Information Sciences*, 178: 1293-1308, 2008

[7] S. Hana, H. Lia and J. Wanga, Resolution of Finite Fuzzy Relation Equations Based on Strong Pseudo-t-norms, *Applied Mathematics Letters,* 19: 752-757, 2006.

[8] I. Perfilieva and L. Nosková, System of Fuzzy Relation Equations with inf-⟶ Composition: Complete Set of Solutions, *Fuzzy Sets and Systems*, 159: 2256-2271, 2008

[9] G. B. Stamou and S. G. Tzafestas, Resolution of Composite Fuzzy Relation Equations Based on Archimedean Triangular Norms, *Fuzzy Sets and Systems*, 120: 395-407, 2001.

[10] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall PTR, 1995.

[11] K. Schittkowski, *NLQPL: A FORTRAN-Subroutine Solving Constrained Nonlinear Programming Problems*, Annuals of Operations Research, 5: 485-500, 1985.

[12] J. Nocedal, S. J. Wright, *Numerical Optimization* (2nd ed.), Berlin, New York: Springer-Verlag, 2006

[13] R. Fletcher, *Practical Methods of Optimization*, John Wiley and Sons, 1987.

[14] P. E. Gill, W. Murray and M. H. Wright, *Practical Optimization*, London. Academic Press, 1981.

[15] M. J. D. Powell, *Variable Metric Methods for Constrained Optimization. Mathematical Programming: The State of the Art*, Springer Verlag, 288-311, 1983.

[16] W. Hock, K. Schittkowski, *A Comparative Performance Evaluation of 27 Nonlinear Programming Codes*, Computing, 30: 335, 1983

[17] J. Nocedal and S. J. Wright, *Numerical Optimization (Second Edition): Springer Series in Operations Research and Financial Engineering*, Springer Science+Business Media, LLC., 2006.

[18] M. Bartholomew-Biggs, *Nonlinear Optimization with Engineering Applications: Springer Optimization and Its Applications*, Springer Science+Business Media, LLC., 2008.

ISBN: 978-989-95079-6-8

A Multi-objective Evolutionary Algorithm for Tuning Fuzzy Rule-Based Systems with Measures for Preserving Interpretability.

M.J. Gacto[1] R. Alcalá, F. Herrera[2]

1 Dept. Computer Science, University of Jaen
Jaen, Spain
2 Dept. Computer Science and Artificial Intelligence, University of Granada
Granada, Spain
Email: mjgacto@ugr.es, {alcala,herrera}@decsai.ugr.es

Abstract— *In this contribution we propose a multi-objective evolutionary algorithm for Tuning Fuzzy Rule-Based Systems by considering two objectives, accuracy and interpretability. To this aim we define a new objective that allows preserving the interpretability of the system. This new objective is an interpretability index which is the union of three metrics to preserve the original shapes of the membership functions as much as possible while a tuning of the membership function parameters is performed. The proposed method has been compared to a single objective accuracy-guided algorithm in two real problems showing that many solutions in the Pareto front dominate to those obtained by the single objective-based one.*

Keywords— Fuzzy Rule-Based Systems, Tuning, Interpretability, Multi-Objective Evolutionary Algorithms.

1 Introduction

Fuzzy modeling usually tries to improve the accuracy of the system without inclusion of any interpretability measure, an essential aspect of Fuzzy Rule-Based Systems (FRBSs). However, the problem of finding the right trade-off between accuracy and interpretability has achieved a growing interest [1].

One of the most widely-used approaches to enhance the performance of FRBSs is the *tuning* of the Membership Functions (MFs). It consists of refining a previous definition of the Data Base (DB) once the Rule Base (RB) has been obtained. Generally, tuning is a variation in the shape of the MFs that improves their global interaction. Classically, the tuning methods refine the parameters that identify the MFs associated to the labels comprising the DB [2].

For this model, to take into account interpretability issues it is necessary to use a measure to quantify the interpretability of the tuned DB. This kind of measure could be used as an additional objective to maintain the interpretability of the fuzzy partitions when the tuning is carried out.

In the literature, many authors improve the difficult trade-off between accuracy and interpretability of FRBSs, obtaining linguistic models not only accurate but also interpretable. We can distinguish two kinds of approaches for managing the interpretability:

1. The complexity of the model [3, 4, 5, 6] (usually measured as number of rules, variables, labels per rule, etc.)
2. Measuring the interpretability of the fuzzy partitions [4, 7, 8] by means of a semantic interpretability measure.

A way for optimizing both objectives (accuracy and interpretability) is the use of the Multi-Objective Evolutionary Algorithms (MOEAs) [9, 10]. Since this problem presents a multi-objective nature the use of MOEAs to obtain a set of solutions with different degrees of accuracy and interpretability is an interesting way to work [3, 5, 6, 8].

In this work, we focus our attention in measuring the interpretability of the fuzzy partitions. We propose a semantics interpretability index for maintaining the interpretability of the system by means of the aggregation of several metrics, with the aim of preserving the original form of the MFs while a tuning is performed. These metrics try to minimize the displacement of the central point of the MFs, besides maintaining the symmetry and the area of the original MFs associated to the linguistic labels. To this end, we apply a specific MOEA to obtain accurate and interpretable linguistic fuzzy models by performing a tuning of the MFs parameters with two main objectives:

- The system error
- The proposed interpretability index.

This algorithm is based on the well known $SPEA2$ [11] and is called SPEA2$_{SI}$ ($SPEA2$ for Semantic Interpretability). In order to improve its search ability, SPEA2$_{SI}$ implements some concepts as incest prevention and restarting [12]. Moreover, it is focused on the most accurate solutions to finally form a wide Pareto front. Thus, SPEA2$_{SI}$ is aimed at generating a complete set of Pareto-optimum solutions with different trade-offs between accuracy and interpretability. We have not considered the well-known NSGA-II [13] algorithm since in [3], approaches based on SPEA2 were shown to be more effective when performing a tuning of the MFs.

To show the good performance of the proposed method it is compared with a single objective accuracy-guided tuning algorithm [14] by applying both of them to initial linguistics models obtained from automatic learning methods. Two real-world problems with different complexities have been considered showing that the solutions of the accuracy based algorithm are dominated by those obtained by SPEA2$_{SI}$. We can see as both objectives required are certainly contradictory as the obtained Pareto fronts moves clearly from the most accurate solutions to the most interpretable ones.

Next section describes the classic tuning used in this work. Section 3 presents the proposed index to control the interpretability of the MFs. Section 4 presents the SPEA2$_{SI}$ algorithm describing its main characteristics and the genetic operators considered. Section 5 shows the experimental study

ISBN: 978-989-95079-6-8

and the results obtained. Finally, in section 6 we point out some conclusions.

2 Preliminaries: Tuning of MFs

This approach, usually called DB tuning, involves refining the MF shapes from a previous definition once the remaining FRBS components have been obtained [15, 16, 17, 18, 19, 20]. The classic way to refine the MFs is to change their definition parameters. For example, if the following triangular-shape MF is considered:

$$\mu(x) = \begin{cases} \frac{x-a}{b-a}, & \text{if } a \leq x < b \\ \frac{c-x}{c-b}, & \text{if } b \leq x \leq c \\ 0, & \text{otherwise} \end{cases}$$

changing the basic parameters — a, b, and c — will vary the shape of the fuzzy set associated to the MF, thus influencing the FRBS performance (See Figure 1). The same yields for other shapes of MFs (trapezoidal, gaussian, sigmoid, etc.).

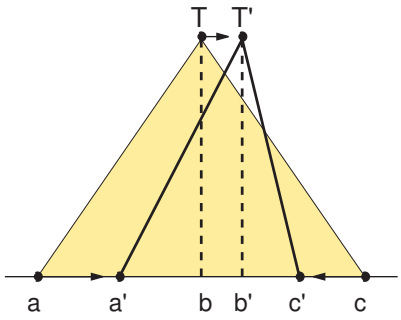


Figure 1: Tuning by changing the basic MF parameters

Tuning involves fitting the characterization of the MFs associated to the primary linguistic terms considered in the system. Thus, the meaning of the linguistic terms is changed from a previous definition (an initial DB). In order to ensure the semantics integrity through the MFs optimization process [1, 21, 22], some researchers have proposed several properties. Considering one or more of these properties several constraints can be applied in the design process in order to obtain a DB maintaining the linguistic model comprehensibility to the higher possible level [15, 23, 24, 25].

In this work we consider strong fuzzy partitions to define the initial MFs. In order to maintain semantics integrity we consider also these constraints by defining the variation intervals for each MF parameter. For each $MF_j = (a_j, b_j, c_j)$ where j=(1,...m) and m is the number of MFs in a given DB, the variation intervals are calculated in the following way:

$$[I^l_{a_j}, I^r_{a_j}] = [a_j - (b_j - a_j)/2, a_j + (b_j - a_j)/2] \ ,$$
$$[I^l_{b_j}, I^r_{b_j}] = [b_j - (b_j - a_j)/2, b_j + (c_j - b_j)/2] \ ,$$
$$[I^l_{c_j}, I^r_{c_j}] = [c_j - (c_j - b_j)/2, c_j + (c_j - b_j)/2] \ .$$

Thanks to these restrictions it is possible to maintain the comprehensibility of MFs to a reasonable level. In any case, it would be very interesting to have a measure for the quality of the tuned MFs. We propose three metrics trying to preserve the original form of the MFs, so improving if possible the trade-off between accuracy and interpretability.

3 An Interpretability Index based on Three Interpretability Metrics

In this section, we propose several metrics to measure the interpretability when a tuning is performed on the DB. At this point, we should remark that these metrics are based on the existence of the variation intervals (integrity constraints) defined in the previous section and, therefore, on the assumption that the initial DB is comprised of triangular uniformly distributed MFs (strong fuzzy partitions) with an associated linguistic meaning. Based on these assumptions, significant changes in the DB can influence negatively in the interpretability. In this way, each metric is proposed to control how good are different desirable aspects of the tuned MFs with respect to the original ones. The metrics proposed are:

- MFs displacements (δ): This metric measures the proximity of the central points of the MFs to the original corresponding ones. It should be higher as closer to the original point.
- MFs symmetry (γ): This metric measures the symmetry of the MFs. It should be higher when the two parts of the support of the MFs (left and right) present more similar lengths.
- MFs area similarity (ρ): This metric measures the similarity of the area of the tuned MFs to the original one. It should be higher when the tuned area and the original area are closer.

In the following subsections the three proposed metrics will be explained in depth.

3.1 MFs displacements (δ)

This metric can control the displacements in the central point of the MFs. It is based on computing the normalized distance between the central point of the tuned MF and the central point of the original MF, and is calculated through keeping the maximum displacement obtained on all the MFs. For each MF_j: $\delta_j = |b_j - b'_j|/I$, where $I = (I^r_{b_j} - I^l_{b_j})/2$ represents the maximum variation for each central parameter. Thus δ^* is defined as: $\delta^* = max_j\{\delta_j\}$.

The δ^* metric takes values between 0 and 1, thereby value near to 1 represent that the MFs present a great displacement. The following transformation is made so this metric represents proximity (maximization):

$$\delta = 1 - \delta^* \ ,$$

3.2 MFs symmetry (γ)

This metric can be used to control the symmetry of the MF shapes. It is based on relating the two parts of the support of the tuned MFs. Let us define that $leftS'_j = |a'_j - b'_j|$ is the length of the left part of the MF support and that $rightS'_j = |b'_j - c'_j|$ is the right part. γ_j is calculated using the following equation:

$$\gamma_j = \frac{min\{leftS'_j, rightS'_j\}}{max\{leftS'_j, rightS'_j\}} \ .$$

Values near to 1 mean that the two parts of the support of the MFs are more similar (higher symmetry). Finally γ is calculated by keeping the minimum value obtained.

$$\gamma = min_j\{\gamma_j\} \ .$$

ISBN: 978-989-95079-6-8

3.3 MFs area similarity (ρ)

This metric can be used to control the area of the MF shapes. It is based on relating the areas of the original and the tuned MFs. Let us define that A_j is the area of the triangle representing the original MF_j, and A'_j is the new area. ρ_j is calculated using the following equation for each MF:

$$\rho_j = \frac{min\{A_j, A'_j\}}{max\{A_j, A'_j\}} \ .$$

Values near to 1 mean that the original area and the tuned area of the MFs are more similar (less changes). The ρ metric is calculated through keeping the minimum value obtained:

$$\rho = min_j\{\rho_j\} \ .$$

3.4 A global Semantics Interpretability index based on the aggregation of the three metrics

We propose an aggregation of the metrics in a global index based on the geometric mean, that is denoted as GM3M index. It is clear that if only one of the proposed metrics has very low values a problem in the interpretability will arise. The aggregation operator should consider this fact:

$$GM3M = \sqrt[3]{\delta * \gamma * \rho}$$

The value of $GM3M$ ranges between 0 (the lowest level of interpretability) and 1 (the highest level of interpretability).

4 Proposed Multi-objective Evolutionary Algorithm

The proposed algorithm performs a parametric tuning to improve the system accuracy as a first objective and uses the $GM3M$ index to try to preserve the interpretability. It is called $SPEA2$ for Semantic Interpretability ($SPEA2_{SI}$) and is based on the well-known $SPEA2$[11] algorithm. In the next subsections the main components of this algorithm are described and the specific characteristics are presented.

4.1 Coding Scheme and Initial Gene Pool

We consider a real coding scheme, being m^i the number of labels of each of the n variables comprising the DB,

$$\begin{array}{l} C_i = (a_1^i, b_1^i, c_1^i, \ldots, a_{m^i}^i, b_{m^i}^i, c_{m^i}^i), \\ \ i = 1, \ldots, n \ , \\ C = C_1 C_2 \ldots C_n \ . \end{array}$$

The initial DB is included as first individual of the initial population. The remaining individuals are generated at random within the corresponding variation intervals defined in the previous section.

4.2 Objectives

In this algorithm we use two objectives. They are:

- The interpretability index ($GM3M$) that is a maximization objective representing the geometric mean of the three proposed metrics (interpretability).
- The Mean Squared Error (MSE) that is a minimization objective (accuracy):

$$\text{MSE} = \frac{1}{2 \cdot |E|} \sum_{l=1}^{|E|} (F(x^l) - y^l)^2,$$

with $|E|$ being the data set size, $F(x^l)$ being the output obtained from the FRBS decoded from such chromosome when the l-th example is considered and y^l being the known desired output. The fuzzy inference system considered to obtain $F(x^l)$ is the *center of gravity weighted by the matching* strategy as defuzzification operator and the *minimum t-norm* as implication and conjunctive operators.

4.3 Main Characteristics of $SPEA2_{SI}$

The proposed algorithm makes use of the $SPEA2$ selection mechanism. However in order to improve the search ability of the algorithm the following changes are considered:

- The proposed algorithm includes a mechanism for incest prevention based on the concepts of CHC [12], for maintaining population diversity. This mechanism avoids premature convergence. Only those parents whose hamming distance divided by 4 is higher than a threshold are crossed. The well-known BLX-0.5 [26] crossover is applied to obtain the offspring.

  Since we consider a real coding scheme, we have to transform each gene considering a Gray Code with a fixed number of bits per gene ($BGene$) determined by the system expert. In this way, the threshold value is initialized as:

  $$L = (\#C * BGene)/4,$$

  where $\#C$ is the number of genes of the chromosome.

  At each generation of the algorithm, the threshold value is decremented by one which allows crossing closer solutions. Following the concept of the CHC algorithm that does not used mutation, in our algorithm we do not include mutation operator.

- A restarting operator is applied by maintaining the most accurate individual, and the most interpretable individual as a part of the new population (external population is forced to be empty) and obtaining the remaining individuals with the tuning parameters generated at random within the corresponding variation intervals. This way preserves both the most accurate and the most interpretable solutions obtained.

  Restarting should be applied when we detect that all crossovers are allowed. However in order to avoid premature convergence we apply the first restart if 50 percent of crossovers are detected at any generation. This condition is updated each time restarting is performed as $\%_{Required} = (100 + \%_{Required})/2$. Moreover, the most accurate solution should be improved before each restarting. To preserve a well formed Pareto front at the end, the restarting is not applied in the last evaluations. The number of evaluations without restart can be estimated as the number of evaluations needed to apply the first restart multiplied by 10. Additionally, restart is disabled if it was never applied before reaching the mid of the total number of evaluations.

ISBN: 978-989-95079-6-8

- In each stage of the algorithm (between restarting points), the number of solutions in the external population ($\overline{P}_{t+1}$) considered to form the mating pool is progressively reduced, by focusing only on those with the best accuracy. To do that, the solutions are sorted from the best to the worst (considering accuracy as sorting criterion) and the number of solutions considered for selection is reduced progressively from 100% at the beginning to 50% at the end of each stage by taking into account the value of L.

  In the last evaluations when restart is disabled. This mechanism, whose main objective is focusing on the most accurate solutions, is also disabled in order to obtain a wide well formed Pareto front.

5 Experiments

To evaluate the goodness of the proposed approach, two real-world problems with different complexities (different number of variables and available data) are considered to be solved (these data sets are available at, *http://www.keel.es/*) [27]:

- Estimating the maintenance costs of medium voltage lines in a town (ELE).
- Predicting the Abalone Age (ABA).

In both cases, the well-known *ad hoc* data-driven learning algorithm of Wang and Mendel [28] is applied to obtain an initial set of candidate linguistic rules. To do so, we will consider strong fuzzy partitions of triangular-shaped MFs. Once the initial RB is generated, the proposed post-processing algorithm can be applied.

Methods considered for the experiments are:

- T method performs a classic MFs parameter tuning by only considering the accuracy of the model as the sole objective [14].
- $SPEA2_{SI}$ is the proposed MOEA method for the classic tuning considering two objectives, precision and the semantics interpretability index (GM3M).

5.1 Experimental Set-up

We consider a *5-fold cross-validation model*, i.e., 5 random partitions of data each with 20%, and the combination of 4 of them (80%) as training and the remaining one as test. For each one of the 5 data partitions, the tuning methods have been run 6 times, showing for each problem the averaged results of a total of 30 runs.

In the case of $SPEA2_{SI}$ the averaged values are calculated considering the most accurate solution from each Pareto obtained. In this way, $SPEA2_{SI}$ can be compared with the single objective method T.

The values of the input parameters considered by T are: population size of 61, 100000 evaluations, 0.6 as crossover probability and 0.2 as mutation probability per chromosome. The values of the input parameters considered by $SPEA2_{SI}$ are: population size of 200, external population size of 61, 100000 evaluations and 30 bits per gene for the Gray codification.

5.2 Results and Analysis

Table 1 shows the results obtained with WM, where $\#R$ stands for the number of rules, $\text{MSE}_{tra/tst}$ for the averaged error obtained over the training/test data, σ for their respective standard deviations and $GM3M$ for the interpretability index. This method obtains the initial knowledge bases that will be tuned by T and $SPEA2_{SI}$. For the WM method the interpretability index takes value 1 that is the highest level of interpretability.

Table 1: Results obtained with WM method

Dataset	#R	MSE_{tra}	σ_{tra}	MSE_{tst}	σ_{tst}	GM3M
ELE	65	56136	1498	56359	4685	1
ABA	68	8.407	0.443	8.422	0.545	1

The results obtained by both post-processing methods are shown in Table 2. In addition we also show δ, γ and ρ that represent the individual values of the metrics comprising, and *t* represents the results of applying a *test t-student* (with 95 percent confidence) in order to ascertain whether differences in the performance of the best results are significant when compared with that of the other algorithm in the table. The interpretation of the *t* column is:

- $\star$ represents the best averaged result.
- $+$ means that the best result has better performance than that of the related row.

Analysing the results showed in Table 2, we can highlight the following facts:

- The proposed method obtains the best results in training and test with respect to T in both problems. In the case of the electrical problem, SPEA2_{SI} improves among 7% and 8.5% in training and test respectively and in the abalone problem it obtains an improvement of 1.5%.
- The most accurate solutions from SPEA2_{SI} improve the accuracy and obtain more interpretable models, with 29%(ELE) and 52%(ABA) of improvement in the interpretability index with respect to the T method.

Figure 2 shows the Pareto front obtained with SPEA2_{SI} method, the solution obtained by T and the initial knowledge base obtained by WM in the same data partition and seed of ELE. We can observe that the obtained Pareto front is quite wide. In fact, the number of non dominated solutions is always equal to the external population size. Moreover, the WM solution coincides with the last point of the Pareto front and the solution obtained with T is dominated by several solutions from SPEA2_{SI}. Furthermore there is not overfitting in the results obtained with the proposed method.

The Pareto front obtained allows selecting solutions with different degrees of accuracy and interpretability. Figure 2 presents that an improvement in the interpretability produces lack of precision and an improvement in the precision produces lack of interpretability. This figure clearly shows that

ISBN: 978-989-95079-6-8

Table 2: Results obtained in both problems

Dataset	Method	MSE_{tra}	σ_{tra}	t-test	MSE_{tst}	σ_{tst}	t-test	GM3M	σ_{GM3M}	t-test	δ	γ	ρ
ELE	T	17020	1893	+	21027	4225	=	0.225	0.046	+	0.058	0.337	0.694
	$SPEA2_{SI}$	**15884**	1191	*	**19257**	2893	*	**0.319**	0.170	*	0.225	0.398	0.648
ABA	T	2.688	0.063	+	2.770	0.242	=	0.144	0.051	+	0.032	0.182	0.636
	$SPEA2_{SI}$	**2.648**	0.051	*	**2.744**	0.276	*	**0.298**	0.153	*	0.177	0.392	0.625

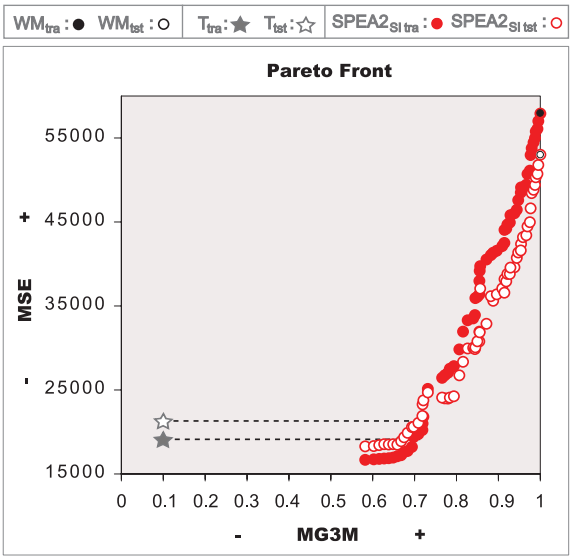


Figure 2: Pareto Front obtained in ELE

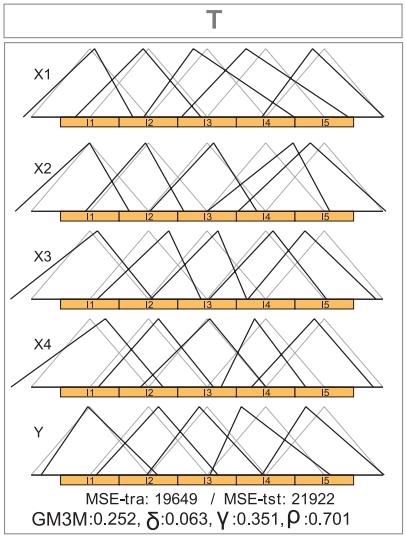


Figure 3: A DB obtained with T method in ELE

both objectives are actually contradictory. In the extremes of the Pareto front an improvement in one objective represents small lost in the other objective. On the contrary in the mid part of the Pareto front improvements in one objective deteriorates the other objective.

Figure 3 presents an illustrative DB obtained with T. Some DBs obtained with $SPEA2_{SI}$, in the same data partition and seed than this shown for T, are shown in Figure 4. This figure includes three DBs, one with the most accurate solution, other with a solution not only accurate but also interpretable and another highly interpretable DB, that obtains 35% of improvement with respect to the WM method with a value of the interpretability index near to 1.

6 Conclusions

The proposed index is useful to preserve the original shape of the MFs, in order to maintain the interpretability and it is a measure of the quality of the DB.

The proposed method obtains wide well formed Pareto fronts that provide a large variety of solutions to select from more accurate solutions to more interpretable ones.

Finally, we can stand out that $SPEA2_{SI}$ algorithm is very competitive and efficient since it is able to maintain the DB interpretability at a better level while accuracy is greatly improved.

Acknowledgment

Supported by the Spanish Ministry of Education and Science under grant no. TIN2008-06681-C06-01.

References

[1] J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, editors. *Interpretability issues in fuzzy modeling*, volume 128 of *Studies in Fuzziness and Soft Computing*. Springer, 2003.

[2] O. Cordón, F. Herrera, F. Hoffmann, and L. Magdalena. *Genetic Fuzzy Systems. Evolutionary tuning and learning of fuzzy knowledge bases*, volume 19 of *Advances in Fuzzy Systems - Applications and Theory*. World Scientific, 2001.

[3] R. Alcalá, M. J. Gacto, F. Herrera, and J. Alcalá-Fdez. A multi-objective genetic algorithm for tuning and rule selection to obtain accurate and compact linguistic fuzzy rule-based systems. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 15(5):539–557, 2007.

[4] José M. Alonso, Luis Magdalena, and Serge Guillaume. Hilk: A new methodology for designing highly interpretable linguistic knowledge bases using the fuzzy logic formalism. *International Journal of Intelligent Systems*, 23(7):761–794, 2008.

[5] A. Botta, B. Lazzerini, F. Marcelloni, and D. C. Stefanescu. Context adaptation of fuzzy systems through a multi-objective evolutionary approach based on a novel interpretability index. *Soft Computing*, 13(5):437–449, 2008.

[6] H. Ishibuchi and Y. Nojima. Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy

ISBN: 978-989-95079-6-8

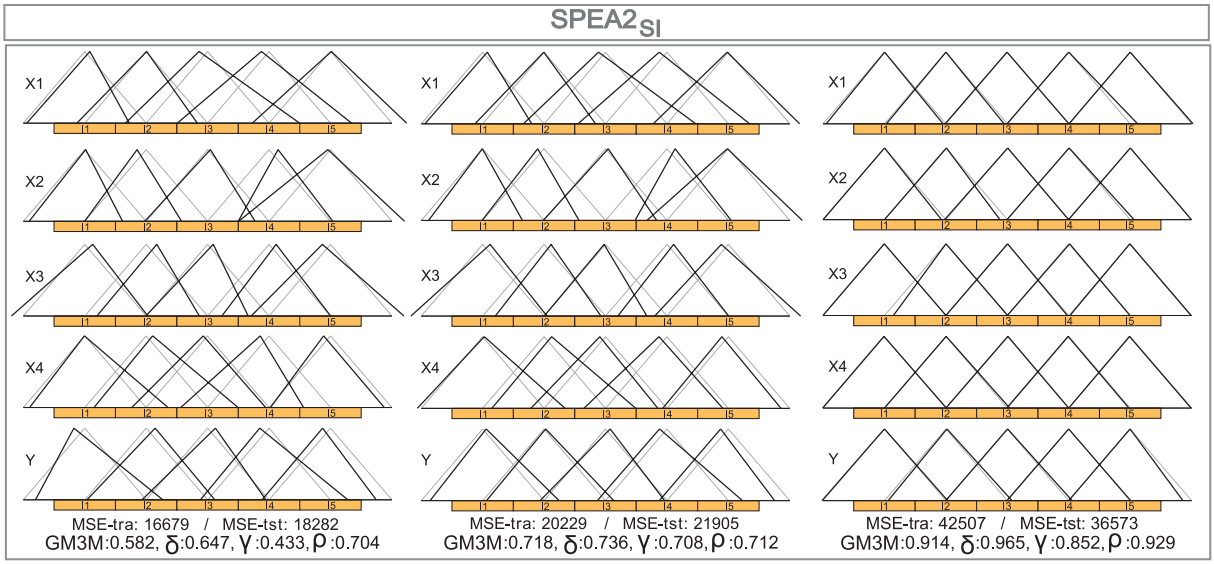


Figure 4: Three DBs obtained in a run of SPEA2$_{SI}$ in ELE

genetics-based machine learning. *International Journal of Approximate Reasoning*, 44(1):4–31, 2007.

[7] J.M. Alonso, O. Cordón, S. Guillaume, and L. Magdalena. Highly interpretable linguistic knowledge bases optimization: Genetic tuning versus solis-wetts. looking for a good interpretability-accuracy trade-off. In *Proceedings of the 2007 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2007)*, pages 901–906, 2007.

[8] P. Fazendeiro and J. V. de Oliveira. A working hypothesis on the semantics/accuracy synergy. In *EUSFLAT-LFA 2005*, pages 266–271, 2005.

[9] C. A. Coello, D. A. Van Veldhuizen, and G. B. Lamont, editors. *Evolutionary algorithms for solving multi-objective problems*. Kluwer Academic Publishers, 2002.

[10] K. Deb. *Multi-objective optimization using evolutionary algorithms*. John Wiley & Sons, NY, USA, 2001.

[11] E. Zitzler, M. Laumanns, and L. Thiele. Spea2: Improving the strength pareto evolutionary algorithm for multiobjective optimization. In *Proc. Evolutionary Methods for Design, Optimization and Control with App. to Industrial Problems*, pages 95–100, Barcelona, Spain, 2001.

[12] L. J. Eshelman. The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination. In G.J.E. Rawlin, editor, *Foundations of genetic Algorithms*, volume 1, pages 265–283. Morgan Kaufman, 1991.

[13] K. Deb, S. Agrawal, A. Pratab, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evolutionary Computation*, 6(2):182–197, 2002.

[14] J. Casillas, O. Cordón, M. J. Del Jesus, and F. Herrera. Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction. *IEEE Trans. Fuzzy Syst.*, 13(1):13–29, 2005.

[15] O. Cordón and F. Herrera. A three-stage evolutionary process for learning descriptive and approximate fuzzy logic controller knowledge bases from examples. *International Journal of Approximate Reasoning*, 17(4):369–407, 1997.

[16] H. B. Gürocak. A genetic-algorithm-based method for tuning fuzzy logic controllers. *Fuzzy Sets and Systems*, 108(1):39–47, 1999.

[17] F. Herrera, M. Lozano, and J. L. Verdegay. Tuning fuzzy logic controllers by genetic algorithms. *International J. of Approximate Reasoning*, 12:299–315, 1995.

[18] Y. Jin, W. von Seelen, and B. Sendhoff. On generating fc^3 fuzzy rule systems from data using evolution strategies. *IEEE Transactions on Systems, Man, and Cybernetics*, 29(6):829–845, 1999.

[19] C.L. Karr. Genetic algorithms for fuzzy controllers. *AI Expert*, 6(2):26–33, 1991.

[20] D. Nauck and R. Kruse. Neuro-fuzzy systems for function approximaton. *Fuzzy Sets and Systems*, 101(2):261–271, 1999.

[21] J. Valente de Oliveira. Semantic constraints for membership function optimization. *IEEE Trans. Syst., Man, Cybern. - Part A: Systems and Humans*, 29(1):128–138, 1999.

[22] J. Valente de Oliveira. Towards neuro-linguistic modeling: constraints for optimization of membership functions. *Fuzzy Sets and Systems*, 106(3):357–380, 1999.

[23] U. Bodenhofer and P. Bauer. A formal model of interpretability of linguistic variables. In Casillas et al. [1], pages 524–545.

[24] F. Cheong and R. Lai. Constraining the optimization of a fuzzy logic controller using an enhanced genetic algorithm. *IEEE Trans. Syst., Man, Cybern. - Part B: Cybernetics*, 30(1):31–46, 2000.

[25] J. Espinosa and J. Vandewalle. Constructing fuzzy models with linguistic integrity from numerical data-AFRELI algorithm. *IEEE Trans. Fuzzy Syst.*, 8(5):591–600, 2000.

[26] L. J. Eshelman and J. D. Schaffer. Real-coded genetic algorithms and interval-schemata. *Foundations of Genetic Algorithms*, 2:187–202, 1993.

[27] J. Alcalá-Fdez, L. Sánchez, S. García, M.J. del Jesus, S. Ventura, J.M. Garrell, J. Otero, C. Romero, J. Bacardit, V.M. Rivas, J.C. Fernández, and F. Herrera. Keel: A software tool to assess evolutionary algorithms for data mining problems. *Soft Computing*, 13(3):307–318, 2009.

[28] L. X. Wang and J. M. Mendel. Generating fuzzy rules by learning from examples. *IEEE Trans. Syst., Man, Cybern.*, 22(6):1414–1427, 1992.

ISBN: 978-989-95079-6-8

On Tolerant Fuzzy c-Means Clustering with L_1-Regularization

HAMASUNA Yukihiro[1] ENDO Yasunori[2] MIYAMOTO Sadaaki[2]

1. Doctoral Program in Risk Engineering, University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan
2. Faculty of Systems and Information Engineering, University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan
Email: yhama@soft.risk.tsukuba.ac.jp, {endo, miyamoto}@risk.tsukuba.ac.jp

Abstract— *We have proposed tolerant fuzzy c-means clustering (TFCM) from the viewpoint of handling data more flexibly. This paper presents a new type of tolerant fuzzy c-means clustering with L_1-regularization. L_1-regularization is well-known as the most successful techniques to induce sparseness. The proposed algorithm is different from the viewpoint of the sparseness for tolerance vector. In the original concept of tolerance, a tolerance vector attributes to each data. This paper develops the concept to handle data flexibly, that is, a tolerance vector attributes not only to each data but also each cluster. First, the new concept of tolerance is introduced into optimization problems. These optimization problems are based on conventional fuzzy c-means clustering (FCM). Second, the optimization problems with tolerance are solved by using Karush-Kuhn-Tucker conditions and an optimization method for L_1-regularization. Third, new clustering algorithms are constructed based on the explicit optimal solutions. Finally, the effectiveness of the proposed algorithm is verified through some numerical examples.*

Keywords— fuzzy c-means clustering, L_1-regularization, optimization, tolerance, uncertainty

1 Introduction

The aim of data mining is to discover important knowledge from a large quantity of data. From this viewpoint, clustering techniques have been actively studied. Clustering is one of the well-known unsupervised classification methods. For example, hard c-means clustering (HCM) is the most basic method. Fuzzy c-means clustering (FCM) is one of the well-known and useful clustering methods. For example, standard fuzzy c-means clustering (sFCM) [1] and entropy regularized fuzzy c-means clustering (eFCM) [2] are representatives. The entropy regularized fuzzy c-means clustering is constructed by regularization with maximum entropy function.

By the way, there are some difficulties of handling a set of data by clustering methods. Some clustering algorithms have been proposed to solve significant problems, for example, data with uncertainty, cluster size, noise or isolated data and so on. When we handle a set of data, data contains inherent uncertainty. For example, errors, ranges or some missing value of attributes are much caused. In these cases, each data is represented by an interval or a set instead of a point. In case of handling data with uncertainty, some significant methods have been proposed [3, 4]. These methods can not only handle data with uncertainty but also obtain high quality results by considering data with uncertainty. Thus, handling data with uncertainty is a very important problem in the field of data mining.

Therefore, some of the authors have proposed the original concept of tolerance [5, 6] which handle data with the above-mentioned uncertainty by using tolerance vector, and constructed some clustering algorithms [7, 8]. In these algorithms, tolerance is defined as hypersphere [5, 6] or hyper-rectangle [7, 8]. In case of hyper-rectangle, the missing value of attributes are handled successfully.

On the other hand, it is difficult to obtain clusters with different size or shape by a single-objective function, e.g., conventional FCM. In general, multi-objective optimization is considered to solve such problems. However, there are some problems by optimization, for example, how to select objective function. We have thought to be solved the above-mentioned problems by introducing a kind of "flexibility" for a pattern space and we have proposed the method to handle such "flexibility" by tolerance vector [9].

It is considered that the constraint for tolerance vector much affects classification results. It means that unsuited parameter makes trivial solutions. Therefore, regularization technique, e.g., Tikhonov's regularization [10] method have been used to solve these ill-posed problems. The quadratic and maximum entropy functions are typical regularization techniques. Moreover, L_1-regularization is the most efficient technique to induce sparseness. In the field of regression models or machine learning, some methods with L_1-regularization have been proposed and given sparse classifiers [12, 13, 14]. In algorithms applied to L_1-regularization, a lot of vari-

ISBN: 978-989-95079-6-8

ables are calculated zero. Non-zero variables are essential to understand classification results.

First, we consider optimization problems of tolerant fuzzy c-means clustering with L_1-regularization. Second, the optimal solutions are derived by Karush-Kuhn-Tucker (KKT) conditions and an optimization method for L_1-regularization . Third, we construct a new clustering algorithm by above-mentioned processes. Finally, the effectiveness of the proposed algorithm is verified through some numerical examples.

2 Preparation

Let data set, cluster and its cluster center be $X = \{x_k \mid x_k = (x_k^1, \ldots, x_k^p)^T \in \Re^p, k = 1 \ldots n\}$, C_i $(i = 1, \ldots, c)$ and $v_i = (v_i^1, \ldots, v_i^p)^T \in \Re^p, (i = 1, \ldots, c)$, respectively. Moreover, u_{ki} is the membership grade of x_k belonging to C_i and we denote the partition matrix $U = (u_{ki})_{k=1\sim n,\ i=1\sim c}$.

2.1 Fuzzy c-means clustering

Fuzzy c-means clustering is based on optimizing an objective function under constraint for membership grade.

We consider following two types of objective functions J_s and J_e.

$$J_s(U,V) = \sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ki})^m \|x_k - v_i\|^2,$$

$$J_e(U,V) = \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ki}\|x_k - v_i\|^2 + \lambda^{-1}\sum_{k=1}^{n}\sum_{i=1}^{c} u_{ki}\log u_{ki}.$$

J_s is a well-known objective function of standard fuzzy c-means clustering (sFCM) proposed by Bezdek [1] and J_e is an entropy regularized fuzzy c-means clustering (eFCM) [2].

The constraint for u_{ki} is as follows :

$$\sum_{i=1}^{c} u_{ki} = 1\ ,\ u_{ki} \in [0,1]\ ,\ {}^{\forall}k.$$

2.2 Regularization

As above-mentioned, regularization is efficient techniques in the field of data mining. Tikhonov's regularization [10] method has been used to solve ill-posed problems. Many data mining algorithms which are applied to regularization have been actively studied to determine a variety of solution [12, 13, 14].

In the field of clustering, many membership regularization techniques have been proposed to obtain a variety of membership functions.

$$\min_U J(U,V) + \lambda\Omega(U),$$

where λ is a regularization parameter and $\Omega(U)$ is a regularization term.

The choice of a regularization term is important to determine a shape of a membership function. The quadratic function and maximum entropy function are typical examples.

$$\Omega(U) = \frac{1}{2}\sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ki})^2,$$

$$\Omega(U) = \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ki}\log u_{ki}.$$

3 The concept of tolerance

In the new concept of tolerance, the vector which attributes not only each data but also cluster center is defined as tolerance vector ε_{ki}. The tolerance κ_{ki} is defined as upper bound of tolerance vector.

From these formulation, the proposed methods can handle data more flexible than conventional methods.

We define κ_{ki}^j as the upper bound of each attribute of tolerance vector $\kappa_{ki} = (\kappa_{11}^1, \ldots, \kappa_{nc}^p) \geq 0$ and tolerance vector $E = \{\varepsilon_{ki} \mid \varepsilon_{ki} = (\varepsilon_{ki}^1, \ldots, \varepsilon_{ki}^p)^T \in \Re^p\ \}$ which mean the admissible range of each data, and the vector within the range of tolerance, respectively. The constraint for ε_{ki}^j is as follows :

$$|\varepsilon_{ki}^j|^2 \leq (\kappa_{ki}^j)^2\ (\kappa_{ki}^j \geq 0), {}^{\forall}k,\ i,\ j.$$

Figure 1 is an illustrative example about the new concept of tolerance defined as hyper-rectangle in $\Re^2$.

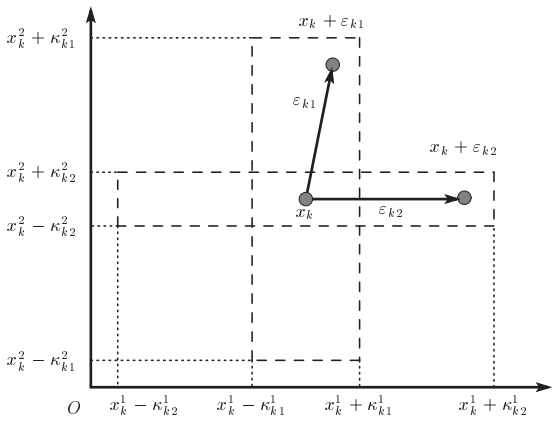


Figure 1: The new concept of tolerance defined as hyper-rectangle in $\Re^2$.

4 Tolerant fuzzy c-means clustering

In this section, we discuss optimization problems for clustering. We formulate the tolerant fuzzy c-means clustering (TFCM) by introducing the notion of tolerance into optimization problems and consider the way

ISBN: 978-989-95079-6-8

to optimize this objective function under the constraints for u_{ki} and ε_{ki}^j. The squared Euclidean-norm is used as dissimilarity, that is,

$$d_{ki} = \|x_k + \varepsilon_{ki} - v_i\|_2^2 = \sum_{j=1}^{p}(x_k^j + \varepsilon_{ki}^j - v_i^j)^2.$$

4.1 Tolerant standard fuzzy c-means clustering

The optimization problem is as follows :

$$J_{ts}(U,E,V) = \sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ki})^m d_{ki}, \tag{1}$$

under the following constraints,

$$\sum_{i=1}^{c} u_{ki} = 1\ ,\ u_{ki} \in [0,1]\ ,\ {}^\forall k, \tag{2}$$

$$|\varepsilon_{ki}^j|^2 \le (\kappa_{ki}^j)^2\ (\kappa_{ki}^j \ge 0), {}^\forall k,\ i,\ j. \tag{3}$$

The goal is to find the solutions which minimize the objective function (1) under the constraints (2) and (3).

From the convexity of (1), we introduce the following Lagrangian function to solve this optimization problem.

The Lagrangian function L_{tsr} is as follows :

$$\begin{aligned} L_{tsr} =& J_{ts}(U,E,V) + \sum_{k=1}^{n}\gamma_k(\sum_{i=1}^{c}u_{ki} - 1) \\ &+ \sum_{k=1}^{n}\sum_{i=1}^{c}\sum_{j=1}^{p}\delta_{ki}^j(|\varepsilon_{ki}^j|^2 - (\kappa_{ki}^j)^2). \end{aligned}$$

Karush-Kuhn-Tucker conditions (KKT conditions) are as follows:

$$\left\{\begin{array}{l} \frac{\partial L_{tsr}}{\partial v_i^j} = 0,\ \frac{\partial L_{tsr}}{\partial \varepsilon_{ki}^j} = 0,\ \frac{\partial L_{tsr}}{\partial u_{ki}} = 0,\ \frac{\partial L_{tsr}}{\partial \gamma_k} = 0, \\ \frac{\partial L_{tsr}}{\partial \delta_{ki}^j} \le 0,\ \delta_{ki}^j\frac{\partial L_{tsr}}{\partial \delta_{ki}^j} = 0,\ \delta_{ki}^j \ge 0. \end{array}\right. \tag{4}$$

First we consider u_{ki}, from $\frac{\partial L_{tsr}}{\partial u_{ki}} = 0$, we have,

$$u_{ki} = \left(\frac{\gamma_k}{m d_{ki}}\right)^{\frac{1}{m-1}}. \tag{5}$$

In addition, from the constraint (2),

$$\sum_{l=1}^{c}\left(\frac{\gamma_k}{m d_{kl}}\right)^{\frac{1}{m-1}} = 1. \tag{6}$$

From (5) and (6), we have,

$$u_{ki} = \frac{\frac{1}{(d_{ki})^{\frac{1}{m-1}}}}{\sum_{l=1}^{c}\frac{1}{(d_{kl})^{\frac{1}{m-1}}}}. \tag{7}$$

If some $x_k + \varepsilon_{ki} = v_i$, we set $u_{ki} = 1/|C'|$. Here, $|C'|$ is number of cluster centers which satisfies $x_k + \varepsilon_{ki} = v_i$.

For v_i^j, from $\frac{\partial L_{tsr}}{\partial v_i^j} = 0$,

$$v_i^j = \frac{\sum_{k=1}^{n}(u_{ki})^m(x_k^j + \varepsilon_{ki}^j)}{\sum_{k=1}^{n}(u_{ki})^m}. \tag{8}$$

For ε_{ki}^j from $\frac{\partial L_{tsr}}{\partial \varepsilon_{ki}^j} = 0$, we can get

$$\varepsilon_{ki}^j = -\frac{(u_{ki})^m(x_k^j - v_i^j)}{(u_{ki})^m + \delta_{ki}^j}. \tag{9}$$

From $\delta_{ki}^j\frac{\partial L_{tsr}}{\partial \delta_{ki}^j} = 0$,

$$\delta_{ki}^j(|\varepsilon_{ki}^j|^2 - (\kappa_{ki}^j)^2) = 0. \tag{10}$$

From (10), we should consider two cases, i.e., $\delta_{ki}^j = 0$ and $|\varepsilon_{ki}^j|^2 = (\kappa_{ki}^j)^2$. First, we consider the case of $\delta_{ki}^j = 0$. In this case, the constraint (3) is not considered. From (9), we can get,

$$\varepsilon_{ki}^j = -(x_k^j - v_i^j).$$

On the other hand, in case that $|\varepsilon_{ki}^j|^2 = (\kappa_{ki}^j)^2$,

$$|\varepsilon_{ki}^j|^2 = |-\frac{(u_{ki})^m(x_k^j - v_i^j)}{(u_{ki})^m + \delta_{ki}^j}|^2 = (\kappa_{ki}^j)^2.$$

From $(u_{ki})^m + \delta_{ki}^j > 0$,

$$\frac{(u_{ki})^m}{(u_{ki})^m + \delta_{ki}^j} = \frac{\kappa_{ki}^j}{|x_k^j - v_i^j|}. \tag{11}$$

From (9), (11),

$$\varepsilon_{ki}^j = \frac{-\kappa_{ki}^j(x_k^j - v_i^j)}{|x_k^j - v_i^j|}.$$

From the above, we can get an optimal solution for ε_{ki}^j as follows :

$$\begin{aligned} \varepsilon_{ki}^j =& -\alpha_{ki}^j(x_k^j - v_i^j), \\ \alpha_{ki}^j =& \min\left\{\frac{\kappa_{ki}^j}{|x_k^j - v_i^j|}, 1\right\}. \end{aligned} \tag{12}$$

4.2 Tolerant entropy regularized fuzzy c-means clustering

The optimization problem is as follows :

$$\begin{aligned} J_{te}(U,E,V) =& \sum_{k=1}^{n}\sum_{i=1}^{c}u_{ki}d_{ki} \\ &+ \lambda^{-1}\sum_{k=1}^{n}\sum_{i=1}^{c}u_{ki}\log u_{ki}. \end{aligned}$$

ISBN: 978-989-95079-6-8

Constraints are same as (2) and (3).
The Lagrangian function L_{er} is as follows :

$$L_{ter} = J_{te}(U,E,V) + \sum_{k=1}^{n}\gamma_k(\sum_{i=1}^{c}u_{ki}-1) + \sum_{k=1}^{n}\sum_{i=1}^{c}\sum_{j=1}^{p}\delta_{ki}^{j}(|\varepsilon_{ki}^{j}|^2-(\kappa_{ki}^{j})^2).$$

KKT conditions are same as (4).
The optimal solutions are derived as follows :

$$u_{ki} = \frac{\exp(-\lambda d_{ki})}{\sum_{l=1}^{c}\exp(-\lambda d_{kl})}, \quad (13)$$

$$v_i^j = \frac{\sum_{k=1}^{n}u_{ki}(x_k^j+\varepsilon_{ki}^j)}{\sum_{k=1}^{n}u_{ki}}, \quad (14)$$

$$\varepsilon_{ki}^j = -\alpha_{ki}^j(x_k^j-v_i^j),$$

$$\alpha_{ki}^j = \min\left\{\frac{\kappa_{ki}^j}{|x_k^j-v_i^j|},1\right\}.$$

5 Tolerant fuzzy c-means clustering with L_1-regularization

In this section, we will consider tolerant fuzzy c-means clustering with L_1-regularization. In this method, constraint for ε_{ki}^j is not considered.

5.1 L_1-regularization

It is well-known that L_1-regularization can induce the strong sparseness of the variables [12, 13, 14]. A lot of ε_{ki}^j become zero, by L_1-regularization described as follows :

$$\Omega(E) = \sum_{k=1}^{n}\sum_{i=1}^{c}\sum_{j=1}^{p}|\varepsilon_{ki}^j|.$$

Here, we will describe objective function with L_1-regularization as J_{tl}. The partial derivatives of J_{tl} which respect to ε_{ki}^j will be uniformly zero, as follows :

$$\left|\frac{\partial J_{tl}}{\partial\varepsilon_{ki}^j}\right| = \nu \qquad \text{if } |\varepsilon_{ki}^j|>0,$$

$$\left|\frac{\partial J_{tl}}{\partial\varepsilon_{ki}^j}\right| < \nu \qquad \text{if } |\varepsilon_{ki}^j|=0.$$

Here, $\nu > 0$ is a regularization parameter. This denotes that if the partial derivatives of J below ν, ε_{ki}^j will be set exactly zero.

5.2 Tolerant standard fuzzy c-means clustering with L_1-regularization

We will consider the following objective function with L_1-regularization.

$$J_{tsl}(U,E,V) = \sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ki})^m d_{ki} + \nu\sum_{k=1}^{n}\sum_{i=1}^{c}\sum_{j=1}^{p}|\varepsilon_{ki}^j|.$$

In the tolerant fuzzy c-means clustering, each tolerance vector can be solved separately. So, we consider the following semi-objective function J_{tsl}^{kij} :

$$J_{tsl}^{kij}(E) = (u_{ki})^m(x_k^j+\varepsilon_{ki}^j-v_i^j)^2 + \nu|\varepsilon_{ki}^j|.$$

To obtain partial derivatives respect to ε_{ki}^j, we will decompose $\varepsilon_{ki}^j = \xi^+ - \xi^-$, where all element of ξ^+ and ξ^- are nonnegative. Thus, the semi-objective function can be written as follows :

$$J_{tsl}^{kij}(E) = (u_{ki})^m(x_k^j+\varepsilon_{ki}^j-v_i^j)^2 + \nu(\xi^+ + \xi^-).$$

The constraints are as follows :

$$\varepsilon_{ki}^j \le \xi^+,$$
$$\varepsilon_{ki}^j \ge -\xi^-,$$
$$\xi^+, \xi^- \ge 0.$$

Introducing the Lagrange multiplier β^+, β^-, δ^+ and $\delta^- \ge 0$, Lagrangian L_{tsl} is as follows :

$$L_{tsl} = (u_{ki})^m(x_k^j+\varepsilon_{ki}^j-v_i^j)^2 + \nu(\xi^+ + \xi^-) + \beta^+(\varepsilon_{ki}^j-\xi^+) + \beta^-(-\varepsilon_{ki}^j-\xi^-) - \delta^+\xi^+ - \delta^-\xi^-.$$

Here,

$$\frac{\partial L_{tsl}}{\partial\xi^+} = \nu-\beta^+-\delta^+, \quad (15)$$

$$\frac{\partial L_{tsl}}{\partial\xi^-} = \nu-\beta^--\delta^-. \quad (16)$$

Since $\delta^+,\ \delta^- \ge 0$, conditions $0 \le \beta^+ \le \nu$ and $0 \le \beta^- \le \nu$ are obtained from (15) and (16), respectively. By using (15) and (16), the Lagrangian L_{tsl} is simplified as follows :

$$L_{tsl} = (u_{ki})^m(x_k^j+\varepsilon_{ki}^j-v_i^j)^2 + \beta\varepsilon_{ki}^j. \quad (17)$$

Here, $\beta = \beta^+ - \beta^-$ and satisfies condition $-\nu \le \beta \le \nu$.
From $\frac{\partial L_{tsl}}{\partial\varepsilon_{ki}^j} = 0$,

$$\frac{\partial L_{tsl}}{\partial\varepsilon_{ki}^j} = 2(u_{ki})^m(x_k^j+\varepsilon_{ki}^j-v_i^j)+\beta = 0.$$

From above,

$$\varepsilon_{ki}^j = -(x_k^j-v_i^j) - \frac{\beta}{2(u_{ki})^m}. \quad (18)$$

Introducing (18) to (17), the Lagrangian dual problem is written as follows :

$$L_{tsld} = -\frac{\beta^2}{4(u_{ki})^m} - \beta(x_k^j-v_i^j).$$

ISBN: 978-989-95079-6-8

This objective function is a quadratic function respect to β. From $\frac{\partial L_{tsld}}{\partial \beta} = 0$, this dual problem is solved as,

$$\beta = -2(u_{ki})^m (x_k^j - v_i^j). \tag{19}$$

From considering (18) and (19), the optimal solution of primal problem is derived. First, if $\beta \leq -\nu$, the optimal solution is $\beta = -\nu$. Second, if $-\nu \leq \beta \leq \nu$, the optimal solution is $\beta = -2(u_{ki})^m (x_k^j - v_i^j)$. Third, if $\nu \leq \beta$, the optimal solution is $\beta = \nu$. Finally, the optimal solution for ε_{ki}^j of L_1-regularized objective function is derived as follows :

$$\varepsilon_{ki}^j = \begin{cases} -(x_k^j - v_i^j) + \frac{\nu}{2(u_{ki})^m} & (\beta \leq -\nu), \\ 0 & (-\nu \leq \beta \leq \nu), \\ -(x_k^j - v_i^j) - \frac{\nu}{2(u_{ki})^m} & (\nu \leq \beta). \end{cases} \tag{20}$$

5.3 Tolerant entropy regularized fuzzy c-means clustering with L_1-regularization

From the same procedure, the optimal solution for ε_{ki}^j of tolerant entropy regularized fuzzy c-means clustering with L_1-regularization is derived as follows :

$$\varepsilon_{ki}^j = \begin{cases} -(x_k^j - v_i^j) + \frac{\nu}{2u_{ki}} & (\beta \leq -\nu), \\ 0 & (-\nu \leq \beta \leq \nu), \\ -(x_k^j - v_i^j) - \frac{\nu}{2u_{ki}} & (\nu \leq \beta). \end{cases} \tag{21}$$

6 Algorithms

Algorithms of TFCM derived in the above section are called as follows. In case of hyper-rectangle, we call these methods TsFCM(R) and TeFCM(R). In case of L_1-regularization, we call these methods TsFCM-L_1R, and TeFCM-L_1R.

Each algorithm of TFCM is calculated according to the following procedure. Eqs. **A**, **B** and **C** used in each algorithm follow **Table 1** .

Algorithm 1

TFCM1 *Set the initial values and parameters.*

TFCM2 *Calculate $u_{ki} \in U$ by Eq.* **A**.

TFCM3 *Calculate $v_i \in V$ by Eq.* **B**.

TFCM4 *Calculate $\varepsilon_{ki} \in E$ by Eq.* **C**. *If convergence criterion is satisfied, stop. Otherwise, go back to* **TFCM2**.

In these algorithms, the convergence criterion is convergence of each variable, value of objective function or number of repetition.

Table 1: The optimal solutions of each algorithm.

Algorithm	Eq.**A**	Eq.**B**	Eq.**C**
TsFCM(R)	(7)	(8)	(12)
TeFCM(R)	(13)	(14)	(12)
TsFCM-L_1R	(7)	(8)	(20)
TeFCM-L_1R	(13)	(14)	(21)

7 Numerical examples

In this section, some numerical examples of classification are shown. In these examples, '$\circ$', '$\triangle$', '$\square$' and '$*$' mean each cluster and cluster centers, respectively. Moreover, tolerance vectors are expressed by arrowed line. The value of each data are normalized between 0 and 10. In addition, $m = 2.0$ in TsFCM and $\lambda = 1.0$ in TeFCM.

A polaris data set is mapped into two dimensional pattern space and consists of 51 points. This data set should be classified into three clusters [1]. Fig. 2 and 3 are classification results of TFCM with L_1-regularization. In Fig. 2, $\nu = 2.0$ and the number of non-zero tolerance vector is 30. In Fig. 3, $\nu = 4.0$ and the number of non-zero tolerance vector is 15. Moreover, Fig. 4 shows the relation between regularization parameter ν and the average of zero parameter ratio out of 1000 trials.

From these results, it is verified that proposed algorithm with L_1-regularization can induce the strong sparseness and its sparseness is controlled by regularization parameter ν.

8 Conclusions

In this paper, we have formulated the optimization problems based on concept of tolerance and derived the optimal solutions of tolerant fuzzy c-means clustering with L_1-regularization. From these results, we have constructed new clustering algorithm. Moreover, we have verified the effectiveness of proposed algorithm through some numerical examples.

The proposed technique is essentially different from the past one from the viewpoint of handling data more "flexible" than conventional methods.

In future works, we will calculate with real data which includes data with uncertainty, e.g., the missing value of attributes. Next, we will consider another type of regularization term or constraint for tolerance vector to induce sparseness.

Acknowledgment

This study is partly supported by Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists and the Grant-in-Aid for Scientific Research (C) and (B) (Project No.21500212 and

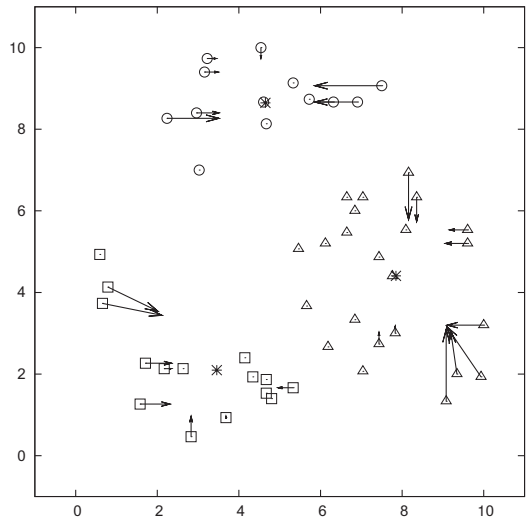


Figure 2: Result of TsFCM with L_1-regularization.

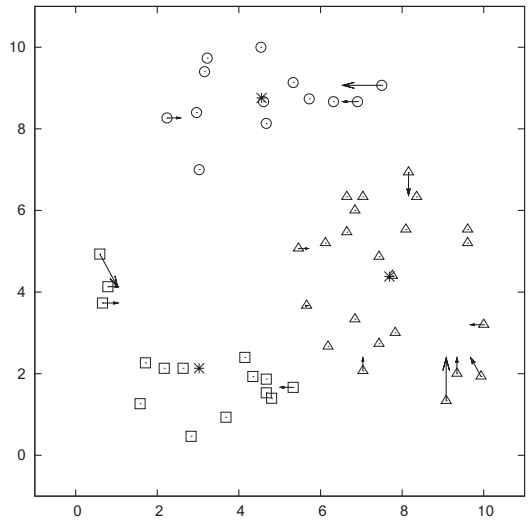


Figure 3: Result of TeFCM with L_1-regularization.

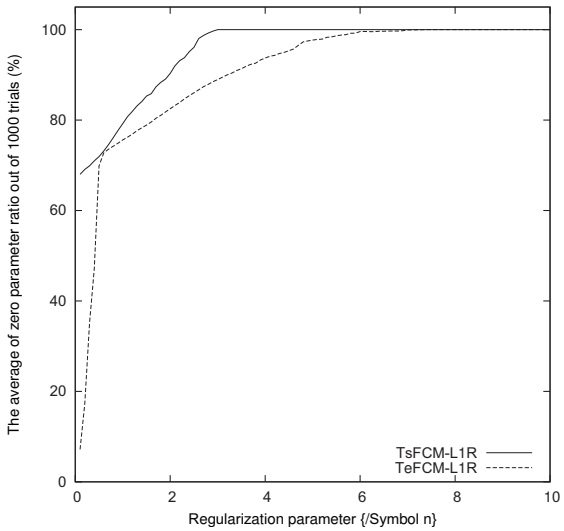


Figure 4: The relation between regularization parameter ν and the average of zero parameter ratio out of 1000 trials

No.19300074) from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

[1] J. C. Bezdek, 'Pattern Recognition with Fuzzy Objective Function Algorithms', Plenum Press,New York, 1981.

[2] S. Miyamoto and M. Mukaidono, "Fuzzy c-means as a regularization and maximum entropy approach", Proc. of the 7th International Fuzzy Systems Association World Congress (IFSA'97), June 25-30, 1997, Prague, Czech, Vol. 2, pp. 86–92, 1997.

[3] W. K. Ngai, B. Kao, C. K. Chui, R. Cheng, M. Chau and K. Y. Yip, "Efficient Clustering of Uncertain Data", Proc. of the Sixth International Conference on Data Mining (ICDM'06), December 18-22, 2006, Hong Kong, China, pp. 436–445, 2006.

[4] O. Takata, and S. Miyamoto, "Fuzzy clustering of Data with Interval Uncertainties", Journal of Japan Society for Fuzzy Theory and Systems, Vol.12, No.5, pp. 686–695, 2000 (in Japanese).

[5] Y. Endo, R. Murata, H. Haruyama and S. Miyamoto, "Fuzzy c-Means for Data with Tolerance", Proc. of International Symposium on Nonlinear Theory and Its Applications, pp. 345–348, 2005.

[6] R. Murata, Y. Endo, H. Haruyama and S. Miyamoto, "On Fuzzy c-Means for data with Tolerance", Journal of Advanced Computational Intelligence and Intelligent Informatics 10(5), pp. 673–681, 2006.

[7] Y. Hasegawa, Y. Endo, Y. Hamasuna and S. Miyamoto, "Fuzzy c-Means for Data with Tolerance Defined as Hyper-rectangles", Modeling Decisions for Artificial Intelligence (MDAI2007), pp. 237-248. Springer, Heidelberg 2007.

[8] Y. Hamasuna, Y. Endo, S. Miyamoto and Y. Hasegawa, "On Hard Clustering for Data with Tolerance", Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, Vol.20, No.3, pp. 388–398, 2008 (in Japanese).

[9] Hamasuna Y., Endo Y. and Miyamoto S., "On Tolerant Fuzzy c-Means", Journal of Advanced Computational Intelligence and Intelligent Informatics (JACIII), (to be published).

[10] Tikhonov, A. N. and V. Y. Arsenin, 'Solutions of Ill-Posed Problems', V. H. Winston, 1977.

[11] P. M. Williams, "Bayesian Regularization and Pruning Using a Laplace Prior", Neural Computation, Vol. 7, No.1, pp. 117–143, 1995.

[12] K. Tsuda and T. Kudo, "Clustering Graphs by Weighted Substructure Mining", Proc. of the 23rd international conference on Machine learning, pp. 953–960, 2006.

[13] R. Tibshirani, "Regression shrinkage and selection via the Lasso", Journal of the Royal Statistical Society, Series B, Vol. 58, No.1, pp. 267-288, 1996.

[14] J. Kazama and J. Tsuji, "Evaluation and extension of maximum entropy models with in equality constraints", In Proc. of EMNLP, pp. 137–144, 2003.

ISBN: 978-989-95079-6-8

A parametric approach to solve quadratic programming problems with fuzzy environment in the set of constraints

Carlos Cruz[1] Ricardo C. Silva[2] José L. Verdegay[1] Akebo Yamakami[2]

1.Department of Computer Science and Artificial Intelligence,
University of Granada, E-18071, Granada, Spain
Email: {carloscruz,verdegay}@decsai.ugr.es
2.Department of Telematics, School of Electrical and Computer Engineering,
University of Campinas, P.O.Box 6101, 13083-970, Campinas-SP, Brazil.
Email: {rcoelhos,akebo}@dt.fee.unicamp.br

Abstract— *Quadratic programming can be seen both as a general approach to linear programming and a special class of nonlinear programming. Moreover, Quadratic Programming problems are of utmost importance in a variety of relevant practical fields, such as, portfolio selection. This work presents and develops a novel fuzzy-sets-based method that solves a class of quadratic programming problems with vagueness in the set of constraints. As vagueness is natural and ever-present in real-life situations requiring solutions, it makes perfect sense to attempt to address them using fuzzy quadratic programming. This kind of problem modeling is being applied in an increasing variety of practical fields especially those with logistics problems. Some illustrative numerical examples illustrating the solution approach are solved and analyzed to show the efficiency of this proposed method.*

Keywords— Fuzzy sets, decision making, fuzzy mathematical programming, quadratic optimization.

1 Introduction

In the early sixties, based on the fact that classical logic does not reflect, to the extent that it should, the omnipresent imprecision in the real world, L. A. Zadeh proposed the Theory of Fuzzy Sets and Fuzzy Logic. Nowadays Fuzzy Logic, or rather Soft Computing, is employed with great success in the conception, design, construction and utilization of a wide range of products and systems whose functioning is directly based on how human beings reason. This is specifically patent in the case of optimization problems, and particularly in so called Mathematical Programming problems. Mathematical programming is an area which solves problems that involve minimization (or maximization) of the objective function in a function domain that can be constrained or not. In this area, Quadratic Programming represents a special class where the objective function is a quadratic function and the constraints are linear. This set of problems can be formalized in the following form:

$$\begin{aligned} \min \quad & \mathbf{c^t x} + \tfrac{1}{2}\mathbf{x^t Q x} \\ \text{s.t.} \quad & \mathbf{A x} \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned} \tag{1}$$

where $\mathbf{c}$ is an n vector, $\mathbf{b}$ is an m vector, $\mathbf{A}$ is an $m \times n$ matrix, and $\mathbf{Q}$ is an $n \times n$ symmetric matrix.

Thus, on the one hand, it is clear that quadratic programming encompasses all linear problems, including applications in scheduling, planning and flow computations, and they may be used to solve some interesting combinatorial optimization problems. On the other hand, quadratic programming is a particular kind of nonlinear programming. There are several classes of problems that are naturally expressed as quadratic problems. Examples of such problems can be found in game theory, engineering modeling, design and control, problems involving economies of scale, facility allocation and location problems, problems in microeconomics amongst others. Several applications and test problems for quadratic programming can be found in [5, 7, 8, 16, 17]. Some traditional methods are available in the literature [2, 24] for solving such problems. An interesting web page about quadratic programming is [6].

Among the several applications, we will present a portfolio selection problem which is an important research field in modern finance. This problem was first introduced by Markowitz [14, 15], and provided a risk investment analysis. Vagueness, approximate values and lack of precision in this problem are very frequent in that context, and quadratic programming problems have shown to be extremely useful in solving a variety of portfolio models. In any case, it is important to point out that the aim of this work is not to solve portfolio models. They are only considered here for the sake of illustrating the fuzzy quadratic programming problems solution approach presented, which in fact is the goal and main aim of this contribution. Some works about portfolio selection problem by using fuzzy approaches can be found in [9, 11, 19, 21, 22].

Moreover, there are some cases where the parameters of the real-world problems are seldom known exactly and have to be estimated by the decision maker. Therefore, the application of Soft Computing, and Fuzzy Logic in particular, has shown, in recent years, great potential for modeling systems which are non-linear, complex, ill-defined and not well understood. Fuzzy Logic is a way to describe this vagueness mathematically and it has found numerous and different applications due to its easy implementation, flexibility, tolerant nature to imprecise data, low cost implementations and ability to model non-linear behavior of arbitrary complexity because of its basis in terms of natural language. The uncertainties can be found in the relation, constants, decision variables or in all parameters of the problem. Some authors have applied Soft Computing methodologies to quadratic programming as can be seen in [1, 3, 4, 12, 13, 18, 20, 23, 25, 26, 27].

With this in mind, the goal of this paper is to present a novel approach that transforms a quadratic programming problem with uncertainties in the coefficients and order relation of the

ISBN: 978-989-95079-6-8

set of constraints into a parametric problem and to obtain a set of optimal solutions of this new problem that belong to the fuzzy solution.

Thus, Problem (1), which is a classic quadratic programming problem, can be rewritten in the following way:

$$\begin{aligned} \min \quad & \mathbf{c^t x} + \tfrac{1}{2}\mathbf{x^t Q x} \\ \text{s.t.} \quad & \tilde{\mathbf{A}}\mathbf{x} \leq^f \tilde{\mathbf{b}} \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned} \tag{2}$$

where $\mathbf{x}$ and $\mathbf{c}$ are $n \times 1$ vectores with real numbers and $\tilde{\mathbf{b}}$ is an $m \times 1$ vector with fuzzy numbers, $\mathbf{Q}$ is an $n \times n$ matrix with real numbers and $\tilde{\mathbf{A}}$ is an $m \times n$ matrix with fuzzy components, and the symbol "$\leq^f$" shall mean just that the decision-maker is willing to permit some violations in the accomplishment of the constraints. Both fuzzy numbers and these violations are measured by membership functions $\mu_i : \mathbb{R} \to [0,1],\ i = 1,\ldots,m$ that for each solution provide the decision maker's degree of satisfaction with the accomplishment of each constraint.

The paper is organized as follows: Section 2 demonstrates that a fuzzy quadratic programming problem can be transformed into a parametric quadratic problem and that the solution obtained by traditional techniques is an acceptable fuzzy solution. To illustrate the approach, Section 3 presents a general portfolio selection problem formulated as a fuzzy quadratic programming; Section 4 presents numerical simulations for the proposed problems and an analysis of the results obtained. Finally, in Section 5 some conclusions are pointed out.

2 A general model for fuzzy quadratic programming

In this section, a novel approach is presented to solve quadratic programming problems with uncertainties in the order relation and coefficients in the set of constraints. This approach transforms this fuzzy quadratic problems into a parametric quadratic programming problem. A quadratic programming problem with this fuzzy environment has been formulated in (2) and the following sub-section will show how the uncertainties can be dealt with.

2.1 Parametric ideas for solving a fuzzy quadratic programming problem

The fuzzy coefficients of the quadratic problem (2) are defined with fuzzy nature, that is, some violations in the accomplishment of such restriction functions are permitted. Therefore these fuzzy parameters can be determined by the decision maker.

It is clear that each membership function will give the degree of membership (satisfaction) such that any $\mathbf{x} \in \mathbb{R}^n$ accomplishes the corresponding fuzzy objective function and constraint upon which it is defined. These membership functions can be formulated as follows

$$\mu_i : \mathbb{R} \to (0,1], \quad i \in I$$

where μ is an linear membership function (and formally can also be a non linear one), and I is the set that contains all fuzzy parameters.

In order to solve this problem in a two-phase method, first let us define for each fuzzy constraint, $i \in I$

$$X_i = \left\{ \mathbf{x} \in \mathbb{R}^n \mid \tilde{A}_i \mathbf{x} \leq^f \tilde{b}_i, \mathbf{x} \geq 0 \right\}.$$

If $\mathbf{X} = \bigcap_{i \in I} X_i$ then the former fuzzy quadratic problem can be addressed in a compact form as

$$\min \{ f(\mathbf{x}) \mid \mathbf{x} \in \mathbf{X} \}$$

It is clear that $\forall \alpha, \gamma \in (0,1]$, an (α,γ)-cut of the fuzzy constraint set will be the classical set

$$X(\alpha,\gamma) = \{ \mathbf{x} \in \mathbb{R}^n \mid \mu_X(\mathbf{x}) \geq \min\{\alpha,\gamma\} \}$$

where $\forall \mathbf{x} \in \mathbb{R}^n$,

$$\mu_X(\mathbf{x}) = \inf \mu_i(\mathbf{x}), i \in I$$

Hence an (α,γ)-cut of the i-th constraint will be denoted by $X_i(\alpha,\gamma)$. Therefore, if $\forall \alpha, \gamma \in (0,1]$,

$$S(\alpha,\gamma) = \{ \mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) = \min\ f(\mathbf{y}),\ \mathbf{y} \in X(\alpha,\gamma) \}$$

the fuzzy solution to the problem will therefore be the fuzzy set defined by the following membership function

$$\mu_{S(\alpha,\gamma)}(\mathbf{x}) = \begin{cases} \sup\{\alpha,\gamma : \mathbf{x} \in S(\alpha,\gamma)\} & \mathbf{x} \in \bigcup_\alpha S(\alpha,\gamma) \\ 0 & \text{otherwise.} \end{cases}$$

Provided that $\forall \alpha \in (0,1]$,

$$X(\alpha,\gamma) = \bigcap_{i \in I} \{ \mathbf{x} \in \mathbb{R}^n \mid [(\tilde{A})_\alpha x]_i \leq r_i(\alpha,\gamma), \mathbf{x} \geq 0, \mathbf{x} \in \mathbb{R}^n \}$$

with $r_i(\alpha,\gamma) = [(\tilde{b})_\alpha]_i + [(\tilde{d})_\alpha]_i (1-\gamma)$.

Thus, the operative solution to the former problem can be found, (α,γ)-cut by (α,γ)-cut, by means of the following auxiliary parametric quadratic programming model,

$$\begin{aligned} \min \quad & \mathbf{c^t x} + \tfrac{1}{2}\mathbf{x^t Q x} \\ \text{s.t.} \quad & \tilde{\mathbf{A}}_\alpha \mathbf{x} \leq \tilde{\mathbf{b}}_\alpha + \tilde{\mathbf{d}}_\alpha (1-\gamma) \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned} \tag{3}$$

Therefore, the fuzzy quadratic programming problem was parameterized at the end of the first phase. In the second phase, the parametric quadratic programming problem is presented as a multiobjective approach which is solved for each of the different α and γ values using any multiobjective optimization technique. We must find efficient solutions to parametric problem for each α and γ that satisfy Karush-Kuhn-Tucker's necessary efficient optimality conditions.

2.1.1 Methods to solve quadratic programming problems with fuzzy relations

By using the parametric idea described above a general method to solve quadratic programming problems with fuzzy relations is presented here, as described in [18]. This method is an extension of the method that was developed to solve fuzzy linear programming problems;

$$\begin{aligned} \min \quad & \mathbf{c^t x} + \tfrac{1}{2}\mathbf{x^t Q x} \\ \text{s.t.} \quad & (Ax)_i \leq b_i + d_i(1-\alpha), i \in I \\ & x_j \geq 0, j \in J, \alpha \in (0,1]. \end{aligned} \tag{4}$$

where d_i is the violation permitted for each constraint i, $\mathbf{Q}$ and $\mathbf{A}$ are an $n \times n$ matrix and an $m \times n$ matrix with real numbers, and $\mathbf{c}$ and $\mathbf{b}$ are an n vector and m vector with real numbers.

ISBN: 978-989-95079-6-8

2.1.2 Methods to solve quadratic programming problems with fuzzy coefficients in the set of constraints

This multiobjective approach was developed in [10] to solve fuzzy nonlinear programming problems. This approach will be used to solve Problem (2) with fuzzy coefficients in the constraints set. The fuzzy solution is obtained by transforming a fuzzy nonlinear programming problem into a parametrical multiobjective nonlinear programming problem in which the parameters $\alpha, \gamma, \beta_{ij} \in [0,1],\ i = 1,\ldots,m, j = 1,\ldots,n+1$ can be treated as new decision variable.

The goal of this parametrical multiobjective problem is to minimize $f(\mathbf{x})$, maximize γ, and maximize and minimize $\alpha, \gamma,\ \beta_{ij},\ i = 1,\ldots,m,\ j = 1,\ldots,n+1$, simultaneously. Therefore, Problem (3) is transformed into a multiobjective quadratic programming problem that is stated as follows:

$$\begin{array}{ll} \min & \left[\mathbf{c^t x} + \frac{1}{2}\mathbf{x^t Q x}, \beta_{11}, 1-\beta_{11}, \ldots, \beta_{m,n+1}, 1-\beta_{m,n+1}\right] \\ \text{s.t.} & \sum_{j=1}^{n}\left((a_\alpha^L)_{ij} + \beta_{ij}\left((a_\alpha^U)_{ij} - (a_\alpha^L)_{ij}\right)\right)x_j \leq (b_\alpha^L + d_\alpha^L)_i + \\ & \quad + \beta_{i,n+1}\left((b_\alpha^U)_i - (b_\alpha^L)_i + \left((d_\alpha^U)_i - (d_\alpha^L)_i\right)(1-\gamma)\right) \\ & \mathbf{x} \geq \mathbf{0}, \alpha, \beta_{ij} \in [0,1],\ i = 1,\ldots,m,\ j = 1,\ldots,n+1 \end{array} \tag{5}$$

where it considers $m(n+3)$ new decision variables α, γ and $\beta_{ij},\ i = 1,\ldots,m,\ j = 1,\ldots,n+1$, to transform the intervals $I_{ij}(\alpha) = [h_{ij}^{-1}(1-\alpha), g_{ij}^{-1}(1-\alpha)]$ into functions of the form $z_{ij}(\alpha, \beta_{ij}) = h_{ij}^{-1}(1-\alpha) + \beta_{ij}(g_{ij}^{-1}(1-\alpha) - h_{ij}^{-1}(1-\alpha))$.

The obtained results for each α and γ value generate a set of solutions $S(\alpha, \gamma)$ and then the Representation Theorem can be used to integrate all these specific alpha-solutions and ending the second phase. Therefore the outlined solution to the parametric method is a valid solution to the fuzzy quadratic problem.

3 Portfolio selection problem

As said previously, in order to illustrate the above described two phases method for solving fuzzy quadratic programming problems, we now focus on general Portfolio Problems. It is important to point out that up to now we have not tried to improve other solution methods for this kind of important problems, but only to show how our solution approach performs. A description of a classical portfolio selection problem that was formulated by Markowitz as a quadratic programming problem is given in [15]. Assume that there are n securities denoted by S_j $(j = 1,\ldots,n)$, then this quadratic problem can be written in the following form:

$$\begin{array}{ll} \min & \mathbf{x^t \Sigma x} \\ \text{s.t.} & \mathbf{x^t E(R)} \geq \rho \\ & \mathbf{1x} = 1 \\ & \mathbf{x} \geq \mathbf{0} \end{array} \tag{6}$$

where $\mathbf{x}$ is an n vector that represents the percentage of money invested in assets, i.e., the proportion of total investment funds devoted to each security; $\mathbf{E(R)}$ is the average vector of returns over m periods because $\mathbf{R} = [r_{ij}]$ is an $m \times n$ matrix that represents the random variables of the returns of assets varying in m discrete times; ρ is a parameter representing the minimal rate of return required by an investor; and $\Sigma = [\sigma_{ij}^2]$ is a covariance $n \times n$ matrix between returns of asset which can be written as:

$$\sigma_{ij}^2 = \sum_{k=1}^{m} \frac{\left(r_{ki} - E(r_i)\right)\left(r_{kj} - E(r_j)\right)}{m-1}. \tag{7}$$

Therefore, the objective of Problem (6) is minimizing the risk variance and the investment diversification subject to a given average return ρ.

The expected return rate, ρ, is a decision maker's value that represents an expert's knowledge, then a fuzzy approach can be used in the constraint of the portfolio selection problem. Problem (6) with the use of fuzzy sets can be formulated in the following form:

$$\begin{array}{ll} \min & \mathbf{x^t \Sigma x} \\ \text{s.t.} & \mathbf{x^t E(R)} \geq^f \rho \\ & \mathbf{1x} = 1 \\ & \mathbf{x} \geq \mathbf{0}. \end{array} \tag{8}$$

Thus, Problem (8) can be rewritten as

$$\begin{array}{ll} \min & \mathbf{x^t \Sigma x} \\ \text{s.t.} & \mathbf{x^t E(R)} \geq \rho - d(1-\alpha) \\ & \mathbf{1x} = 1 \\ & \mathbf{x} \geq \mathbf{0}. \end{array} \tag{9}$$

4 Numerical experiments

The problems we use to evaluate this method are one hypothetic mathematical formulation and one fuzzy portfolio selection problem with the fuzzy approach described in Section 3. Nevertheless, they are efficient in validating the realized study. First, in subsection 4.1 we will show the formulation of the problems. Then in subsection 4.2 the computational results and a comparative analysis of the classic methods and the iterative methods responses will be presented.

The tests were all performed on a PC with two 2.26GHZ Intel® Core(™) 2 Duo processors, 4GB RAM running Ubuntu 8.10 operational system. All the problems presented in this work were resolved using **fminimax** function to solve constraint programming problems of *ToolBox Optimization* of MATLAB® 7.4.0 program.

4.1 Formulation of the numerical examples

We present some real-world and theoretical problems found in the literature with a view to validating the proposed algorithm. We simulate two quadratic programming problems both with uncertainties in the order relation and coefficients in the set of contraints.

Table 1 provides one theoretical quadratic problem that is described in [8]. The optimal solution to the problems without uncertainties is presented in the columns $\bar{\mathbf{x}}'$ and $f(\mathbf{x})$ of table 1.

In order to show the performance of our method, we use the set of historical data shown in Table 2 introduced by Markowitz. The columns 2-10 represent American Tobacco, A.T.&T., United States Steel, General Motors, Atcheson&Topeka&Santa Fe, Coca-Cola, Borden, Firestone and Sharon Steel securities data, respectively. The returns on the nine securities, during the years 1937-54, are presented in Table 2.

This example will consider performances of portfolios with respect to "return" thus defined. This assumes that a dollar of realized or unrealized capital gains is exactly equivalent to a

ISBN: 978-989-95079-6-8

dollar of dividends, no better and no worse. This assumption is appropriate for certain investors, for example, some types of tax-free institutions. Other ways of handling capital gains and dividends, which are appropriate for other investors, can be viewed in [15].By computing the average value of all the years of each column of random variables of Table 2, we obtained the expected values of each return of the securities that are described in (10). Then, using Equation (7), we computed the covariance matrix that is presented in (11).

4.2 Results and Analysis

Here we show the results obtained for the problem by the fuzzy quadratic programming method introduced in Section 2. The problem described in this work was solved by using linear membership functions as presented by Problem (3). Table 3 shows the solution of the theoretical problem described by Table 1, while the solution of the real-world portfolio selection problem, described by Table 2, is presented in Table 4.

Table 3 presents the results of the hypothetical problem with fuzzy coefficients and relation in the constraints set. Each row describes parametrical solutions to different α-cut levels of the fuzzy coefficients and each column represents parametrical solutions to different violations of the fuzzy relation in the set of constraints. Now, by applying the Representation Theorem in these parametrical solutions, we can define a fuzzy solution that describes a satisfactory solution of the fuzzy quadratic programming problem shown in Table (1).

Table 4 presents the results of the fuzzy portfolio selection problem to different α-cut levels of the fuzzy coefficients and violations of the fuzzy relation in the set of constraints. Thus, a fuzzy solution can be defined by applying the Representation Theorem in these parametrical solutions. This fuzzy solution describes a satisfactory solution of the fuzzy portfolio selection problem.

5 Conclusions

Fuzzy quadratic programming problems are of utmost importance in an increasing variety of practical fields because real-world applications inevitably involve some degree of uncertainty or imprecision, for example in logistics management. One of these problems is the portfolio selection problem, where the imperfect knowledge of the returns on the assets and the uncertainty involved in the behaviour of financial markets may be introduced by means of fuzzy quantities and/or fuzzy constraints. In this context this paper has presented an operative and novel method for solving Fuzzy Quadratic Programming problems which is carried out by performing two phases that finally provide the user with a fuzzy solution. The method has been validated by solving a number of practical problems. The solutions obtained aid the authors to follow along this line of research to try to solve real problems in practice, in such a way that oriented Decision Support Systems involving Fuzzy Quadratic Programming problems can be built.

A parametric approach that solve a fuzzy quadratic problem was proposed. The set of optimal solutions obtained by one parametric approach to each $\alpha, \beta, \gamma \in [0, 1]$ constructs the fuzzy solution by using the Representation Theorem. The authors aim is to extend the line of investigation involving Fuzzy Quadratic Programming problems in order to try to solve other practical problems.

Acknowledgments

The authors want to thank the support provided by the Brazilian agency CAPES and the Spanish projects TIN2005-08404-C04-01 (70% of which are FEDER funds) and MINAS (TIC-00129-JA).

References

[1] AMMAR, E., KHALIFA, H.A.: Fuzzy portfolio optimization a quadratic programming approach. Chaos, Solitons and Fractals **18**, 1045–1054 (2003)

[2] BEALE, E.: On quadratic programming. Naval Research Logistics Quarterly **6**, 227–244 (1959)

[3] BECTOR, C.R., CHANDRA, S.: Fuzzy mathematical programming and fuzzy matrix games, *Studies in Fuzziness and Soft Computing*, vol. 169. Springer, Berlin (2005)

[4] BELLMAN, R.E., ZADEH, L.A.: Decision-marking in a fuzzy environment. Management Science **17**(4), B141–B164 (1970)

[5] FLOUDAS, C.A., PARDALOS, P., ADJIMAN, C., ESPOSITO, W.R., GÜMÜS, Z.H., HARDING, S.T., KLEPEIS, J.L., MEYER, C.A., SSHWEIGER, C.A.: Handbook of test problems in local and global optimization, *Nonconvex Optimization and its Applications*, vol. 33. Kluwer Academic Publishers, Dordrecht (1999)

[6] GOULD, N., TOINT, P.: A quadratic programming page. www.numerical.rl.ac.uk/qp/qp.html

[7] GOULD, N.I.M., TOINT, P.L.: A quadratic programming bibliography. RAL Numerical Analysis Group Internal Report 2000-1, Department of Mathematics, University of Namur, Bruxelles, Belgium (2000)

[8] HOCK, W., SCHITTKOWSKI, K.: Test examples for nonlinear programming codes, *Lecture Notes in Economics and Mathematical Systems*, vol. 187. Spring-Verlag (1981)

[9] INUIGUCHI, M., RAMIK, J.: Possibilistic linear programming: A brief review of fuzzy mathematical programming and a comparison with stochastic programming in portfolio selection problem. Fuzzy Sets and Systems **111**(1), 3–28 (2000)

[10] JIMÉNEZ, F., CADENAS, J., SÁNCHEZ, G., GÓMEZ-SKARMETA, A., VERDEGAY, J.: Multi-objective evolutionary computation and fuzzy optimization. International Journal of Approximate Reasoning **43**(1), 59–75 (2006)

[11] LEON, T., LIERN, V., VERCHER, E.: Viability of infeasible portfolio selection problems: A fuzzy approach. European Journal of Operational Research **139**, 178–189 (2002)

[12] LIU, S.T.: Quadratic programming with fuzzy parameters: A membership function approach (2007). Chaos, Solitons and Fractals. Article in press. doi:10.1016/j.chaos.2007.07.054

[13] LIU, S.T.: Solving quadratic programming with fuzzy parameters based on extension principle. In: FUZZ-IEEE 2007, pp. 1–5. IEEE International Conference Fuzzy Systems, London, UK (2007)

[14] MARKOWITZ, H.: Portfolio selection. The Journal of Finance **7**(1), 77–91 (1952)

[15] MARKOWITZ, H.M.: Portfolio Selection: Efficient Diversification of Investments, second edn. Blackwell Publisher, Massachusetts (1991)

ISBN: 978-989-95079-6-8

Table 1: Fuzzy quadratic programming problem

$f(\mathbf{x})$	$\mathbf{x}_{initial}$	Constraints	Violation	Classic solution	
				$\overline{\mathbf{x}}^T$	$f(\mathbf{x})$
$9-8x_1-6x_3-4x_3+$ $+2x_1^2+2x_2^2+x_3^2+2x_1x_2+2x_1x_3$	$[0.5;0.5;0.5]^T$	$x_1+x_2-2x_3-3 \leq^f 0$ $x_i \geq^f 0, i=1,2,3$	$d_1=\widetilde{0.3}$	$[4/3;7/9;4/9]^T$	1/9

Table 2: Fuzzy portfolio selection problem

	#1	#2	#3	#4	#5	#6	#7	#8	#9
Year	Am.T	A.T&T.	U.S.S.	G.M.	A.T.&S.	C.C.	Bdm.	Frstn.	S.S.
1937	-0.305	-0.173	-0.318	-0.477	-0.457	-0.065	-0.319	-0.4	-0.435
1938	0.513	0.098	0.285	0.714	0.107	0.238	0.076	0.336	0.238
1939	0.055	0.2	-0.047	0.165	-0.424	-0.078	0.381	-0.093	-0.295
1940	-0.126	0.03	0.104	-0.043	-0.189	-0.077	-0.051	-0.09	-0.036
1941	-0.28	-0.183	-0.171	-0.277	0.637	-0.187	0.087	-0.194	-0.24
1942	-0.003	0.067	-0.039	0.476	0.865	0.156	0.262	1.113	0.126
1943	0.428	0.300	0.149	0.255	0.313	0.351	0.341	0.580	0.639
1944	0.192	0.103	0.260	0.290	0.637	0.233	0.227	0.473	0.282
1945	0.446	0.216	0.419	0.216	0.373	0.349	0.352	0.229	0.578
1946	-0.088	-0.046	-0.078	-0.272	-0.037	-0.209	0.153	-0.126	0.289
1947	-0.127	-0.071	0.169	0.144	0.026	0.355	-0.099	0.009	0.184
1948	-0.015	0.056	-0.035	0.107	0.153	-0.231	0.038	0	0.114
1949	0.305	0.030	0.133	0.321	0.067	0.246	0.273	0.223	-0.222
1950	-0.096	0.089	0.732	0.305	0.579	-0.248	0.091	0.650	0.327
1951	0.016	0.090	0.021	0.195	0.040	-0.064	0.054	-0.131	0.333
1952	0.128	0.083	0.131	0.390	0.434	0.079	0.109	0.175	0.062
1953	-0.010	0.035	0.006	-0.072	-0.027	0.067	0.21	-0.084	-0.048
1954	0.154	0.176	0.908	0.715	0.469	0.077	0.112	0.756	0.185

$$\mathbf{E}(\mathbf{R}) = \begin{bmatrix} 0.0659 & 0.0616 & 0.1461 & 0.1734 & 0.1981 & 0.0551 & 0.1276 & 0.1348 & 0.1156 \end{bmatrix} \tag{10}$$

$$\Sigma = \begin{pmatrix} 0.0565 & 0.0228 & 0.0303 & 0.0518 & 0.0172 & 0.0341 & 0.0257 & 0.0464 & 0.0383 \\ 0.0228 & 0.0155 & 0.0199 & 0.0259 & 0.0085 & 0.0106 & 0.0153 & 0.0265 & 0.0221 \\ 0.0303 & 0.0199 & 0.0905 & 0.0663 & 0.0470 & 0.0141 & 0.0111 & 0.0836 & 0.0445 \\ 0.0518 & 0.0259 & 0.0663 & 0.1011 & 0.0546 & 0.0307 & 0.0220 & 0.0775 & 0.0388 \\ 0.0172 & 0.0085 & 0.0470 & 0.0546 & 0.1354 & 0.0136 & 0.0221 & 0.0683 & 0.0476 \\ 0.0341 & 0.0106 & 0.0141 & 0.0307 & 0.0136 & 0.0437 & 0.0119 & 0.0254 & 0.0229 \\ 0.0257 & 0.0153 & 0.0111 & 0.0220 & 0.0221 & 0.0119 & 0.0305 & 0.0229 & 0.0184 \\ 0.0464 & 0.0265 & 0.0836 & 0.0775 & 0.0683 & 0.0254 & 0.0229 & 0.1024 & 0.0553 \\ 0.0383 & 0.0221 & 0.0445 & 0.0388 & 0.0476 & 0.0229 & 0.0184 & 0.0553 & 0.0839 \end{pmatrix} \tag{11}$$

ISBN: 978-989-95079-6-8

Table 3: Results of the first phase of the hypothetical problem.

γ/α	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.0
1.0	0.1111	0.1045	0.0982	0.0920	0.0860	0.0803	0.0747	0.0693	0.0642	0.0592	0.0544
0.9	0.1037	0.0973	0.0911	0.0851	0.0794	0.0738	0.0684	0.0632	0.0583	0.0535	0.0489
0.8	0.0958	0.0896	0.0836	0.0779	0.0723	0.0669	0.0618	0.0568	0.0521	0.0475	0.0432
0.7	0.0874	0.0814	0.0757	0.0701	0.0648	0.0597	0.0548	0.0501	0.0456	0.0413	0.0372
0.6	0.0784	0.0727	0.0672	0.0620	0.0569	0.0521	0.0474	0.0430	0.0388	0.0348	0.0311
0.5	0.0689	0.0635	0.0583	0.0534	0.0486	0.0441	0.0398	0.0357	0.0319	0.0282	0.0248
0.4	0.0588	0.0538	0.0490	0.0444	0.0400	0.0359	0.0320	0.0283	0.0248	0.0216	0.0186
0.3	0.0483	0.0437	0.0393	0.0351	0.0312	0.0275	0.0240	0.0208	0.0178	0.0150	0.0125
0.2	0.0374	0.0333	0.0294	0.0258	0.0224	0.0192	0.0163	0.0136	0.0112	0.0090	0.0070
0.1	0.0265	0.0230	0.0197	0.0167	0.0139	0.0114	0.0092	0.0072	0.0054	0.0039	0.0026
0.0	0.0160	0.0132	0.0108	0.0085	0.0066	0.0048	0.0034	0.0022	0.0013	0.0006	0.0002

Table 4: Results of the first phase of the portfolio selection problem.

γ/α	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.0
1.0	0.0343	0.0332	0.0322	0.0313	0.0305	0.0297	0.0290	0.0284	0.0278	0.0273	0.0268
0.9	0.0305	0.0298	0.0291	0.0286	0.0281	0.0275	0.0270	0.0266	0.0261	0.0257	0.0253
0.8	0.0282	0.0277	0.0272	0.0268	0.0263	0.0259	0.0255	0.0251	0.0247	0.0243	0.0239
0.7	0.0265	0.0261	0.0257	0.0253	0.0249	0.0245	0.0241	0.0238	0.0234	0.0230	0.0227
0.6	0.0266	0.0247	0.0243	0.0240	0.0240	0.0236	0.0233	0.0226	0.0223	0.0220	0.0216
0.5	0.0252	0.0249	0.0245	0.0245	0.0229	0.0234	0.0223	0.0216	0.0213	0.0210	0.0207
0.4	0.0240	0.0237	0.0234	0.0230	0.0227	0.0225	0.0222	0.0218	0.0215	0.0202	0.0209
0.3	0.0232	0.0228	0.0225	0.0222	0.0218	0.0215	0.0212	0.0209	0.0206	0.0203	0.0200
0.2	0.0220	0.0217	0.0215	0.0212	0.0209	0.0206	0.0204	0.0201	0.0198	0.0196	0.0193
0.1	0.0212	0.0209	0.0207	0.0204	0.0201	0.0199	0.0196	0.0194	0.0192	0.0189	0.0187
0.0	0.0204	0.0202	0.0199	0.0197	0.0195	0.0192	0.0190	0.0188	0.0186	0.0184	0.0182

[16] MAROS, I., MÉSZÁROS, C.: A repository of convex quadratic programming problems. Department Technical Report DOC 97/6, Department of Computing, Imperial College, London, UK (1997)

[17] SCHITTKOWSKI, K.: More test examples for nonlinear programming codes, *Lecture Notes in Economics and Mathematical Systems*, vol. 282. Spring-Verlag (1987)

[18] SILVA, R.C., VERDEGAY, J.L., YAMAKAMI, A.: Two-phase method to solve fuzzy quadratic programming problems. In: FUZZ-IEEE 2007, pp. 1–6. IEEE International Conference Fuzzy Systems, London, UK (2007)

[19] TANAKA, H., GUO, P., TÜRKSEN, B.: Portfolio selection based on fuzzy probabilities and possibity distributions. Fuzzy Sets and Systems **111**, 387–397 (2000)

[20] TANG, J., WANG, D.: An interactive approach based on a genetic algorithm for a type of quadratic programming problems with fuzzy objective and resources. Computers and Operations Research **24**(5), 413–422 (1997)

[21] VERCHER, E., BERMÚDEZ, J.D., SEGURA, J.V.: Fuzzy portfolio optimization under downside risk measures. Fuzzy Sets Syst. **158**(7), 769–782 (2007)

[22] WATADA, J.: Fuzzy portfolio selection and its applications to decision making. Tatra Mountains Mathematics Publication **13**, 219–248 (1997)

[23] WERNERS, B.: Interative multiple objective programming subject to flexible constraints. European Journal of Operational Research **31**, 342–349 (1987)

[24] WOLFE, P.: The simplex method for quadratic programming. Econometrica **27**, 382–398 (1959)

[25] XU, Z., DA, Q., CHEN, Q.: Priority approach based on quadratic programming model to fuzzy preference relation. Journal of Southeast University (English Edition) **21**(1), 108–110 (2005)

[26] ZIMMERMANN, H.J.: Fuzzy programming and linear programming with several objective functions. Fuzzy Sets and Systems **1**, 45–55 (1978)

[27] ZIMMERMANN, H.J.: Fuzzy mathematical programming. Computers & Operations Research **10**(4), 291–298 (1983)

ISBN: 978-989-95079-6-8

Interval-Valued Fuzzy System for Segmentation of Prostate Ultrasound Images

Miguel Pagola[1] Edurne Barrenechea[1] Aranzazu Jurio[1] Mikel Galar[1] Pedro Couto[2]
Farhang Sahba[3] Hamid R. Tizhoosh[4]
1.Automática y Computación, Universidad Pública de Navarra, Pamplona (Spain)
Email: miguel.pagola@unavarra.es
2.CITAB, Universidade Tras-os-Montes e Alto Douro, Vilareal, (Portugal)
Email: pcouto@utad.pt
3.University of Toronto, Toronto (Canada)
Email: farhang@ieee.org
4.Systems Design Engineering, University of Waterloo, Ontario (Canada)
Email: tizhoosh@uwaterloo.ca

Abstract— *In this paper we introduce an application of interval-valued systems to the segmentation of prostate ultrasound images. The system classifies each pixel as prostate or background. The input variables are the values of each pixel in different processed images as proximity, edginess and enhanced image. The system has 20 rules and is trained with ideal images segmented by an expert. Interval-valued fuzzy systems have been used due to their potential to capture uncertainty in a more robust way compared to ordinary fuzzy systems.*

Keywords— Interval-valued fuzzy sets, ultrasound image, image segmentation

1 Introduction

Ultrasound images are used in clinical settings to evaluate anomalies, tissues and organs. Ultrasound imaging is a quite common modality due to its low cost, portability and harmlessness to human body. Transrectal ultrasound images of the prostate in male patients are frequently used for both diagnosis and treatment purposes. Detecting the outline of the prostate – prostate segmentation – is therefore the first step regardless the following step is of diagnostic nature or treatment planning as for brachytherapy of prostate cancer.

Ultrasound image segmentation is strongly affected by the lower quality of these images. Speckle noise, shadows and prostate inhomogeneity make the segmentation a complicated task. There exist several methods to segment ultrasound images, a complete review can be found in [12]. Some methods use machine learning techniques to extract the objects/lesions. In those methods ideally segmented images created by an expert radiologist are used to train the system. In [18] reinforcement learning is used to train an agent devoted to segment prostate ultrasound images. Zhang et al. [20] optimize the weights of a weighed wavelet to detect microcalcifications in mamographic images. Neural networks were used in [13] to identify possible injuries in liver images and in [5] a genetic algorithm optimizes the weights of a self-organized neural network (Kohonen net).

One of the problems of the Neural Networks is the black box reasoning. In medical applications it is very useful to know how a certain task has been done the process of segmenting the image. Hence, we propose to use a fuzzy logic system to segment ultrasound images since they provide a convenient way of interpreting the tasks/results.

Due to the complexity of the ultrasound images we are going to use interval-valued fuzzy sets enabling us to represent the uncertainty that is within these images. In [9] Mendel proposes an adaptation of the fuzzy rule learning algorithm to interval type 2 fuzzy rules. Note what he called interval type 2 fuzzy sets are the same as interval-valued fuzzy sets in some cases (when $a = 1$, please see section 2). The objective of this paper is to develop an interval-valued fuzzy logic system (from now IT2FLS following the notation given in [9]) to segment prostates in transrectal ultrasound images.

Why to use interval-valued fuzzy systems instead of classical fuzzy systems? When we design a fuzzy system that is going to be trained using typical machine learning algorithms, we must choose the input variables, the output of our system and the number of rules. If we train the system with the past data, we can understand this process as function fitting problem, in which the fuzzy rule base system is a parametrized function and the training process is the modification of said parameters. This means that the output of the system fits the training data. If the system uses interval-valued fuzzy sets instead of classical fuzzy sets, the number of parameters to define an interval fuzzy rule is larger than the one needed for a classical fuzzy rule. So, in the training process, with the same data, the interval fuzzy system has more parameters, more degrees of freedom, which means it can be adjusted in a better way to the data. Some researchers suggest that this is an advantage compared to ordinary fuzzy systems. Therefore it means that for the same linguistic complexity (number of rules and number of variables) interval-valued fuzzy systems can, at least in theory, achieve better accuracy. We have made a comparative study to verify this hypotheses.

This work is organized as follows: first we present an introduction of interval-valued fuzzy logic systems. In section 3 we review Mendel's algorithm to generate an IT2FLS from training data. Later, in section 4 we present the model that we propose to segment ultrasound images. Finally we show some experimental results, conclusions and future research.

2 Interval-valued fuzzy systems

An interval-valued fuzzy set constitutes that the membership degree of every element to the set is given by a closed subinterval of interval [0,1]. The concept of type 2 fuzzy sets was introduced by Zadeh [19] as a generalization of an ordinary

ISBN: 978-989-95079-6-8

fuzzy set. The membership degree of an element to a type 2 fuzzy set is a fuzzy set in [0,1].

An interval type 2 fuzzy set $\overline{\overline{A}}$ in U is defined as

$$\overline{\overline{A}} = \{(u, A(u), \mu_u(x)) | u \in U, A(u) \in L([0,1])\},$$

where $A(u) = [\underline{A}(u), \overline{A}(u)]$ is a membership function; i.e., a closed subinterval is $[0,1]$, and function $\mu_u(x)$ represents the fuzzy set associated with the element $u \in U$ obtained when x is within $[0,1]$; $\mu_u(x)$ is given in the following way:

$$\mu_u(x) = \begin{cases} a \text{ if } \underline{A}(u) \leq x \leq \overline{A}(u) \\ 0 \text{ otherwise} \end{cases},$$

where $0 \leq a \leq 1$.

In [8]–[10], it is proved that an interval type 2 fuzzy set is the same as an interval-valued fuzzy set if $a = 1$.

Example: We can represent an interval-valued fuzzy set by means of an upper bound membership function and a lower bound membership function. In this work, all of the membership functions are going to be represented by a Gaussian function:

$$\mu_k^l(x_k) = exp\left[-\frac{1}{2}\left(\frac{x_k - m_k^l}{\sigma_k^l}\right)\right] \ \sigma_k^l \in [\sigma_{k1}^l, \sigma_{k2}^l] \quad (1)$$

In fig. 1 we show the membership function with parameter $m_k^l = 0.4$. We take $\sigma_{k1}^l = 0.05$ to represent the lower bound and $\sigma_{k2}^l = 0.1$ for the upper bound.

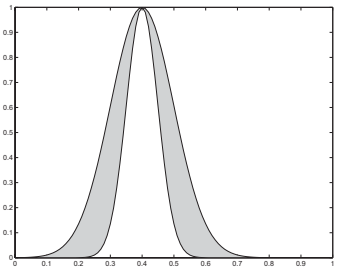


Figure 1: Interval-valued membership function with a Gaussian primary membership function and $m = 0.4$, $\sigma_{k1}^l = 0.05$ y $\sigma_{k2}^l = 0.1$.

An IT2FLS is a rule-based system in which the sets that represent the antecedents and consequents are interval-valued fuzzy sets (or interval type 2 fuzzy sets). In fig. 2 we depict the most important modules of an IT2FLS (see [9]).

The input of the system is a set of values. The fuzzfier transforms the inputs into interval-valued fuzzy sets. Then, the main part of the system, using the rules carries out the inference to generate conclusions, represented by interval-valued fuzzy sets. To use these conclusions (decisions) in the real world, the defuzzifier transforms these sets into crisp values. Commonly, before defuzzification, we can reduce the sets from interval-valued fuzzy sets to classical fuzzy sets, such a way typical defuzzifying techniques could be used to obtain crisp values.

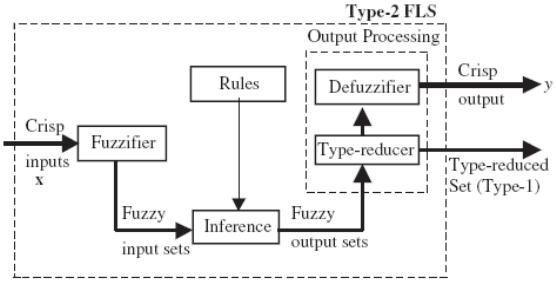


Figure 2: Interval Type 2 Fuzzy Logic System

Considering an IT2FLS with p inputs $x_1 \in X_1, \cdots, x_p \in X_p$ and one output $y \in Y$. Then we assume that the system has M rules in the following way:

$$R^l: \text{ IF } x_1 \text{ is } F_1^l \text{ AND } \ldots \text{ AND } x_p \text{ is } F_p^l, \text{ THEN } y \text{ is } G^l \text{ con } l = 1, \ldots, M \quad (2)$$

where $F_1^l, \ldots, F_p^l, G^l$ are interval-valued fuzzy sets. Each rule is interpreted as an implication:

$$R^l : F_1^l \times F_2^l, \cdots \times F_p^l \rightarrow G^l = A^l \rightarrow G^l \text{ with } l = 1, \ldots, M \quad (3)$$

R^l is described by the membership function $\mu_{R^l}(\mathbf{x}, y) = \mu(x_1, \ldots, x_p, y)$ where,

$$\mu_{R^l}(\mathbf{x}, y) = \mu_{A^l \rightarrow G^l}(\mathbf{x}, y) = \left[\bigcap_{i=1}^{p} \mu_{F_i^l}(x_i)\right] \bigcap \mu_{G^l}(y) \quad (4)$$

Using the extension of the Zadeh's compositional rule to interval-valued fuzzy sets, the consequent is calculated as follows:

$$\mu_{B^l}(y) = \mu_{A^l \circ R^l}(y) = \bigcup_{x \in X} \left[\mu_{A_x}(\mathbf{x}) \bigcap \mu_{R^l}(\mathbf{x}, y)\right] \text{ with } y \in Y, l = 1, \ldots, M \quad (5)$$

There exist several works regarding the inference with interval-valued fuzzy rules [2, 6], but in this work we are going to use the method proposed by Mendel in [9] (for a more detailed study see [10]).

3 Design of Interval-valued fuzzy systems from data

In this section we show how we can design an IT2FLS from training data.

Given a collection of N pairs of input-output data $(\mathbf{x}^{(1)}, y^{(1)}), (\mathbf{x}^{(2)}, y^{(2)}), \cdots (\mathbf{x}^{(N)}, y^{(N)})$ where $\mathbf{x}^{(t)}$ is the input vector and $y^{(t)}$ is the output value of the t training pair, we define the error of the IT2FLS for the t-th input as:

$$e^{(t)} = f_{s2}(\mathbf{x}^{(t)}) - y^{(t)} \quad t = 1, \ldots, N \quad (6)$$

where f_{s2} is the output of the IT2FLS. Such a value depends on the parameters that define the interval-valued fuzzy sets of the antecedents and consequents of the corresponding rules (m_k^l, σ_{k1}^l, σ_{k2}^l, y_l^j and y_r^j, with $k = 1, \ldots, p$, for p inputs, $j = 1, \ldots, M$, for M rules). Also y_l^j and y_r^j represent the bounds of the membership function of the consequent after

ISBN: 978-989-95079-6-8

the type reduction.
We define the quadratic error of the IT2FLS in t-th input as:

$$E^{(t)} = \frac{1}{2}[e^{(t)}]^2 \quad t = 1, \ldots, N \tag{7}$$

and the performance of the system as

$$E = \sum_{i=1}^{N} E^{(t)} \quad t = 1, \ldots, N \tag{8}$$

The training of an IT2FLS aims at finding the optimal values of the parameters that define the system such that equations (7) or (8) are minimized. If we minimize equation (7) (such a method is explained in [9]), the learning process is done in an incremental way, because the parameters are adjusted iteratively after processing every training pair. If we minimize equation (8), as we have done in this work, the learning process is done as a batch processing, because the parameters are adjusted after processing all of the training pairs. When we process all of the N training pairs and modify the parameters it is called a training epoch. In this work, in every epoch the parameters are adjusted proportional to the error gradient $\frac{\partial E}{\partial m_k^l}$, $\frac{\partial E}{\partial \sigma_{k1}^l}$, $\frac{\partial E}{\partial \sigma_{k2}^l}$, $\frac{\partial E}{\partial y_l^j}$ and $\frac{\partial E}{\partial y_r^j}$.

This quantity is calculated using the *resilient backpropagation* algorithm (RPROP, [14]). This algorithm automatically adjusts its own parameters during the training process, and is very easy to implement and can achieve a high convergence speed.

4 Proposed method

The objective of our method is to classify each pixel of the image. That is, decide if each pixel belongs to the important area (prostate) to segment or if it belongs to the background. Since the size and the position of the prostate are subject to change, the user must select the central point of the region (also called the seed point) to avoid false detections (and to focus on the segmentation itself). Therefore, the method is semi-supervised due to user suppling some relevant information. However, automation of this step is quite possible and has been repeatedly reported in literature.

The IT2FLS that we propose has 5 inputs, one output and consists of 20 rules. The inputs are:

1. pos_i: The distance, in pixels, of the pixel considered from the central point in the horizontal axes.
2. pos_j: The distance, in pixels, of the pixel considered from the central point in the vertical axes.
3. $dist$: The proximity of the pixel w.r.t. to the central point, calculated via a flooding algorithm.
4. edg: Edginess of each pixel.
5. mgr: Average gray level of the pixels neighborhood (e.g. 5×5 neighborhoods) in the enhanced image.

In the following subsections we present the techniques used to obtain the values of enhancement, proximity and edginess.

4.1 Enhanced image

The algorithm used to enhance ultrasound images is the one proposed by Sahba et al. [15, 16, 17], in which fuzzy rules such as the following have been used:

IF the pixel does not belong to the prostate,
THEN leave it unchanged
IF the pixel belongs to the prostate AND is dark,
THEN make it darker
IF the pixel belongs to the prostate AND is gray,
THEN make it dark
IF the pixel belongs to the prostate AND is bright,
THEN make it brighter

We use a simplified version of these rules in form of

IF the pixel *belongs* to the object AND is *dark*,
THEN make it *darker*,

or in a even more simple formulation and to save time we can use rules such as:

IF the pixel *belongs* to the object,
THEN make it *darker*,

where the degree of "belonging" of each pixel to the object is a function of its distance to the central point of the object or the inside of an initial/coares segment as proposed by Sahba et al. The main idea of enhancement is to eliminate the noise in the images and enhance the gray levels of selected area (regional contrast enhancement). First the noise is eliminated using a median filter (7×7 or 9×9). Then each pixel is fuzzyfied depending on its intensity with a membership function that is constructed taking into account the mean level of gray of the surroundings and the position of the selected point.

4.2 Proximity image

The proximity image represents the proximity of every pixel to the central point (similar to [7, 15, 16]), but taking into account the edges that separate the different regions of the original image. First we calculate the edges of the enhanced image using the Canny algorithm [4]. Then, starting from the central pixel selected by the user the algorithm labels the pixels with their distance to the central pixel step by step. In the first step the neighbors are labeled with distance 1 and so on (Fig. 3). The pixels marked as edge by the Canny edge detector are used as walls and cannot be labeled, so the proximity values generated are related with areas of the image.

4.3 Edginess

To create the edginess image, we calculate what is commonly called *false edges*. In [3] we presented a method to obtain false edges by means of t-norms and t-conorms. For every pixel a neighborhood matrix is constructed (3×3, 5×5, etc.). Applying t-norms to the elements of the matrix we obtain a lower bound of an interval. Applying t-conorms we obtain the upper bound of that interval. We called the length of the interval, that is the difference between the upper bound and the lower bound of the interval, false edge. The most known case is obtained when using minimum as the t-norm and maximum as the t-conorm.

ISBN: 978-989-95079-6-8

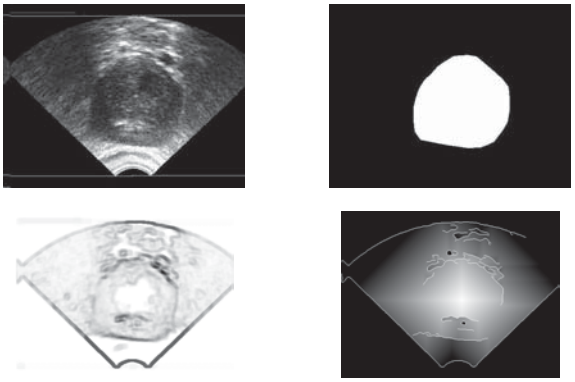


Figure 3: Top left to bottom right: Original image, ideally segmented image, edginess image, proximity image

5 Experimental results

To evaluate the performance of the IT2FLS we have a set of ten prostate ultrasound images with their corresponding ideal segmentation created by an expert. Each image has a central point of the prostate provided by the expert (for many prostate images automated detection of central point is relatively easy [1]). We use two of these images for the training and the other six for the validation. From the two training images we select randomly 200 pixels which are the training data pairs. This training data set is used to adjust the parameters of the system as described in section 3. The data is also split into two groups, the training set (80% of the data pairs) and the validation set (20%). To evaluate the performance of the IT2FLS we use an overlap measure S_A between the areas of the IT2FLS result and the ideally segmented image.

$$S_A = \frac{|Ideal \cap IBF|}{|Ideal \cup IBF|}, \tag{9}$$

where $Ideal$ is the binary image segmented by the expert, IBF is the binary image obtained using our method, $\cap$ and $\cup$ are the intersection and the union between crisp sets, respectively. Due to the initial values of the parameters of the system are selected randomly before training, we can obtain different possible solutions. In Table 1 we show the best ones after 20 trials. In Fig. 4 and 5 we show the binary images obtained by IT2FLS.

Table 1: Area overlap of segmented images with ideally segmented prostates.

Image	1	2	3	4	5	6	7	8	9	10
IVFS S_A(%)	84	61	74	75	77	75	62	64	71	72
Fuzzy S_A	86	67	77	78	77	80	71	72	70	73

Table 2: Percentage of convergence.

System	Percentage of convergence
Fuzzy	90 %
IVFS	30 %

In the experimental results we show that interval-valued fuzzy rule systems perform similar to classical fuzzy systems. The point is that the convergence of IVFS systems is really poor and also the mean performance achieved by these systems is a bit worse than ordinary fuzzy sets. It means that the extra adjustable parameters, if we dont use a specific learning algorithm with the IVFS system, are not worth in this case.

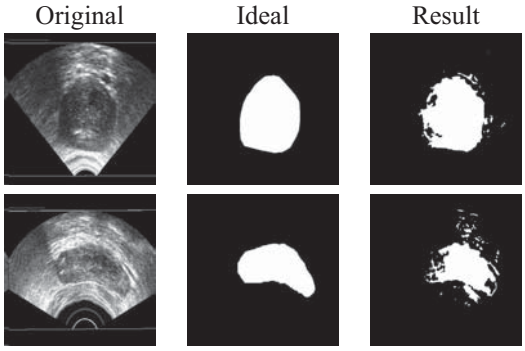


Figure 4: Comparison between ideally segmented images and IT2FLS segmented images used for training.

6 Conclusions and future research

We have proposed a new method to segment ultrasound images using an interval-valued fuzzy system. The system has 5 inputs, 20 rules and one output and can be trained using images previously segmented by an expert. An average overlap of 72% between the segmented area and the ideal segmentation has been reached. Nevertheless the results of the classical fuzzy systems are a bit better than the IVFSs system, mainly due to poor convergence of the learning algorithm.Taking into account the size of the proposed system on one hand and the challenging nature of prostate ultrasound segmentation on the other hand, the results can be regarded as promising.

The results show poor spatial consistency, which could be improved by adding some constraints or rules regarding shape and boundaries of the areas to the system. Also, we want to obtain interval input values to capture the uncertainty existing in the ultrasound images and deal with it via the IT2FLS in order to obtain better results.

Acknowledgment - This paper has been partially supported by the National Science Foundation of Spain, Reference TIN2007-65981

References

[1] J. Awad, Prostate Segmentation and Regions of Interest Detection in Transrectal Ultrasound Images. PhD Thesis, University of Waterloo, Canada, 2007.

[2] H. Bustince, “Indicator of inclusion grade for interval-valued fuzzy sets. Application to approximate reasoning based on interval-valued fuzzy sets”, Int. Journal of Approximate Reasoning, vol. 23, pp. 137–209, 2000.

[3] H. Bustince, E. Barrenechea, M. Pagola, R. Orduna, Construction of interval type 2 fuzzy images to represent images in grayscale. False edges. FUZZY IEEE 2007.

[4] J. Canny, A Computational approach to edge detection, IEEE Transactions on Pattern Recognition and Machine Intelligence, 8(6), 679-698 (1986).

ISBN: 978-989-95079-6-8

[5] Z. Dokur and T. Olmez, Tissue segmentation in ultrasound images by using genetic algorithms, *Expert Systems with Applications*, 34(4), 2739-2746 (2008)

[6] M.B. Gorzalczany, "A method of inference in approximate reasoning based on interval-valued fuzzy sets", *Fuzzy Sets and Systems*, vol. 21, pp. 1–17, 1987.

[7] M. Mancas, B. Gosselin, B. Macq, Segmentation Using a Region Growing Thresholding. In "Image Processing: Algorithms and Systems IV". Edited by Dougherty, Edward R.; Astola, Jaakko T.; Egiazarian, Karen O. Proceedings of the SPIE, Volume 5672, pp. 388-398, 2005.

[8] J.M. Mendel, R. I. Bob John. Type-2 fuzzy sets made simple, *IEEE Transactions on Fuzzy Systems*, **10**(2), pages 117–127, 2002.

[9] J.M. Mendel. Uncertain Rule-Based Fuzzy Logic Systems, *Prentice-Hall, Upper Saddle River*, NJ, 2001.

[10] J.M. Mendel. Advances in type-2 fuzzy sets and systems, *Information Sciences*, Volume 177(1), pp. 84–110, 2007.

[11] S. Mitra, Y. Hayashi, Neuro-fuzzy rule generation: survey in soft computing framework, IEEE Transactions on Neural Networks, 11(3), 748 - 768 May (2000).

[12] J.A. Noble and D. Boukerroui, Ultrasound image segmentation: A survey, IEEE Transactions on Medical Imaging 25 (2006), pp. 987–1010.

[13] K. Ogawa, M. Fukushima, K. Kubota and N. Hisa, Computer-aided diagnostic system for diffuse liver-diseases with ultrasonography by neural networks, IEEE Trans. on Nuclear Science 45(6) 3069-3074, 1998.

[14] M. Riedmiller, H. Braun. A direct adaptive method for faster backpropagation learning: The RPROP algorithm. *Proceedings of the IEEE International Conference on Neural Networks*, San Francisco,1993

[15] F. Sahba, H.R. Tizhoosh, M.M. Salama, A coarse-to-fine approach to prostate boundary segmentation in ultrasound images, BioMedical Eng. OnLine 2005, 4:58

[16] F. Sahba, H.R. Tizhoosh, M.M.A. Salama, Segmentation of Prostate Boundaries Using Regional Contrast Enhancement. The IEEE International Conference on Image Processing (ICIP), Volume 2, pp. 1266-9, 2005.

[17] H.R. Tizhoosh, Observer-Dependent Image Enhancement, Proceeding of FUZZ-IEEE 2001.

[18] H.R. Tizhoosh. Image thresholding using type II fuzzy sets, *Pattern recognition*, 38(12), pp. 2363-2372, 2005.

[19] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning – I, *Information Science*, **8**, 199–249, 1975.

[20] W. Zhang, H. Yoshida, R.M. Nishikawa, K. Doi, Optimally weighted wavelet transform based on supervised training for detection of microcalcs in digital mammograms, Medical Physics 25(6), 1998, pp. 949–956.

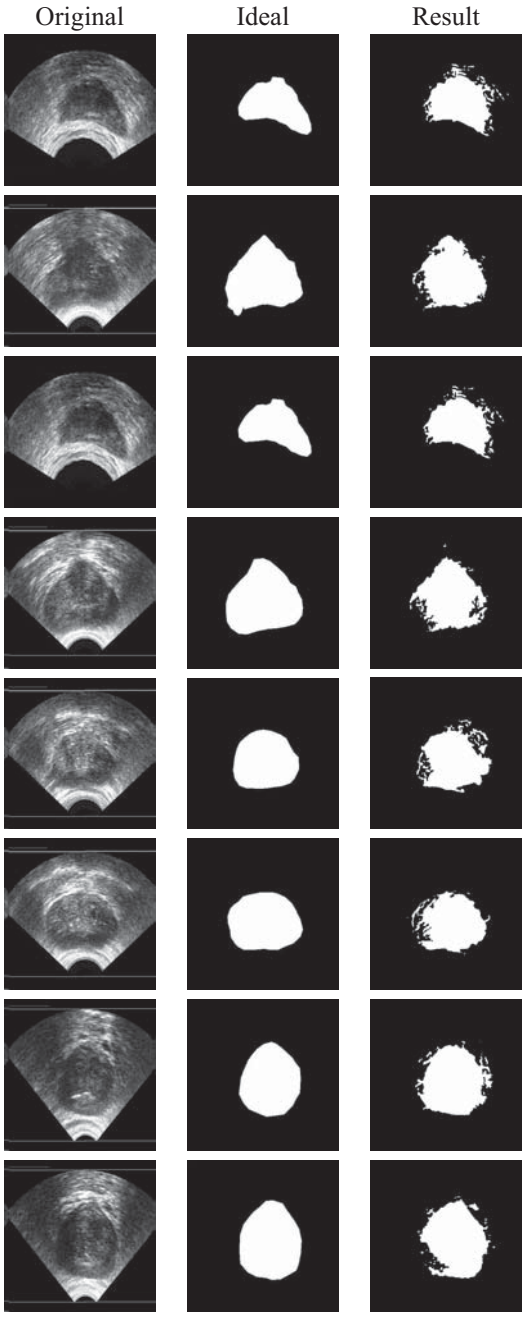


Figure 5: Comparison between ideally segmented images and IT2FLS segmented images. These are test images.

ISBN: 978-989-95079-6-8

A toward framework for generic uncertainty management

Ernesto Damiani Paolo Ceravolo Marcello Leida

Università degli Studi di Milano, Dipartimento di Tecnologie dell'Informazione
via Bramante 65, 26013 Crema (CR), Italy
Email: {ernesto.damiani, paolo.ceravolo, marcello.leida}@unimi.it

Abstract*— *The need for an automatic inference process able to deal with information coming from unreliable sources is becoming a relevant issue both on corporate networks and on the open Web. Mathematical theories to reason with uncertain information have been successfully applied in several situations, but each one of these models is tailored to deal with a specific semantics of uncertainty. In this paper, we put forward the idea of using explicit representations of the different types of uncertainty for partitioning the inference process into parts. By coordinating multiple independent reasoning processes, we are sometimes able to apply a specific model to each type of uncertain information, and recombine the final results via a suitable reconciliation process. We validated our approach applying it to the classic* ***schema matching problem*, and using the* ***Ontology Alignment Evaluation Initiative****, (OAEI) tests to assess the results.*

Keywords— Uncertainty, Ontology Matching, Reasoning, Rules.

1 Introduction

The problem of automatic inference is one of the most challenging problem in computer science [1], becoming even harder when knowledge is uncertain, due to lack of reliability of the source of information, approximation, dependencies and other factors. While many mathematical models for reasoning on uncertain information have been proposed, the general problem of handling and interpretation of uncertain knowledge is still to be solved. In this paper, we put forward the idea of using explicit representations of the different uncertainties present in the knowledge base according to different uncertainty models, coordinating multiple independent reasoning processes. By splitting the inference process into parts, we are able to apply a specific model to each type of uncertain information, recombining the final results via a suitable reconciliation process. Although the interoperability among multiple inference models been studied [2, 3], in the literature we are not aware of any hybrid reasoning processes which can handle the flexible integration of different models. As a proof of concept of this approach we present a semantics-aware matching strategy, that we apply to the well-known problem of ontology alignment [4]. The paper is structured as follows: Section 2 introduces the problem of uncertain information in knowledge management, briefly presenting the different types of uncertainty and the mathematical models used for the inference process. The section also introduces the need for an explicit representation of the various types of uncertainty, referring to the Ontology of Uncertainty [5] proposed by W3C's UR3W-XG incubator group[1]. Section 3 presents a case study, applying our technique to the classic *schema matching problem*, testing it via the Ontology Alignment Evaluation Initiative (OAEI) tests and comparing the results to the participants to the OAEI 2007 contest [4]. Conclusions and future work on our framework are presented in Section 5.

[1] http://www.w3.org/2005/Incubator/urw3/

2 Uncertain Information Representation and Reasoning

Experience has shown that the open Web and other platforms for hosting user-generated content can provide little quality control at content production time. As a result, most publicly available information can be considered *uncertain* to some degree. In order to clarify the notion of uncertainty, it is important to distinguish between **degrees of truth** and **degrees of uncertainty** in the information [6]. A degree of truth can be defined as the degree of compatibility between a statement and a knowledge base, which is limited to what the system knows about reality: a statement S is true if this assumption agrees with the set of statements in the knowledge base. Instead, Uncertainty of a statement arises when the knowledge base does not provide sufficient information to decide if a statement is true or false. Therefore uncertainty falls at a meta-level with respect to truth [7]. In case of truth values, we briefly mention two major theories [7]: *Classical two-valued logic* and *Fuzzy Logic*. In the first case, a statement's truth value can only assume one of two values [8], namely `0` or `1`. In the second case [9], truth values belong to the entire interval $[0, 1]$. For uncertainty representation, we distinguish between *Probability* and *Possibility theory*. The degree of probability associated to a statement is a typical example of gradual uncertainty. In sentences like: *"The player tossing a coin* wins *with 50% probability"*, "The player *wins*" is a true statement: it cannot happen that a player "half wins". The "50%" at the end of the sentence is not the statement truth value, but its level of uncertainty. *Possibility theory* is an alternative to probability theory, which separates the uncertainty of statements in possibilities and necessities [7]. Uncertainty can be classified as **Epistemic**, if it comes from the limited knowledge of the agent that generates the assertion or **Aleatory** if it is intrinsic in the observed world. Depending on the features of the agent that generates uncertain statements, is possible to identify two different types of uncertainty: **Objective** if the uncertainty derives from a repeatable observation and **Subjective** if the uncertainty in the information is derived from an informal evaluation. Furthermore, uncertainty can depend on the type of statement it is associated to: **Ambiguous**, **Inconsistent**,**Vague**, **Incomplete** and **Empiric**.

For our purposes, uncertainty can be represented as an annotation about a statement, expressing the level of certainty about it. We shall call "uncertain information" the triple

ISBN: 978-989-95079-6-8

(S, t, l) formed by a statement S, its truth value t and its corresponding uncertainty level l. It is important to remark that in many practical scenarios, we may encounter statements whose uncertainty levels have diverse semantics, especially when uncertain information is generated by multiple unsupervised processes. For example, Web-based weather forecasting services provide uncertain information in different forms; the uncertainty can be possibilistic or probabilistic (*Cloudy:"50%"*, *Rain:"10%"*) while truth value ranges can be classical (*Rainy*, *Fair*) or fuzzy (*Partially Covered*, *Heavy Rain*). Traditional approaches to uncertain reasoning support extensions of logic models dealing with inference on statements and their truth values, including mathematical theories able to deal with the uncertainty levels. In many cases, however, handling uncertainty hs an impact on the complexity and even the decidability of the inference problem. Scientific areas very active in the integration of classical logic with mathematical theories dealing with uncertainty is Artificial Intelligence and Knowledge Representation. Recently the effort was concentrated on the languages for the Web, such as in particular Semantic Web standards. For example, several probabilistic extensions of *Description Logics* (DL) [10] and *First Order Logic* (FOL) are available. Here, we shall focus the discussion on Description Logics, which are the logical model underpinning the OWL-Lite and OWL-DL ontology languages [11] used on the Semantic Web [12]. OWL-Lite and OWL-DL correspond respectively to $\mathcal{SHIF}(\mathbf{D})$ and $\mathcal{SHOIN}(\mathbf{D})$ Description Logics respectively, which are known to be tractable. A sound and complete fragment of $\mathcal{SROIQ}(\mathbf{D})$ is $\mathcal{SHOQ}(\mathbf{D})$ [13]. This fragment can be extended to handle uncertainty; in [14] Lukasiewicz defines a Probabilistic Description Logic $\mathcal{P} - \mathcal{SHOQ}(\mathbf{D})$. Based on Lukasiewicz's Probabilistic Description Logics, Klinov in [15] has implemented the probabilistic reasoner Pronto, that can reason with ontologies where a probability interval is assigned to statements, specifying the probability that a certain statement is true. As far as truth-values are concerned, the work of Straccia and Bobillo [16] extends the classical two-valued Description Logics to fuzzy sets. Namely, the authors present an extension of $\mathcal{SHIF}(\mathbf{D})$ Description Logics to the fuzzy case, dealing with different definitions for the logic operators (Zadeh logic, Lukasiewicz logic and Classical Logic); this approach, moreover, provides support to backward reasoning in case of Classical Logic semantics.

At first sight, one might hope that uncertainty representations and truth value ranges can be freely mixed according to the characteristics of the problem at hand. On the Web, vague information is usually modelled using fuzzy truth-values, knowledge uncertainty due to incomplete or defective observations is represented by Probability theory, and uncertainty arising from common sense knowledge and guessing can be handled with Possibility theory [17].

Unfortunately, the problem of dealing simultaneously with probability-based uncertainty and fuzzy truth values has been widely treated in literature but, as stated in [7] probability and possibility theories are not fully compositional with respect to all the logical connectives, without a relevant loss of expressiveness. This consideration leads to the consequence that uncertain calculi and degrees of truth are not fully compositional either. Nevertheless, some work in this direction has been proposed, by imposing restrictions to the expressiveness of the logics. The most relevant studies are: [18, 19] where the authors define *Probabilistic Description Logics Programs* (PDLP) by combining stratified fuzzy Description Logics programs with respect to degrees of probabilities in a unified framework. In [20] a definition of possibilistic fuzzy Description Logics has been proposed by associating weights, representing degrees of uncertainty, to the fuzzy Description Logic formulas. An extension of the fuzzy Description Logics in the field of Possibility theory has been presented also in [21] by annotating logic axioms with possibilities and necessity measures; by extending the approach presented in [20].

It is also important to underline that different approaches can be used to tackle the same type of uncertainty (i.e. in case of incomplete information is possible to use Possibilistic or Probabilistic theory) the choice of the best theory depends on the context. For instance, in [22] the authors use Dempster-Shafer's beliefs theory to resolve inconsistencies.

2.1 Ontology of Uncertainty

As mentioned in Section 2 uncertainty is generated from different situations and represented under different semantics. it is possible to create a classification of assertions based on several criteria: *nature*, *derivation*, *temporal validity* and *type*. Nature of uncertainty can be divided in *epistemic* and *aleatory*; *objective* and *subjective*; based on the temporal validity of a statement that can be valid for a period of time or can be valid always. A statement is *contingent* if refers to a particular situation or instant; in the second case a statement refers to situations that summarize trends (e.g. laws of physics, common sense knowledge, statistical knowledge) and is classified as *generic* statement. The nature of uncertainty, moreover, also depends on the statement it is attached to. Statements can be *ambiguous* in case the statement can be represented in different worlds with more than one interpretation, *inconsistent* if there is no possible world where the statement can hold, *vague*, and *incomplete* in case the knowledge about the observed world do not provide enough information to take a decision. Finally, a statement is *empiric* when is satisfied at least in one world. A first effort toward capturing all aspects of uncertainty is the *Ontology of Uncertainty*, published by the UR3W-XG incubator group[5]. This ontology tries to capture the nature, type and source of uncertainty that are specific of an assertion and allows moreover to relate the assertion to the correct computational model of inference. Statements, in the Ontology of Uncertainty, are represented by the concept `Sentence`, that provides information about the source (`Agent`) of the assertion, the subject (`World`) of the assertion and the semantics information about the `Uncertainty` model related to the assertions. There is not much to say about the two concepts `Agent` and `World`, which respectively represent the producer and the subject of an assertion. More interesting is the case of concept `Uncertainty`, the central concept of the ontology. This concept is related to all the various elements used to classify a `Sentence` under precise semantics. Other concepts related to `Uncertainty` are then used to describe type, derivation, validity and nature of the statement[2]. The ontology provides a generic meta-model rep-

[2]Currently, the Ontology of Uncertainty defined by the URW3-XG does not include a concept of `Validity` related to the temporal

ISBN: 978-989-95079-6-8

resented in OWL-DL [11] for representing uncertainty associated to various assertions and provides some use case scenarios, where, according to the semantics of the uncertainty, the correct inference model is selected. Unfortunately, the document [5] produced by the URW3-XG incubator group does not specify how to deal with situations where more than one model is involved in the inference process.

2.2 Using the Ontology of Uncertainty to support Reasoning

The Ontology of Uncertainty provides information on which mathematical model of uncertainty can be employed for managing a specific set of statements. When more than one model is involved in the reasoning process, the problem of integrating the results of multiple inference processes arises : if the subsets of statements handled by each model are disjoint, i.e. inferences are independent form each other, however, there are no particular problems in re-conciliating the results of the various reasoning processes;s. Some work in this direction has been carried out by the *Rule Interchange Format* (RIF[3]). In [2] the authors propose a framework for sharing information between three different models of uncertainty, where the fuzzy linguistic truth values are propagated through the three models in a non-monotonic way, by exploiting the extension principle [23] and aggregation of linguistic values. This approach is promising but is grounded to fixed fuzzy values (linguistic truth) that are used by all the different models and then aggregated according to non-monotonic rules.

In our approach, instead, we make use of the Ontology of Uncertainty as a way to model different types of uncertainty in a unified framework. The inference process involves three different steps: the first step is to partition the knowledge base in subsets according to the specific model, the second step is to carry out independent inferences, and the third one aggregates the results of the independent inference processes, following the *First Inference Then Aggregation* (FITA) approach [24], which also supports parallel reasoning. In our strategy the various reasoning processes are independent; we use the Ontology of Uncertainty classification to divide the various matching relations according to the uncertainty model to be used for the reasoning process. This way, the reasoners can be modelled as parallel processes. Directives on how to divide the heterogeneous knowledge base and how to recombine the results of the different reasoning processes are explicitly specified as DL-Safe Horn rules [25].

The knowledge base containing the information for our matching strategy is composed by a set of statements, generated independently by different sources, seen as instances of the concept `Sentence` in the Ontology of Uncertainty. To each statement S, information about uncertainty is associated by instantiating the ontology concept `Uncertainty` that defines the correct semantics. DL-Safe rules and SPARQL queries are largely involved in this process. A first set of DL-Safe rules is used to associate the statements to the correct type of uncertainty, and a set of SPARQL queries is used to divide the knowledge base according to the reasoning model associated to the statement. Finally a third set of rules is used to aggregate the results of the various reasoners. Currently, our approach requires the manual definition of rule sets. However, while the first set of rules is largely application dependent, the set of SPARQL queries and the third set of rules can be reused in many applications[4]. A detailed application of our approach is presented in Section 3.

3 A Semantics-aware Matching Strategy

Schema Matching is the time-honored problem of identifying the relations between the entities of two data source schemata. In case these schemata are represented as ontologies, this problem is also known as Ontology Alignment. In the literature different matching operators for a wide range of situations are available: the most exhaustive survey is [26].

Recent proposals tackle the schema matching problem by considering more than one matching operator at once and combining the final results through a matching strategy [27]. A **Matching Strategy** can be defined as *the process of transformation from a set of Matching Relations* Mr *to a new set* Mr', while a **Matching Operator** can be defined as *a function that takes as input two schema elements and creates as output a Matching Relation between the two elements*.

So far, even if some logic-based approaches are available [28, 29], most strategies proposed in literature neither consider explicitly the semantics of the various matching operators nor the different meanings of the relations that they generate. Instead, our matching strategy explicitly models the semantics of different matching operators as a Description Logic. In our approach, matching relations are stored as instances of a domain ontology, describing our application scenario. This domain ontology is extended with the *Ontology of Uncertainty* [5], which is used to associate the respective uncertainty to the different types of assertions. The *classification* of the various uncertainty types is performed by applying SWRL rules [30, 25] to the knowledge base. Once the classification is performed, the knowledge base is divided (by a *splitting* process) according to the specific uncertainty model. Each one of this inference models is used to perform a classification of the various relations in order to discover the most reliable ones. Finally a *reconciliation* process aggregates the best results[5].

3.1 The Matching Ontology

A matching relation mr_i is a $1:1$ relation associating two elements (concepts or attributes) of the two ontologies to align by a relation r from a collection of set theory operators ($\equiv$,$\subset$,$\supset$,$\cap$,$\neq$) and a degree of matching δ represented as:

$$mr_i = \langle e_h, e_k, r, \delta \rangle \tag{1}$$

The relation r between the elements depends on the particular feature that is analysed by the matching operator: as an example, in case of *JaroWinkler* matching operator, the features are composed by the label of the elements e_h and e_k to

validity of an assertion, but its addition is straightforward.

[3]The mission of the Rule Interchange Format (RIF) Working Group is to produce W3C Recommendations for rules interchange. `http://www.w3.org/2005/rules/wiki/RIF_Working_Group`

[4]Although they may require some fine tuning. This especially true of the third set, which is closely related to the aggregation procedure.

[5]The *classification*, *splitting* and *reconciliation* process in our system are rule driven, but this is not a strict requirement: this phases of the strategy can also be inferred from a Description Logics reasoning process used to classify various instances.

ISBN: 978-989-95079-6-8

match, while in case of an *Instance-based* matching operator the features are composed by the instances related to the elements to match.

We generally consider a matching operator as a process that generates a matching relation mr_i represented by the relation r between the elements e_h, e_k in input, associated with the strength δ of the relation. A matching operator is defined by a function f used to extract a particular feature that is analysed by the operator, a function θ which generates the relation r and a function ϕ that is in charge of the creation of the value δ, which represents the *strength* of the relation:

$$mo = \langle f, \theta, \phi \rangle \qquad (2)$$
$$mo(e_h, e_k) \rightarrow \langle e_h, e_k, r, \delta \rangle$$

Obviously, different matching operators carry different semantics that needs to be considered in the matching generation process. Also, the same relation can be generated by different matching operators; but in this case, the syntax of the relations is identical while the semantics is different. Consequently, different theories can be adopted, in order to infer the most suitable matching relations: the relations generated by the data type based matching operator can be considered as *possibilities*, while the relations generated by the instance based and string based matching operator can be modelled as *necessities* [6]. As we have seen, the Matching Ontology models the information created by the various matching operators, as well as information describing each specific matching operator. The attributes, relations and concepts from the ontologies to align are stored in the Matching Ontology as instances of the concept `Element`, represented by a unique identifier (e.g. URI) which is used by the system to retrieve the element in the original ontologies. The concept `MatchingOperator`, in the Matching Ontology, represents a generic matching operator; the specific matching operators (e.g. JaroWinkler, WordNet, Instances) are modelled as subclasses of this concept. The various matching relations are defined as instances of the concept `MatchingRelation`, which is composed by the concept `RelatedElements` that has a `subject` and an `object` that represent the two elements that have been related by one or more matching operators; and by a `confidence` value representing the strength of the matching relation. Subclasses of the concept `Matching Relation` are created in order to specialize the matching relation according to the specific relation (e.g. `Equivalent`, `Subset`, `Superset`, `Disjoint`, `Intersection`).

Instances of the Matching Ontology are generated during the matching process. The various matching operators generate a set of matching relations in the form $\langle e_h, e_k, r, \delta \rangle$; these relations are stored in the Matching Ontology with a reference to the instance of the matching operator that has generated the relation. When the Matching Ontology is generated in this way, some information is certain (e.g. the information provided by the original ontologies), and some other information in uncertain under different semantics (e.g. the matching relations that can disagree or that may have a probability degree to consider). This consideration motivates the use of additional annotations able to model the semantics of uncertainty of the various matching relations.

3.2 Managing uncertainty in the matching strategy

In a scenario like the one described in Section 3.1, we use the Ontology of Uncertainty to identify situations where is important to explicitly describe the type of uncertainty related to an assertion. First of all, the two ontologies (Matching Ontology and Ontology of Uncertainty) have to be linked somehow. Assertions in the Ontology of Uncertainty are represented by the concept `Sentence`. A sentence is `saidBy` an `Agent`, which we identify with the matching operator; moreover a sentence have also an object of the assertion (`saidAbout`) which is represented by the concept `World`. The two concepts `Agent` and `World` are then the linking point between the two ontologies; the link is defined by declaring `MatchingRelation` and `MatchingOperator` as sub-concepts of `World` and `Agent` respectively. Here, the Ontology of Uncertainty is used basically to drive the reasoning process: each type of uncertainty is processed by its specific reasoner and a final process, based on SWRL rules, integrates the results of the various reasoners. The application flow of the Matching Strategy operates as follows: a process takes as input the uncertain knowledge base generated by the matching operators, afterwards it divides the assertions according to their uncertainty and each sub part of the ontology is processed by its specific reasoner: in the system we consider a Probabilistic Description Logic reasoner, such as *Pronto* [15], a fuzzy Description Logic reasoner such as *FuzzyDL* [16] and a Defeasible Logic Reasoner such as *DR-Prolog* [3], but other models can be easily added. In our scenario the sources of information to be analysed, according to different uncertainty models are independent and no intersections among them has to be managed. This particular case allows a straightforward use of the Ontology of Uncertainty to drive the reasoning process, although in general the assumption of independence among the source of information is a lucky case.

The first part of the matching strategy is to assign to the various assertions (`Sentence`), the correct information about their uncertainty semantics. This information is classified according to a set of pre-defined SWRL rules that assigns the correct semantics in relation to several factors. The assignment is based on: (i) the `Agent` that has generated the relation; as an example: some agents can generate objective or subjective assertions: we can identify objective statements as necessities and subjective as possibilities; (ii) the presence of a degree of probability: as an example, sentences with a degree of probability can be handled with probabilistic theory models or with possibility theories. In this case is important to identify which statements need to be modelled with probability theory and which ones need to be modelled with possibility theory. (iii) the level of inconsistency among matching relations: as an example, if a sentence asserts that two elements are equivalent and another matching relation asserts that they are disjoint; (iv) the trustiness level of the matching operator: some operators are more reliable than others; (v) the level of detail of the assertion: the assertion created by a Data Type matching operator is more vague then an assertion created by a Regular Expression matching operator; (vi) the Data Type

[6] Note that this does not have a direct implication on the mappings to be created by the systems; for instance, it is clear that a necessary matching of strings does not imply a necessary mapping.

ISBN: 978-989-95079-6-8

of the elements to match can be used to establish priorities between operators.

$$\begin{aligned} uncertainty : Sentence(?sentence) \wedge \\ matching : MatchingRelation(?matrel) \wedge \\ uncertainty : saidBy(?sentence, ?operator) \wedge \\ matching : MatchingOperator(?operator) \wedge \\ uncertainty : saidAbout(?sentence, ?matrel) \wedge \\ uncertainty : Objective(?derivation) \wedge \\ uncertainty : Uncertainty(?uncer) \wedge \\ uncertainty : hasUncertainty(?sentence, ?uncer) \rightarrow \\ uncertainty : derivationType(?uncer, ?derivation) \end{aligned} \tag{3}$$

There are one or more rules for each specific uncertainty type, nature, model, derivation and temporal validity. An example of a rule that we use in this step of the matching strategy is reported in (3). In our strategy the rules are applied to the matching ontology with the use of a rule engine such as Jess[7]; the rules we use have to be restricted to the DL-Safe [25] subset to ensure tractable complexity of the reasoning processes. By applying the rules, all the matching relations are associated to their respective uncertainty.

At this point, a set of sub-knowledge bases is created by dividing the various instances of the concept `sentence` according to their uncertainty model. A SPARQL [32] query is used in this case to select the instances of the concept `sentence` that respect the desired restrictions. As an example the following SPARQL query returns the instances of `Sentence` that are associated to a probability theory model.

```
SELECT ?sentence ?type ?derivation ?nature ?model
WHERE {?sentence uncertainty:hasUncertainty ?uncertainty.
?uncertainty uncertainty:nature ?nature.
?uncertainty uncertainty:derivationType ?derivation.
?uncertainty uncertainty:uncertaintyType ?type.
?uncertainty uncertainty:uncertaintyModel ?model.
?probabilistic rdf:type uncertainty:Probability.
FILTER (?model = ?probabilistic)}
```

Once the set of assertions has been partitioned, the parallel reasoning processes can be launched. The reasoning processes are performed locally, exploiting the information provided by each assertion and the information provided by the ontologies to align. According to the information that has been provided to each reasoner, the process has to return back to the matching strategy the set of assertions that they believe to be the most reliable ones. Each reasoning process returns the results as instances of its representative concept (sub-concepts of the concept `Sentence`).

When the parallel reasoning processes come to an end, results are propagated back to the matching ontology by a reconciliation process. This process can be another reasoning process; in the case of our matching strategy we make use of SWRL rules to aggregate the results. Basically the reconciliation process follows the priority between the various reasoners that need to be made explicit. Some models are more reliable than others and this preference is defined in our system by SWRL rules. In our matching strategy we have to deal with inconsistencies because different relations on the same pair of elements can be classified as reliable relations by different reasoning processes. In case the preferences between inference models can not solve this situation we handle inconsistencies by assigning preferences between relations and operators, according to contextual factors: (e.g., analysing the data types of the elements to match. *Equivalence* relations have the highest priority in case of Strings, because Instance-based equivalence between integers is less reliable than the one between Strings). Rules are also used to propagate the best matching relations to other elements of the ontologies exploiting structural information from the original ontologies to match. This case can again be managed using a defeasible rules system such as DR-Prolog [3], which provides different precedences between rules, to help the decision process in inconsistent situations.

[7] in the Java prototype we developed, we have used Jess Rule engine (`http://herzberg.ca.sandia.gov/`) with the support of JessTab [31] and Protégé (`http://protege.stanford.edu/`) to translate SWRL rules and assertions from OWL to Jess and backwards.

4 Experimental Evaluation

Let us now describe the use of OAEI as a validating test for our data integration system. The comparison has been carried out with the results of the 2007 contest [4]. The majority of the participants of the context are based on a linear combination of several matching operators. In some cases this aggregating function is adaptive (Asmov [33], Prior+ [34]), while in some other cases it is fixed and defined by several experimentation (RiMOM [35], Sambo [36], Soda [37], Ola [38], TaxoMap [39], X-Som [40]). Some approaches are based on Possibility theory such as DSSim [28] and OWL-CM [41]. The test is performed using a palette of five semantically different matching operators (*JaroWrinkler Matching Operator, WordNet Label Matching Operator, Description Matching Operator, Type-based Matching Operator* and *Instance-based Matching Operator*). The results are stored in a matching ontology and the most reliable relations are extracted by our matching strategy. Each relation is associated to the respective matching operator. The reasoners that we used are a Fuzzy Description Logic reasoner [16] for the matching relations classified as probabilities (e.g. JaroWinkler matching operator) and a classic Description Logic reasoner [42] to process certain matching relations that the rules classify as necessities (e.g. String matching operator). The matching strategy has been developed as a Java prototype that we used to run the tests.

4.1 Results

The testbed can be divided in five subcategories:

101-104 This test set is the easiest set of tests. The first task is to match the reference ontology 101 with itself, the second test requires to generate a matching from the reference ontology to an ontology totally irrelevant (102 is a wine ontology). The ontology 103 represents a language generalization (unavailable constraints are replaced by their generalization): this ontology is an OWL-Lite generalization of the ontology 101. Finally the ontology 104 represents a language restriction with respect to the reference ontology 101. The constraints that are not available in OWL-Lite are simply removed. In the case of this first set of tests, our algorithm does not generate perfect alignments. In the case of 103 the propagation mechanism of the strategy does not perform well because of the generalization of the ontology. The various matching operators generate several reliable relations that the strategy can not recombine correctly in relation to their priority.

201-210 In this set of test cases, the structure of ontology

ISBN: 978-989-95079-6-8

Test #	Name	Prec.	Rec.	fMeas.
101	Reference Alignment	1	1	1
102	Irrelevant Ontology	NaN	NaN	NaN
103	Language Generalization	0,81	0,56	0,66
104	Language Restriction	0,94	0,94	0,94

is preserved. Syntactical changes are introduced: labels and identifiers are replaced by random names, misspellings, synonyms, language translation and moreover the comments in some cases have been suppressed. Our matching strategy obtains good results in this set of tests, because of the variety of the palette of matching operators that are considered in the strategy. In some cases (202, 209, 210) the Recall value is low, this because in case a matching operator does not create a possible relation this can not be automatically generated by the strategy. SWRL rules are used to propagate the best results to related elements, but if the matching relation is not created by any matching operator the propagation mechanism is not effective.

Test #	Name	Prec.	Rec.	fMeas.
201	No Names	0,92	0,90	0,91
202	No Names, No Comments	0,88	0,15	0,26
203	No Comments	0,97	0,97	0,97
204	Naming Conventions	0,92	0,92	0,92
205	Synonyms	0.91	0,90	0,90
206	Translation	0,96	0,70	0,95
207		0,98	0,70	0,97
208		0,96	0,72	0,82
209		0,87	0,34	0,49
210		0,92	0,12	0,22

221-247 In this case the set of tests can be divided into two subgroups: **221-231** and **232-247**. The first subgroup contains several structural modifications, such as the hierarchy that is flattened or expanded, and individuals, restrictions and data types that are suppressed. Each one of the documents in this subgroup has been modified by a structural change. Because of the fact that the labels and comments are preserved, the modifications have little influence on our system. The use of *Description Matching Operator* and *Name Matching Operator* allows the strategy to find most of the correct alignments using just the labels and comments information. In the second subgroup (232-247), the modifications made by combinations of the single modifications used in 221-231. Our system obtains good results for 232-247 as well.

Test #	Name	Prec.	Rec.	fMeas.
221	No Specialization	0,92	0,92	0,92
222	Flattened Hierarchy	0,92	0,92	0,92
223	Expanded Hierarchy	0,93	0,93	0,93
224	No Instance	0,93	0,93	0,93
225	No Restrictions	0,91	0,91	0,91
228	No Properties	0,97	0,97	0,97
230	Flattened Classes	0,95	0,96	0,95
231	Expanded Classes	0,93	0,93	0,93
232		0,95	0,95	0,95
233		1	1	1
236		1	1	1
237		0,92	0,92	0,92
238		0,91	0,91	0,91
239		0,93	0,97	0,95
240		0,85	0,88	0,87
241		1	1	1
246		0,97	1	0,98
247		0,91	0,94	0,93

Note that the observations about the Recall value we made about the previous tests, here are not valid since the ontologies provide sufficient information to the matching operators that can create reliable relations. Also in case of 222 where there is no hierarchy the results are satisfactory because of the information provided by the ontology.

248-266 This set of documents represents the most challenging case. This set combines structural and syntactical suppressions. The single challenges represented by the two previous set of documents are mixed in this set. All labels and identifiers are replaced by random names, and the comments are also suppressed. In this case our system does not perform well because no hierarchical representation is preserved so the strategy can not propagate the few correct matching found. This results in a low level of Recall. However, not enough information is provided in the ontologies, and the matching strategy can only find few alignments. The tests from 254 to 262, are the most difficult since almost all literal (labels and comments) and structural information are removed. In this case the propagation of the results can not take place because the relations that the strategy discovers can not be associated to other elements because of the structural information that is missing. When some structural information is preserved, the strategy can exploit this information in order to create possible matches, starting from the relations discovered by the single matching operators.

Test #	Name	Prec.	Rec.	fMeas.
248		0,88	0,14	0,25
249		0,74	0,14	0,24
250		0,92	0,33	0,49
251		0,86	0,13	0,22
252		0,65	0,11	0,19
253		0,88	0,14	0,25
254		0,90	0,27	0,42
257		0,92	0,33	0,49
258		0,86	0,13	0,22
259		0,65	0,11	0,19
260		0,82	0,31	0,45
261		0,69	0,27	0,39
262		0,90	0,27	0,42
265		0,82	0,31	0,45
266		0,69	0,27	0,39

301-304 This test set is composed by real ontologies contextually related to the reference ontology 101. The ontologies in this test set represent bibliographical information and they have been defined independently each other. This test represents a real world case of ontology alignment. Our strategy performs in the average with respect to the other systems evaluated. In the case of ontologies 301 our approach finds most of the correct alignments, but it also returns some wrong results. The alignment results for 302 and 303 are far from satisfactory. The reason is that these ontologies do not provide individuals and with shallow class hierarchy, where classes and properties are not related. In this case as well, the recall value is low: the matching operators are very sensible to the noise in data; moreover without a useful hierarchy the few matching relations identified by the matching operators can not be propagated or enforced with the support of hierarchical information. The ontology 304 has similar structure and vocabularies to the reference ontology 101 and in this case, the results are slightly better than the previous alignments.

Test #	Name	Prec.	Rec.	fMeas.
301	Real: BibTeX/MIT	0,9	0,6	0,71
302	Real: BibTeX/UMBC	0,79	0,46	0,58
303	Real: Karlsruhe	0,83	0,49	0,62
304	Real: INRIA	0,81	0,62	0,70

ISBN: 978-989-95079-6-8

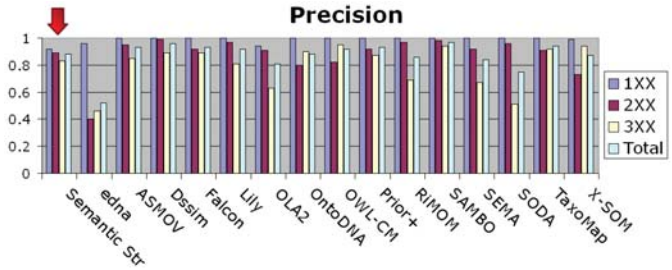


Figure 1: Graph of the Precision values of the different matching algorithms.

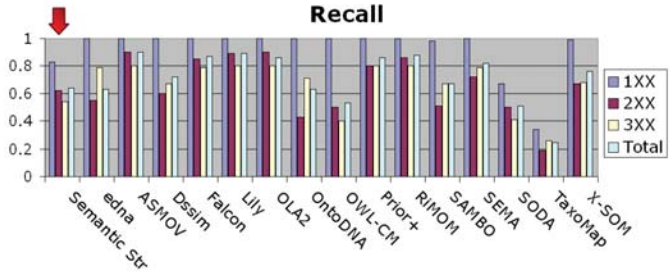


Figure 2: Graph of the Recall values of the different matching algorithms.

As one can see from the graphs in Figure 1 and 2 , our algorithm has acceptable results but under the best scores. These results are anyway reasonably encouraging, because our strategy is neither entailed to a particular reasoning model nor to a set of specific matching operator. More important our strategy does not imply a training phase that is typically required by the other systems. Our semantics-aware strategy is a new approach to ontology matching problem that provide a approach generally valid and not dependent on the domain. At this stage the results are very sensible to factors independent to the strategy; such as the quality of the matching operators and the available reasoners.

For this test we used a Probabilistic Description Logics reasoner and a traditional Description Logic reasoner in case of matching relations without a confidence degree. Moreover the inconsistencies, in the knowledge base, are not treated by a reasoning process but with the use of SWRL rules. As soon as new reasoners or new models appears we can exploit their use with our strategy. Moreover or strategy is highly customizable: if a new matching operator is plugged in the matching strategy is just necessary to create the related concept in the matching ontology and the SWRL rules reacting to the semantics of the new matching operator.

In case of the first series of test the strategy do not obtain the best results but, instead provides results under the average. Even the most simple matching operators, such as Edna, performs better than our strategy. The reason of this behaviour has been identified in the confusion of the strategy that considers the relations generated by different operator all reliable at the same level. We had run the same test only with JaroWinker matching operator and we obtained an average value of Precision 1 and Recall 1 for all the first series; this result confirm our previous assumption. In the case of the second and the third series the results are in the average of the approaches presented to the contest. The graph in Figure 1 shows the comparison of the Precision of our strategy with respect to the other algorithms. As is possible to see the values are very closed to 1, except in some isolated cases (Edna). These results show that the various strategies generate correct relations. The graph in Figure 2 shows the comparison of the Recall of our strategy with respect to the other algorithms. As is possible to see the values this time instead are far from the limit. Only few strategies (Asmonv, Falcon, Lily, OLA2) obtain results around 0.8, while the others still remain over the $0,5$. This generally low value of Recall means that the algorithm does not propagate well the good matching relations by exploiting the structure of the ontologies.

5 Conclusions

In this paper we presented preliminary work on a framework for managing different types of uncertainty and a possible application to the Schema Matching problem. The Ontology of Uncertainty, proposed by the W3C's UR3W-XG incubator group, provides a vocabulary to annotate different sources of information with different types of uncertainty. We argue that such annotations should be clearly mapped to corresponding reasoning and representation strategies. This mapping allows the system to analyse the information on the basis of its uncertainty model, running the inference process according to the respective uncertainty. In this way we can also deal with complex situations that do not tailor to the traditional strategies: as an example, matching relations provided by a user with her specific level of expertise. In our scenario the sources of information analysed, according to different uncertainty models are independent and no intersection among them has to be managed. This particular case allows a straight-forward use of the Ontology of Uncertainty to drive the reasoning process, although in general the assumption of independence among the source of information is a lucky case. The need of additional work on the Ontology of Uncertainty is necessary in order to support reasoning processes when combinations of uncertainty models are applied to a single source of information. This outlook is promising in order to provide more expressive frameworks for reasoning under different types of uncertainties, but definitely the need of more research in this direction is evident.

Acknowledgements

This work was partly funded by the Italian Ministry of Research under FIRB contract n. RBNE05FKZ2_004, TEKNE.

References

[1] Maurizio Lenzerini. Data integration is harder than you thought. In *CooplS '01: Proceedings of the 9th International Conference on Cooperative Information Systems*, pages 22–26, London, UK, 2001. Springer-Verlag.

[2] X. Luo, C. Zhang, and N. R. Jennings. A hybrid model for sharing information between fuzzy, uncertain and default reasoning models in multi-agent systems. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(4):401–450, 2002.

[3] Grigoris Antoniou and Antonis Bikakis. Dr-prolog: A system for defeasible reasoning with rules and ontologies on the semantic web. *IEEE Trans. Knowl. Data Eng.*, 19(2):233–245, 2007.

ISBN: 978-989-95079-6-8

[4] Jérôme Euzenat, Antoine Isaac, Christian Meilicke, Pavel Shvaiko, Heiner Stuckenschmidt, Ondrej Svab, Vojtech Svatek, Willem Robert van Hage, and Mikalai Yatskevich. Results of the ontology alignment evaluation initiative 2007. In *Proc. 2nd ISWC 2007 international workshop on ontology matching (OM), Busan (KR)*, pages 96–132, 2007.

[5] I Kenneth J. Laskey, Kathryn B. Laskey, Paulo C. G. Costa, Mieczyslaw M. Kokar, Trevor Martin, and Thomas Lukasiewicz. Uncertainty reasoning for the world wide web. *W3C Incubator Group Report*, 2008.

[6] B. de Finetti. La logique de la probabilité. In *Actes du Congres International de Philosophie Scientifique a Paris 1935, Tome IV*, pages 1–9. 1936.

[7] Didier Dubois and Henri Prade. Can we enforce full compositionality in uncertainty calculi. In *In Proc. of the 11th Nat. Conf. on Artificial Intelligence (AAAI-94*, pages 149–154. AAAI Press / MIT Press, 1994.

[8] George Boole. The calculus of logic, 1848.

[9] L.A. Zadeh, K.S Fu, K. Tanaka, and M. Shimura. *Fuzzy sets and their applications to Cognitive and Decision Processes*. 1975.

[10] Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*, 2003.

[11] Deborah L. McGuinness and Frank van Harmelen. Owl web ontology language overview. *W3C Recommendation*, 2004.

[12] Shiyong Lu, Ming Dong, and Farshad Fotouhi. The semantic web: Opportunities and challenges for next-generation web applications. *International Journal of Information Research*, 7:2002, 2002.

[13] I. Horrocks and U. Sattler. Ontology reasoning in the SHOQ(D) description logic. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence*, 2001.

[14] Thomas Lukasiewicz. Expressive probabilistic description logics. *Artif. Intell.*, 172(6-7):852–883, 2008.

[15] Pavel Klinov. Pronto: Probabilistic dl reasoning with pellet, 2008.

[16] Fernando Bobillo and Umberto Straccia. fuzzydl: An expressive fuzzy description logic reasoner. In *2008 International Conference on Fuzzy Systems (FUZZ-08)*, pages –. IEEE Computer Society, 2008.

[17] Janusz Kacprzyk and Slawomir Zadrozny. Computing with words in intelligent database querying: standalone and internet-based applications. *Inf. Sci.*, 134(1-4):71–109, 2001.

[18] Thomas Lukasiewicz and Umberto Straccia. Description logic programs under probabilistic uncertainty and fuzzy vagueness. In *ECSQARU*, pages 187–198, 2007.

[19] Thomas Lukasiewicz and Umberto Straccia. Uncertainty and vagueness in description logic programs for the semantic web, 2007.

[20] Didier Dubois, Jérôme Mengin, and Henri Prade. Possibilistic uncertainty and fuzzy features in description logic. A preliminary discussion. In E. Sanchez, editor, *Fuzzy logic and the semantic web*, pages 101–113. http://www.elsevier.com/, 2006. DMenP001.

[21] Fernando Bobillo, Miguel Delgado, and Juan Gómez-Romero. Extending fuzzy description logics with a possibilistic layer. In *URSW*, volume 327 of *CEUR Workshop Proceedings*, 2007.

[22] Andriy Nikolov, Victoria S. Uren, Enrico Motta, and Anne N. De Roeck. Using the dempster-shafer theory of evidence to resolve abox inconsistencies. In *URSW*, volume 327, 2007.

[23] L.A. Zadeh. Fuzzy sets. *Information Control*, 8:338–353, 1965.

[24] Hiroshi Maeda and Yuji Nobusada. A study on the parallel computability of multi-fold approximate reasoning. *Fuzzy Sets Syst.*, 97(2):129–144, 1998.

[25] Boris Motik, Ulrike Sattler, and Rudi Studer. Query answering for owl-dl with rules. In *Journal of Web Semantics*, pages 549–563, 2004.

[26] Erhard Rahm and Philip A. Bernstein. A survey of approaches to automatic schema matching. *VLDB Journal: Very Large Data Bases*, 10(4):334–350, 2001.

[27] Pavel Shvaiko and Jérôme Euzenat. *Ontology matching*. Springer-Verlag, Heidelberg (DE), 2007.

[28] Miklos Nagy, Maria Vargas-Vera, and Enrico Motta. Dssim-ontology mapping with uncertainty. In *Ontology Matching*, 2006.

[29] Andrea Calì, Thomas Lukasiewicz, Livia Predoiu, and Heiner Stuckenschmidt. A framework for representing ontology mappings under probabilities and inconsistency. In *URSW*, 2007.

[30] Ian Horrocks, Peter F. Patel-Schneider, Harold Boley, Said Tabet, Benjamin Grosof, and Mike Dean. Swrl: A semantic web rule language combining owl and ruleml. *W3C Member Submission*, 2004.

[31] Henrik Eriksson. Using jesstab to integrate protégé and jess. *IEEE Intelligent Systems*, 18(2):43–50, 2003.

[32] Eric Prud'hommeaux and Andy Seaborne. Sparql query language for rdf. *W3C Member Submission*, 2008.

[33] Yves Jean-Mary and Mansur Kabuka. Asmov results for oaei 2007. In *Proc. 2nd Ontology matching workshop, Busan (KR)*, pages 141–150, 2007.

[34] Ming Mao and Yefei Peng. The prior+: Results for oaei campaign 2007. In *2nd International Workshop on Ontology Matching (OM) at 6th International Semantic Web Conference (ISWC), Busan, Korea, 2007*, 2007.

[35] Yi Li, Qian Zhong, Juanzi Li, and Jie Tang. Result of ontology alignment with rimom at oaei'07. In *The Second International Workshop on Ontology Matching*, 2007.

[36] Patrick Lambrix and He Tan. Sambo-a system for aligning and merging biomedical ontologies. *Web Semant.*, 4(3):196–206, 2006.

[37] Sami ZGHAL, S. Ben Yahia, Engelbert MEPHU NGUIFO, and Y. Slimani. Soda: an owl-dl based ontology matching system. In *OAEI'07 - OM'07 - ISWC'07*, Busan (Korea), nov 2007.

[38] Jérôme Euzenat, David Loup, Mohamed Touzani, and Petko Valtchev. Ontology alignment with ola. pages 333–337, 2004.

[39] Haifa Zargayouna, Brigitte Safar, and Chantal Reynaud. TaxoMap in the OAEI 2007 alignment contest . In *Ontology Alignment Evaluation Inititiative (OAEI) 2007 Campaign - Workshop on Ontology Matching*, 2007.

[40] Carlo Curino, Giorgio Orsi, and Letizia Tanca. X-som: A flexible ontology mapper. In *DEXA '07: Proceedings of the 18th International Conference on Database and Expert Systems Applications*, pages 424–428, Washington, DC, USA, 2007. IEEE Computer Society.

[41] Boutheina Ben Yaghlane and Najoua Laamari. Owl-cm: Owl combining matcher based on belief functions theory.

[42] Evren Sirin, Bijan Parsia, Bernardo C. Grau, Aditya Kalyanpur, and Yarden Katz. Pellet: A practical owl-dl reasoner. *Web Semantics: Science, Services and Agents on the World Wide Web*, 5(2):51–53, June 2007.

ISBN: 978-989-95079-6-8

Fuzzy Descriptions to Identify Temporal Substructure Changes of Cooccurrence Graphs

Matthias Steinbrecher[1] Rudolf Kruse[1]

1. Department of Knowledge Processing and Language Engineering
Otto-von-Guericke University of Magdeburg
Universitätsplatz 2, 39106 Magdeburg, Germany
E-mail: {msteinbr, kruse}@ovgu.de

Abstract— *Cooccurrence graphs easily grow very dense when applied to represent binary association patterns of large amounts of data. Therefore, postprocessing is needed to extract valuable information from them. We propose an approach to identify subgraphs of cooccurrence graphs that show a certain temporal behavior. This behavior is described with linguistic variables and fuzzy connectives defined over the change rate domains of certain graph measures. These measures assess graph properties whose change over time the user is interested in. To justify our proposed method, we are going to present evidence from a real-world dataset.*

Keywords— Cooccurrence graphs, Temporal change, Fuzzy description

1 Introduction

Frequent pattern mining has become a prominent method for identifying patterns in large volumes of data and led to algorithms for postprocessing these patterns [1] or transfer the underlying ideas to other data structures such as graphs [2, 3, 4]. Since the search for frequent patterns necessarily has to deal with subsets of input data, one easily runs into the problem of combinatorial explosion which is reflected by the common problem of finding more patterns than there are input data. We have addressed this issue in previous work [5], arguing that the user requires tools that allow to filter the results in order to identify only those results that meet his criteria (such as interestingness, novelty, etc.). In addition to that we also provided arguments and empirical evidence [6] that patterns usually do not arise all of a sudden but evolve or disappear rather slowly as time passes. In consequence we proposed a method that allows the user to specify linguistically (in terms of fuzzy variables) the temporal behavior of the values of association rules' evaluation measures that he is interested in. The presented algorithm thinned out the entire rule set retaining just those rules that matched the users' concepts (to some degree).

In this paper we develop this idea further while keeping the way of describing temporal behavior (explained in the background section) but transferring it to a different area of application.

This area of application comprises the identification of interesting substructures in cooccurrence graphs. These graphs arise quite naturally wherever, theoretically speaking, one is interested in the fact that two entities share some property with respect to a so-called location (which not necessarily has to be a spatial artifact but often is). Two authors being cited by the same paper [7], two persons having visited the same website [8, 9, 10], or two crimes being committed at the same location are just three examples of cooccurrences. We will elaborate the possible applications later in the future work section. All these examples can be represented by an undirected graph with the node set comprising all possible locations and the weighted edges representing the cooccurrences. A typical graph of a real-world application is depicted in Fig. 1. It represents 100 locations in a 3D gaming environment. Whenever two locations have been visited by the same players, an edge is inserted (which also gets assigned a weight drawn as the width of the edge representing the number of common visitors, but this is omitted here). It is pretty obvious that a user needs some assistance tools that allow him to identify interesting substructures (edge combinations). This becomes even more important if we have multiple such graphs representing the cooccurrences of different time frames (e. g. weeks or months). The focus of this paper is to present a straightforward yet powerful approach how to identify common substructures in a collection of cooccurrence graphs by means of linguistics expressions that address the temporal change in these patterns.

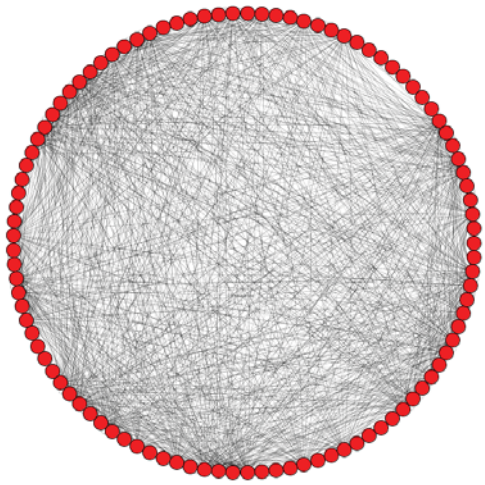


Figure 1: A typical cooccurrence graph with 100 nodes as it arose in a small-sized application (online gaming players that having visited the same locations in a 3D world). It is clear that this representation calls for some means of postprocessing in order to extract usable information from it.

The remainder of this paper is organized as follows: Section 2 introduces the notation used throughout the paper and revisits the linguistic filtering introduced in [6]. Section 3 mo-

ISBN: 978-989-95079-6-8

tivates and presents the method of extracting substructures of cooccurrence graphs that exhibit a certain user-specified temporal behavior. To underpin our proposal, we use a real-world dataset to illustrate the different stages of analysis in Section 4. Since this dataset reveals substructures that could not have been better generated manually, we refrain from handcrafting an artificial dataset and use selected subsets instead. We conclude our paper in Section 5 and propose other applications as well as possible promising extensions.

2 Background and Nomenclature

2.1 Graph Notations

In this paper we are going to deal exclusively with undirected graphs which we model as a tuple $G = (V, E)$ with vertices V and edge set E with

$$E \subseteq V \times V \setminus \{(v, v) \mid v \in V\},$$

and the constraint

$$(u, v) \in E \;\Rightarrow\; (v, u) \in E$$

to emphasize the undirected character. We will interpret the graphs as cooccurrence graphs where edges determine the number of cooccurrences (of whatever kind). This is taken into account with an edge weight function for every edge $e = (u, v)$:

$$w : E \to \mathbb{N}_0 \;\text{ with }\; w(e) = w(u, v) = w(v, u).$$

In the figures, this weight is represented as the edge width, thus we use the notion *width* and *weight* interchangeably. Given a subset $W \subseteq V$, we can induce a subgraph $G_W = (W, E_W)$ with

$$E \supseteq E_W = \{(u, v) \mid u, v \in W \wedge (u, v) \in E\}.$$

In the remainder we will sometimes use such a subset W in the context of a graph; it is G_W that we then refer to. A threshold θ defines the subgraph $G_\theta = (V, E_\theta)$ with

$$E_\theta = \{(u, v) \mid (u, v) \in E \wedge w(u, v) \geq \theta\},$$

i. e., as the graph containing only edges with a weight greater or equal to θ. Both operations can of course be combined, i. e., $G_{W,\theta}$ represents the subgraph of G induced by the node set W after having removed all edges with weight less than θ.

Since we will deal with sequences of graphs, we denote the temporal index as a superscript. All graphs share the same node set V and differ only in their edge sets or edge weights or both. Given a sequence $G^{(1)}, \ldots, G^{(n)}$ of graphs, we define the sum of these graphs as follows: $G_\Sigma = (V, E_\Sigma)$ with

$$E_\Sigma = \bigcup_{i=1}^{n} E^{(i)} \;\text{ and }\; w_\Sigma(u, v) = \sum_{i=1}^{n} w^{(i)}(u, v).$$

2.2 Linguistic Filtering Revisited

As described in [6] it is not only important to find patterns that meet some predefined constraints (such as minimum support or confidence) but also interpret these patterns in terms of temporal change. If a pattern describes a problem in some domain or an interesting customer behavior, it might be of interest to find such evolving patterns early. The underlying idea is as follows: given a pattern, we devise a set of evaluation measures that characterize the particular pattern (in [6] we used association rule measures such as support, lift, confidence, etc.). Next, the time series of a user-selected subset of these measures is calculated for every pattern. Each time series in turn is aggregated to a single value representing the overall trend, if any. The domain of this aggregate (i. e., the change rate domain) is equipped with an adequate fuzzy partition. Given a fuzzy rule antecedent (representing the user's intention of what temporal behavior of which measure(s) he is interested in), for every rule a membership degree to this concept is computed and an ordered list of rules according to these degrees is returned.

3 Spatio-temporal Filtering

3.1 Motivation

The objective of our approach is to answer questions of the following type (given a sequence of cooccurrence graphs):

> "First, what are interesting candidates for subgraphs that it would be worth looking at over time?"

and

> "Second, given a (still intractable large) set of subgraphs, which graphs become more sparse and less balanced over time?"

Before we turn to the algorithmic part of our approach, we need to negotiate which types of substructures within the graphs are most interesting to users. We will exploit the edge weights for this purpose. Several measures are needed to quantify for every subgraph aspects such as size, completeness, edge balance, etc.

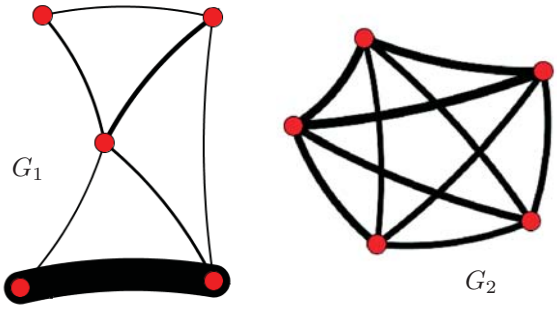
If the cooccurrence graphs represent visits of different web pages within the same online shop portal, then it might be desirable to know whether customers are able to use the web portal as intended by the owners. Are there dead ends where users are stuck? What are the "hot spot" sites, i. e., the pages that attract the most users and are visitors able to find the recently introduced shortcut to related pages? How do the accesses to the support area of the site change after renewing the navigational aids, etc.

We explicitly stress that subgraphs that are heavily interconnected with large edge weights only provide us with a *hint* that there may be an interesting visiting pattern. However, we can never conclude transitivity just from the cooccurrence graph! This is due to the fact that it only represents *binary* cooccurrences. Even a fully connected graph does not tell us anything about individual events. The sets of cooccurring events whose cardinality is represented by the edge weights even might be mutually disjoint. However, these subgraphs are found to be valuable hints that are worth being investigated.

3.2 Graph Measures

Focussing on the before-mentioned type of aspects one can identify highly connected subgraphs with large edge weights to be one type of substructures that are most interesting to users. Another type may be single edges just connecting two nodes or substructures that are highly interconnected but with

ISBN: 978-989-95079-6-8



Measure		G_1	G_2
size	Size	5.0	5.0
comp	Completeness	0.7	1.0
wght	Total Weight	5082.0	7906.0
avg	Average Weight	726.0	790.6
med	Median Weight	257.0	770.0
dev	Std. Dev.	1191.1	129.3

Figure 2: Two graphs with the same number of nodes. G_1 lacks three edges to be complete, therefore the completeness is just 0.7, whereas the clique G_2 yields 1.0. The rather large difference between average weight and median weight for G_1 (in contrast to G_2) indicate an imbalanced edge widths distribution which is strengthened by the large standard deviation value. The two graphs obviously justify this finding. The layout only acts as a visual cue and thus does not have any influence on the graph measures.

a large imbalance in the edge weights. The latter might represent two active sets of websites (large edge weights) between which users are able to navigate back and forth (numerous edges in between but with small weights since not every user is likely to use the offered navigational freedom).

The last arguments call for measures that on the one hand capture the mentioned properties of subgraphs and on the other hand allow to build a fuzzy partition on their domains since we are not going to ask for subgraphs *with 9 nodes and edges with weights greater than 100* but for *large* subgraphs with *moderately sized* edges. We use the following set of (sub)graph measures to quantify different aspects:

$$\begin{aligned}
&\text{Size} && \operatorname{size}(G_W) = |W| \\
&\text{Completeness} && \operatorname{comp}(G_W) = \frac{2\,|E_W|}{|W|^2 - |W|} \\
&\text{Edge Weight} && \operatorname{wght}(G_W) = \sum_{e \in E_W} w(e)
\end{aligned}$$

The size simply represents the number of nodes of the subgraph, whereas completeness refers to the relative number of edges compared to the maximal number. Zero represents an isolated graph (no edges) while a value of 1 designates a clique. Finally, the edge weight simply returns the sum of all edge weights without giving any clue about the distribution of these weights among the edges. Therefore, three additional measures are used: $\operatorname{avg}(G_W)$ calculates the arithmetic mean of all edge weights, $\operatorname{med}(G_W)$ returns the median of the weights and $\operatorname{dev}(G_W)$ represents the standard deviation of the weights. Fig. 2 illustrates these intentions with two graphs of the same size. Note, that these are subgraphs from real-world data and no artificial graphs that were crafted to meet the requirements.

3.3 *Candidate Graph Generation*

As we are now equipped with measures to assess certain aspects of subgraphs that we would like to track over time, the remaining question is how to determine such candidate graphs? It is clear that a brute-force approach (testing all subsets of nodes as potential subgraph node sets) fails immediately due to runtime problems, even for small node sets. We therefore promote the following heuristic: The graphs of all time frames are added as shown in Section 2 to arrive at the sum graph G_Σ (or simply the cooccurrence graph if we ignore the time frames). Next, a threshold θ is chosen and the graph's components (disconnected subgraphs) $\mathcal{C}_{G_\Sigma} = \{C_1, \ldots, C_j, \ldots, C_m\}$ of $G_{\Sigma,\theta}$ are taken as the candidate subgraphs. The choice of θ can be entirely left to the user (e. g. by offering a graphical preview tool that shows the components instantly whenever the user selects a new threshold via a slider) or θ may be determined in such a way to limit either the number of components or the (average) size of the components.

3.4 *Matching Against Linguistic Concepts*

Whatever way of determining the granularity of components is chosen, we are left with a set of mutual disjoint node sets $\mathcal{C}_{G_\Sigma}$ that are used to create a sequence of subgraphs $\left\langle G^{(i)}_{C_j} \right\rangle$, $i = 1, \ldots, n$, $j = 1, \ldots, m$ (one sequence for every subgraph induced by the node set) that are evaluated against the user-specified temporal behavior description. A time series is generated for every measure referenced in this user description. The temporal change within this time series is computed and the degree of membership to the user description is calculated. We will employ a simple regression approach, i. e., we fit a regression line into the time series and interpret its slope as an indicator of decrease, stability and increase.

The example concept from the motivation of this section is repeated here:

"Completeness is decreasing and std. deviation is increasing"

Translated into a linguistic concept, the user may specify

$$\langle\, \Delta_{\text{comp}} \texttt{ is decr} \,\wedge\, \Delta_{\text{dev}} \texttt{ is incr} \,\rangle\,,$$

which is evaluated to

$$\top\left(\mu^{(\text{decr})}_{\Delta_{\text{comp}}}(C_j), \mu^{(\text{incr})}_{\Delta_{\text{dev}}}(C_j)\right),$$

where $\top$ represents a t-norm modelling the fuzzy conjunction (we use $\top_{\min}(a,b) = \min\{a,b\}$ in this paper). Fig. 3 depicts an example subgraph consisting of 9 nodes. Five time frames are shown with the respective edge weights. The graph is obviously becoming less dense with time, i. e., the completeness is decreasing. The chart for this measure is depicted in Fig. 4. In analogy to this, Fig. 5 shows the increasing deviation of the edge weights which is attributed to the emergence of the strong cooccurrence (the sudden appearance in this case can be explained with time frames that were too large to appropriately cover the short period during which this strong cooccurrence emerged). If we equipped the change rate domains

ISBN: 978-989-95079-6-8

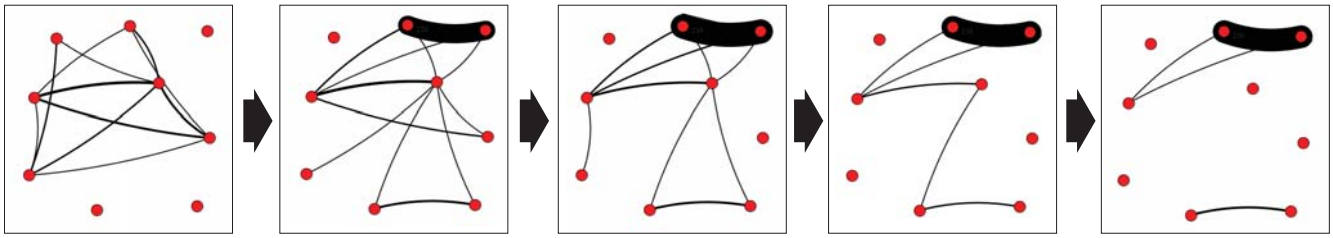


Figure 3: The temporal evolution of the graphs induced by a set of nine nodes. The number of edges is decreasing with time resulting in an almost isolated graph. Simultaneously, the edges that are remaining grow more and more unbalanced, i. e., the deviation of the edge weights is increasing. Both time series of the corresponding measures comp and dev are shown in Fig. 4 and Fig. 5, respectively.

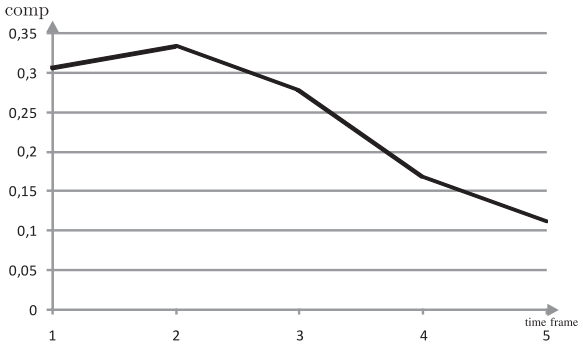


Figure 4: Time series with decreasing trend for the completeness of the edge weights for the five graphs of the time frames depicted in Fig. 3.

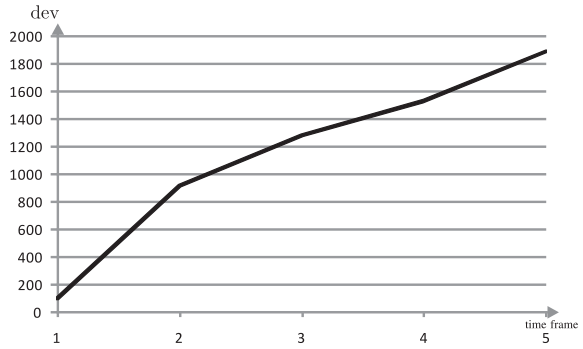


Figure 5: Time series with increasing trend for the standard deviation of the edge weights for the five graphs of the time frames depicted in Fig. 3.

(i. e., domains of the slopes of the regressions lines of the two time series) with appropriate fuzzy partitions (as we will do it in the experiments section) we could calculate the membership degree of this node set to the above-mentioned linguistic concept.

Summarizing, we state the following procedure:

1. Given a sequence $G^{(0)}, \ldots, G^{(n)}$ of cooccurrence graphs with their sum graph being G_Σ.
2. Based on an appropriate value θ, we calculate the candidate graph node sets $\{C_1, \ldots, C_m\}$ which are the vertices of the components of $G_{\Sigma,\theta}$.
3. The user provides a set of linguistic descriptions (fuzzy rule antecedents) that refer to the temporal change of the graph measures.
4. Provide fuzzy partitions for every domain of the change rate of the measures used in the descriptions of step 3.
5. Evaluate for every graph G_{C_j} the degree of membership to the linguistic concepts of step 3.
6. For every linguistic concept sort the graphs in descending order with respect to their membership degrees.

4 Experiments

Now that we have a tool at hand that allows us to determine the structural changes of subgraphs over time, we demonstrate the applicability with a real-world dataset that was already used to illustrate the examples above.

This dataset contains player contacts at certain locations within a 3D environment over a time period of six months. We carefully selected a subset of 100 such locations and discretized month-wise in order to result in a dataset large enough to justify the need for filtering while simultaneously being able to extract subgraphs that exhibit a structure and behavior that could not have been crafted more exemplary by hand. Therefore, we refrain from creating an artificial dataset with just the same structures.

Fig. 6 shows the sum graph of the dataset, i. e., the cooccurrences of six months among 100 locations. The edge weights indicate some subgraphs that are worth looking closer at. The threshold θ has been chosen to be 1000 to induce the candidate node sets.

We will match two linguistic concepts against these graph candidates: first, we are interested in decay, i. e., in graphs that show kind of a dissolving behavior, translating into a decreasing completeness and decreasing total weight. In the sample data this might indicate locations whose attractiveness is diminishing. A second concept that we would like to assess is that of an establishing pattern. An increasing average edge weight and deviation (of edge weight) might point out a phase of initial apparent random visiting of multiple locations which accumulates into a strong favored visiting pattern.

4.1 Concept 1: Decreasing Completeness and Weight

In order to evaluate the membership degrees to the linguistic concept

$$\langle\, \Delta_{\mathrm{comp}} \texttt{ is decr} \wedge \Delta_{\mathrm{wght}} \texttt{ is decr} \,\rangle,$$

ISBN: 978-989-95079-6-8

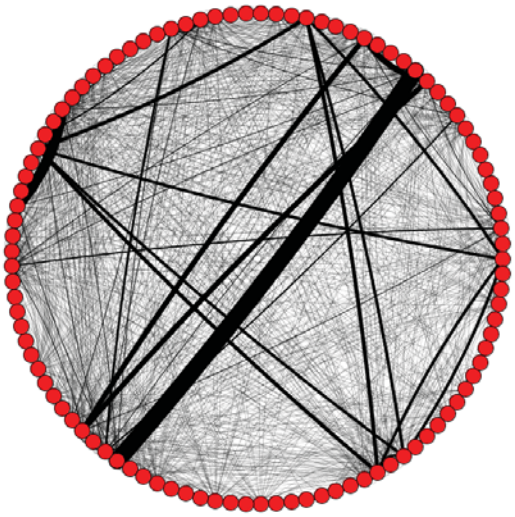


Figure 6: The sum graph of six months of visiting history of players in a 3D environment. We will match the major components (extracted via a user-specified threshold) against two linguistic concepts.

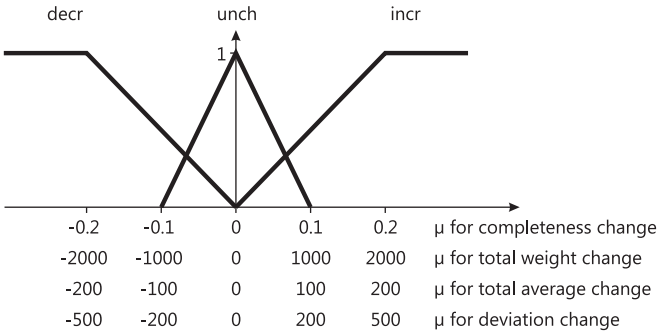


Figure 7: Fuzzy partitions for the four graph measures used in the two experiments.

we need to declare a fuzzy partition on the change rate domains of comp and wght. Fig. 7 displays all used fuzzy partitions. We apply three fuzzy sets. Note the asymmetric slopes of the borders. This setup has proven to be useful in this context since "unchanged" has a more strict semantic to users than the adjectives "decreasing" and "increasing". The respective values that determine the particular fuzzy partition can be read from the four different horizontal scales. These values have been selected with respect to the dataset since the quantity that renders a slope to be highly decreasing or increasing differs, of course, from dataset to dataset.

If we apply the linguistic concept to the candidate graphs and select the one with the highest membership (ignoring the remaining ones here for brevity), the subgraph whose history is depicted in the upper row in Fig. 8 scores 0.71. Most of the high degree can be attributed to the rapid loss of visits in the last two months. The membership degree was evaluated via

$$\min\{\mu^{(\text{decr})}_{\Delta_{\text{comp}}}(C_1), \mu^{(\text{decr})}_{\Delta_{\text{wght}}}(C_1)\} = \min\{0.71, 0.84\} = 0.71,$$

with C_1 being the set containing the five nodes. Data inspection revealed a newly set up structure which was heavily frequented shortly after opening followed by abating excitement.

4.2 Concept 2: Increasing Average and Deviation

We follow the same procedure to find the subgraph that scores best on the concept

$$\langle\, \Delta_{\text{avg}} \texttt{ is incr} \,\wedge\, \Delta_{\text{dev}} \texttt{ is incr} \,\rangle\,.$$

The lower part of Fig. 8 shows the resulting graph with a score of

$$\min\{\mu^{(\text{incr})}_{\Delta_{\text{avg}}}(C_2), \mu^{(\text{incr})}_{\Delta_{\text{dev}}}(C_2)\} = \min\{0.83, 0.89\} = 0.83,$$

The graph shows an establishing link between two nodes in parallel with a weakening in the remaining edges thus rendering the graph history becoming more unbalanced.

5 Conclusion and Future Work

In this paper we discussed the need to postprocess cooccurrence graphs if they grow very dense in practical use. We put forward a heuristic that allows to restrict the candidate node sets, and transferred the fuzzy concept matching approach from [6] to the graphical setting. We provided empirical evidence of the applicability by analyzing a real-world dataset containing six months of game player visits to 100 locations in a 3D environment.

As indicated in the introductory section, there are many other scenarios for which such an analysis might be interesting. We are most interested in applying our method to a web click stream analysis.

Since the used real-world dataset contained time stamps for every event, it is possible to generate a directed graph that also indicates which nodes were the source and target nodes for different users. This requires, of course, some heuristic with which to decide what should be the maximal period in which a user has to commit two visits to different nodes in order to assume that it was a transition and not just two independent visits.

A shortcoming of the presented approach is the heuristic to generate the candidate node sets. It worked well on the underlying datasets but one can imagine that it will return fewer and more interconnected components the more the threshold θ is reduced. Such subgraphs are also referred to as *giant connected components* [11]. To address this problem, we intend to phrase the problem in the setting of the emerging area of graph mining, more specific finding common substructures in a *single* graph [2, 12]. However, we will need a more specific subgraph definition since we have to account for the edge weights and not only the edge presence. This in turn automatically leads to another area of investigation: devising more sophisticated graph measures such as the edge betweenness centrality [13].

References

[1] Rakesh Agrawal, Tomasz Imielinski, and Arun N. Swami. Mining Association Rules between Sets of Items in Large Databases. In Peter Buneman and Sushil Jajodia, editors, *Proc. ACM SIGMOD Int. Conf. on Management of Data, Washington, D.C., May 26-28, 1993*, pages 207–216. ACM Press, 1993.

[2] Mathias Fiedler and Christian Borgelt. Subgraph support in a single graph. In *Proc. IEEE Int. Workshop on Mining Graphs and Complex Data (MGCS 2007 at ICDM 2007, Omaha, NE)*, pages 399–404. IEEE Press, Piscataway, NJ, USA, 2007.

ISBN: 978-989-95079-6-8

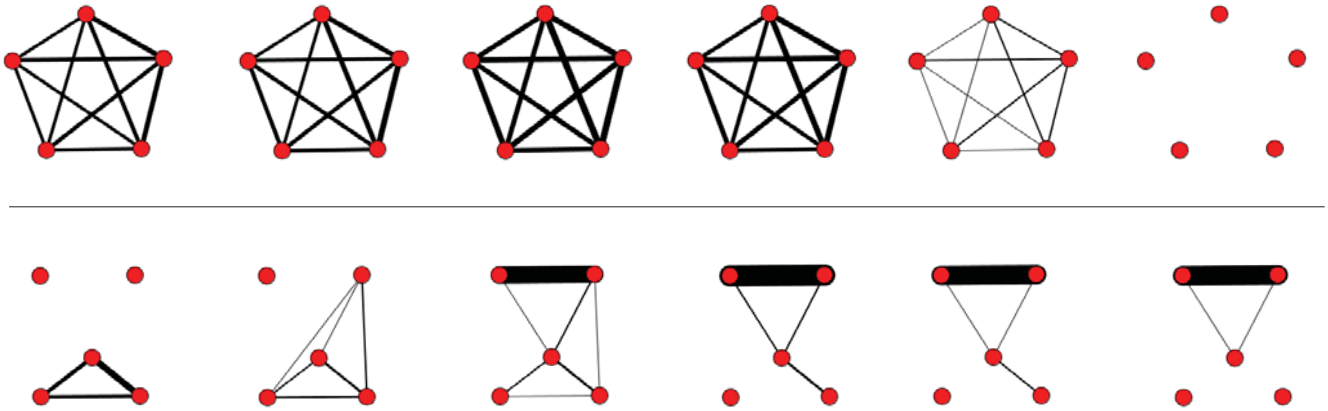


Figure 8: The histories of the two subgraphs that scored highest in the two experiments. The upper row shows a six-month development of five locations that were heavily visited but declined rather rapidly towards the end. It yielded a membership degree of 71% to the concept "completeness is decreasing and total weight is decreasing". The lower row depicts the best-scoring subgraph of the second experiment resulting in a membership degree of 83% to the concept "average is increasing and deviation is increasing."

[3] Christian Borgelt. On canonical forms for frequent graph mining. In *Workshop on Mining Graphs, Trees, and Sequences (MGTS'05 at PKDD'05, Porto, Portugal)*, pages 1–12, Porto, Portugal, 2005. ECML/PKDD'05 Organization Committee.

[4] Tobias Werth, Alexander Dreweke, Marc Wörlein, Ingrid Fischer, and Michael Philippsen. Dagma: Mining directed acyclic graphs. In IADIS, editor, *Proceedings of the ECDM 2008 (IADIS European Conference on Data Mining (ECDM))*, pages 11–17. IADIS PRESS, 2008.

[5] Rudolf Kruse, Christian Borgelt, Detlef D. Nauck, Nees Jan van Eck, and Matthias Steinbrecher. The role of soft computing in intelligent data analysis. In *Proc. 16th IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE'07, London, UK)*, pages 9–17. IEEE Press, Piscataway, NJ, USA, 2007.

[6] Matthias Steinbrecher and Rudolf Kruse. Identifying temporal trajectories of association rules with fuzzy descriptions. In *Proc. Conf. North American Fuzzy Information Processing Society (NAFIPS 2008)*, pages 1–6, New York, USA, May 2008.

[7] Levent Bolelli, Seyda Ertekin, and C. Lee Giles. *Knowledge Discovery in Databases: PKDD 2006*, volume 4213/2006, chapter Clustering Scientific Literature Using Sparse Citation Graph Analysis, pages 30–41. Springer Berlin / Heidelberg, 2006.

[8] Prasanna Desikan and Jaideep Srivastava. *Advances in Web Mining and Web Usage Analysis*, volume 3932/2006 of *Lecture Notes in Computer Science*, chapter Mining Temporally Changing Web Usage Graphs, pages 1–17. Springer Berlin / Heidelberg, 2006.

[9] M. E. J. Newman. Modularity and community structure in networks. *PROC.NATL.ACAD.SCI.USA*, 103:8577, 2006.

[10] Aaron Clauset, M. E. J. Newman, and Cristopher Moore. Finding community structure in very large networks. *Physical Review E*, 70:066111, 2004.

[11] Béla Bollobás. *Random Graphs*, chapter The Evolution of Random Graphs — the Giant Component, pages 130–159. Cambridge Univ Press, 2nd edition, 2001.

[12] Björn Bringmann and Siegfried Nijssen. *Advances in Knowledge Discovery and Data Mining*, volume 5012/2008 of *Lecture Notes in Computer Science*, chapter What Is Frequent in a Single Graph?, pages 858–863. Springer Berlin / Heidelberg, 2008.

[13] U. Brandes and C. Pich. Centrality estimation in large networks. *International Journal of Bifurcation and Chaos 17*, 7:2303–2318, 2007.

ISBN: 978-989-95079-6-8

Online Measurement of Oily-water Using Rule-based Fuzzy Concentration Map and Halogen Fluorescence

Gyeongdong Baek[1], Moon-Ho Choi[2], Tae-Ryong Jeon[1], Jaewoo Cho[1], and Sungshin Kim[1]

1 School of Electrical Engineering, Pusan National University, Busan, Korea
Email: {gdbaek, jtr, jwjw777, sskim}@pusan.ac.kr
2 MARSEN CO. Ltd., Busan, Korea
Email: master@marsen.co.kr

Abstract—*The aim of this study is to develop oil content measurement system using transmitted and scattered intensity. Two fluorescence analyses would be suggested with methods that it can use rapidly, continuously, conveniently and accurately in long-distance cruise vessels. The present paper consists of the structure of fluorescence measurement hardware and an interpretative analysis of experimental data. Proposed rule-based fuzzy concentration map offers the possibility of the conversion of fluorescence into oil content. This measurement method would be applied oil discharge monitoring control equipment in cruise vessels.*

Keywords—oil content meter, oil discharge monitoring system, fluorescence analysis, rule-based fuzzy concentration map.

1 Introduction

Maritime pollution controls are effected through the working of the MARPOL Convention and its associated Annexes, as developed by the IMO(international maritime organization) Maritime Environmental Protection Committee, MEPC. The vessel have an OFE(oil filtering equipment) on board. From 6 July 1998, all ships of 400 gross tonnages and above, regardless of age, must be fitted with 15ppm OFE. Some ships were able to use 100ppm OFE, but this is no longer acceptable. According to regulation 10(3)(b) of Annex I processed bilge water from machinery spaces is only allowed to be discharged into the sea through OFE system in a Special Area of Annex I to MARPOL 73/78 Convention [1].

The problem of detecting oil in water has practical significance, especially in the environment [2-3]. Optical measurement techniques are analytical methods commonly used in instrumentation for online measurement of oily water [4-5]. Online optical sensors have been used to monitor for example the coagulation and filtration processes in the water industry [6]. Online nephelometers have been used to measure oily water distributions [7]. In a study of smart nephelometer, artificial neural networks have been used to interpret online nephelometer data [8].

This paper, after working the sensing system with the halogen fluorescence, will attempt to analyze characteristic of the oil content. By comparison with the transmitted and scattered intensity of 25*mg/l* and 1000*mg/l* samples, proposed rule- based fuzzy concentration map is classified into different kinds of oil content. The following section provides a experimental system. Section 3 presents proposed rule-based fuzzy concentration map, section 4 and 5 show experimental result and contains concluding remarks.

2 Experimental System

A fluorescence measurement system is shown in Fig. 1. And system prototype is shown in Fig. 2. The experimental system consists of light emitting part, light receiving part, signal measurement part, and analyzing part.

2.1 Light Emitting Part

The component of the hydrocarbon of oil had strong fluorescence absorbance at the wavelength of about between 600*nm* and 900*nm* [9-10]. The halogen lamp provides the necessary spectral response for emission at between 200*nm* and 900*nm* [11]. So halogen light exhibits high absorption in

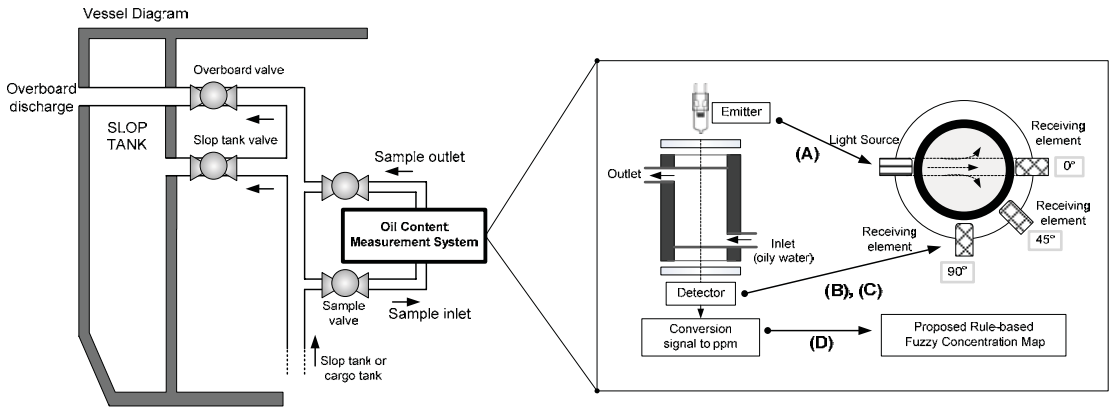


Figure 1: Schematic of fluorescence measurement system,
(A) light emitting part, (B-C) light receiving and signal measurement parts, (D) analyzing part.

ISBN: 978-989-95079-6-8

oil, system prototype was to use halogen lamp as the source of light.

2.2 *Light Receiving Part*

Since the required photo-detector of oily water should detect wavelengths about between 600*nm* and 1000*nm*, the most suitable sensor was found to be the photodiode that spectral response range is between 320*nm* and 1100*nm*. As in the case of the system prototype, voltages are being measured. Since photodiodes generate low current, it is connected to a resistor to transform the measured current to voltage as shown in Fig. 4. And resistor should have a large value in order to attain high voltages the data acquisition board, because there is no amplifier in this system. The transmitted and scattered fluorescence intensities are measured. Each sensor is positioned at 0˚, 45˚, 90˚ as shown in Fig. 1.

2.3 *Signal Measurement Part*

The signal sources of photodiodes are floating or nonreferenced signal source [12]. A floating signal source is not connected to the building ground system, but has an isolated ground-reference point. For examples of floating signal sources are outputs of thermocouples, battery-powered devices, optical isolators, and isolation amplifiers. Since the voltage level of the floating source can move, data acquisition board then saturates, causing erroneous readings. So resistors are used as illustrated in Fig. 3. These resistors, called bias resistors, provide a DC path. V_m is the measurement voltage. In this experiment, light oil samples were prepared with two types, 25*ppm* and 1000*ppm*. The data acquisition were being measured 30 seconds, and sampling rate of 2*kHz*.

2.4 *Analyzing Part*

The experimental results are shown in Fig. 4-5. There is a significant disparity between transmitted and scattered. It shows that the more oil content increases, the more transmitted intensity is decreased. And scattered intensities are increased. Given experimental results, the component of the oil contamination may be seen as absorbing and scattering wavelength based on the hydrocarbon. Although there is not a stark contrast between samples with marked oil content, data-sets have fluorescence patterns to classify. So attention was directed to conversion from fluorescence intensity to oil content using proposed rule-based fuzzy concentration map.

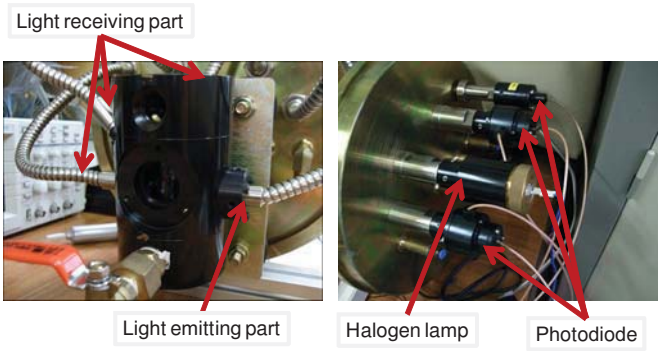


Figure 2: System prototype; Left: front of system, optical fibre to the sensors; Right: back of system, halogen lamp and photodiodes.

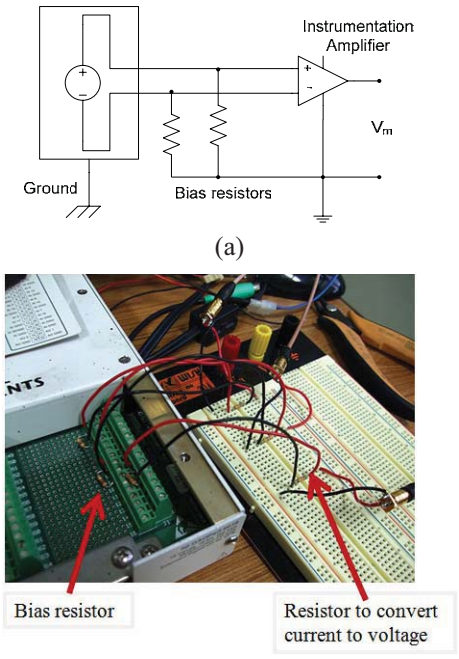


Figure 3: (a) Schematic of connecting a floating signal source, (b) Practical application.

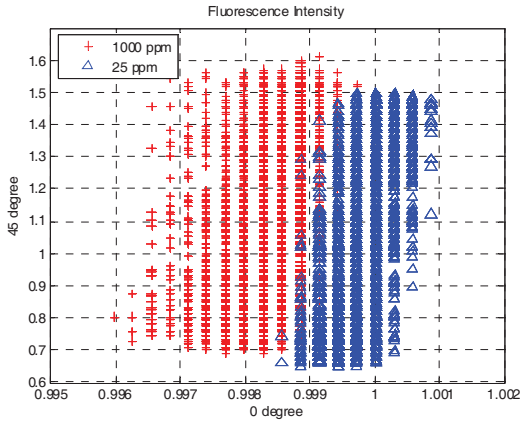


Figure 4: Intensity variations of transmitted (0°) and scattered (45°) in light oil.

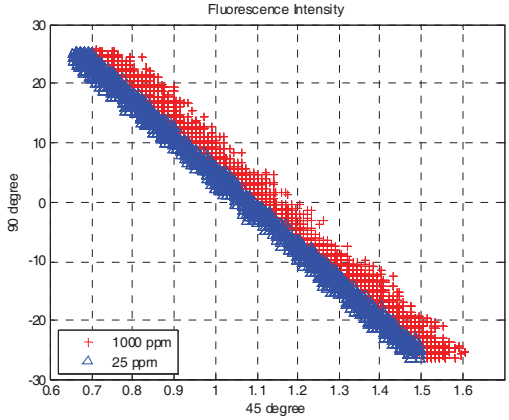


Figure 5: Intensity variations of scattered (45° and 90°) in light oil.

ISBN: 978-989-95079-6-8

3 Rule-based Fuzzy Concentration Map

Proposed algorithm consist of two variables of fuzzy inference system`s output and one fusion parameter as shown in Fig. 6. ρ_t, ρ_c, ρ_f are predicted oil content, λ is fusion effect ratio, and s_0, s_{45}, s_{90} are optical sensor signals that is positioned 0˚, 45˚, 90˚. The hypothesized relationship between transmitted intensity and scattered intensity is written:

$$\rho_f = (1-\lambda)\rho_t + \lambda\rho_c + \varepsilon \tag{1}$$

where ρ_t, ρ_c denote the independent variables of sub predicted oil content of ANFIS output, ρ_f denotes the final estimated oil content, λ is the fusion effect ratio, and ε is the noise term reflecting other factors that influence content.

The following sections provide proposed structure, fuzzy concentration map and multiple regression analysis.

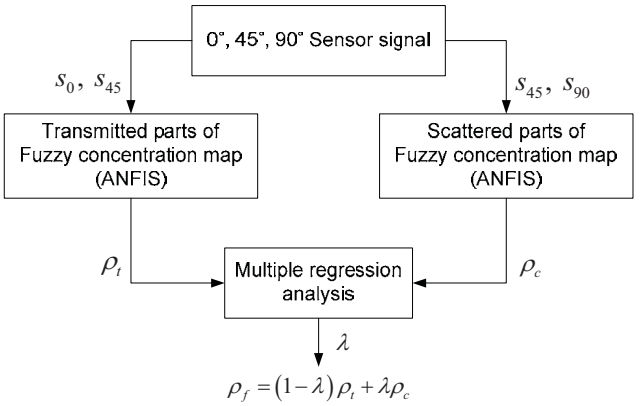


Figure 6: Proposed rule-based fuzzy concentration map.

3.1 Fuzzy Concentration Map

Proposed rule-based fuzzy concentration map presents a modeling method for oil content measuring in the form of ANFIS [13,14]. ANFIS(adaptive neuro-fuzzy inference system) represent TSK(Takagi-Sugeno-Kang) fuzzy models as shown in Fig. 7(a). ANFIS are adaptive networks that are functionally equivalent to TSK fuzzy inference system as shown in Fig. 7(b). *Prod* is the product of all incoming signal or *T*-norm operator that *AND* can be used. *Norm* is the normalized firing weights. The fuzzy concentration map shown in Fig. 7(c) classifies the input space of sensor signal into *n* regions which are assigned a labelled class of oil content. Every *n* region is an adaptive node with a node equation (2).

$$O_i = \bar{w}_i f_i = \bar{w}_i \left(p_i x + q_i y + r_i\right) \tag{2}$$

where $\bar{w}_i$ denotes the normalized firing weight from *i*, and p_i, q_i, r_i are the parameter set of *i* region. These are referred to as consequent parameters.

The fuzzy concentration map predicts oil content of transmitted and scattered part, which computes the overall output as the summation of all *n* regions in (3).

$$\rho_{ppm} = \sum_i O_i = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{3}$$

The ANFIS can be trained by hybrid learning algorithm with combining back propagation and least squares method. In the forward pass the algorithm uses least squares method to identify the consequent parameters. In the backward pass the errors are propagated backward and the premise parameters are updated by gradient descent.

In this paper, proposed method consists of two ANFIS. The two ANFIS have training dataset that is (0˚ sensor, 45˚ sensor, target *ppm*) and (45˚ sensor, 90˚ sensor, target *ppm*). Training was made using MATLAB software.

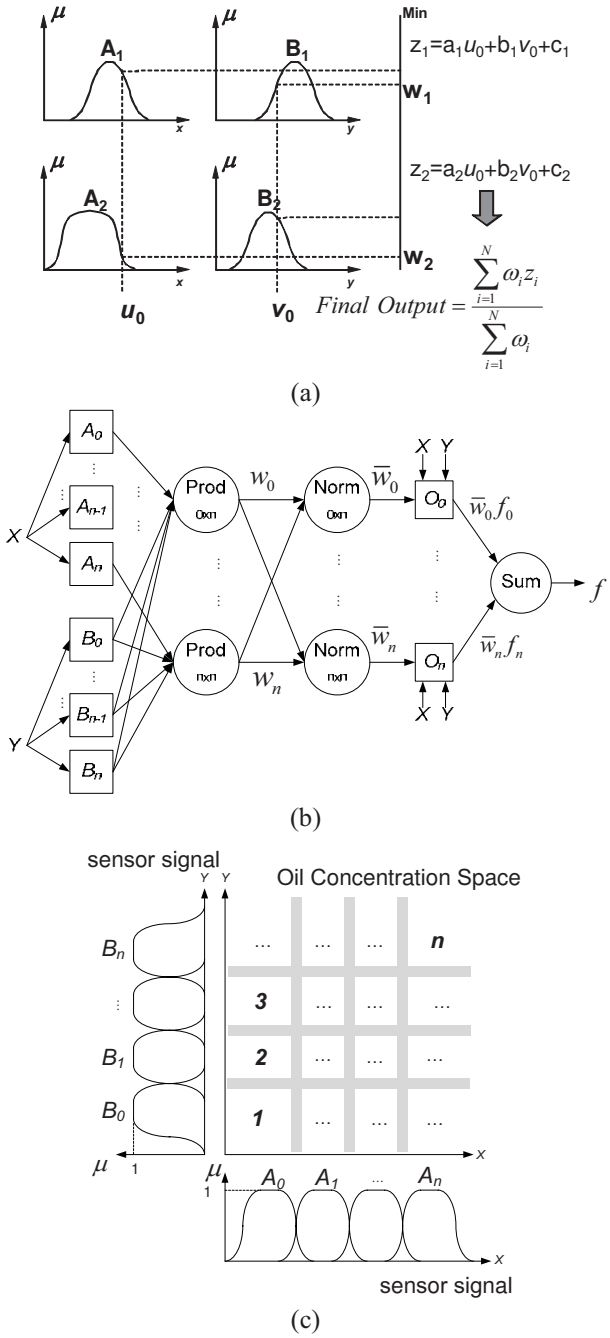
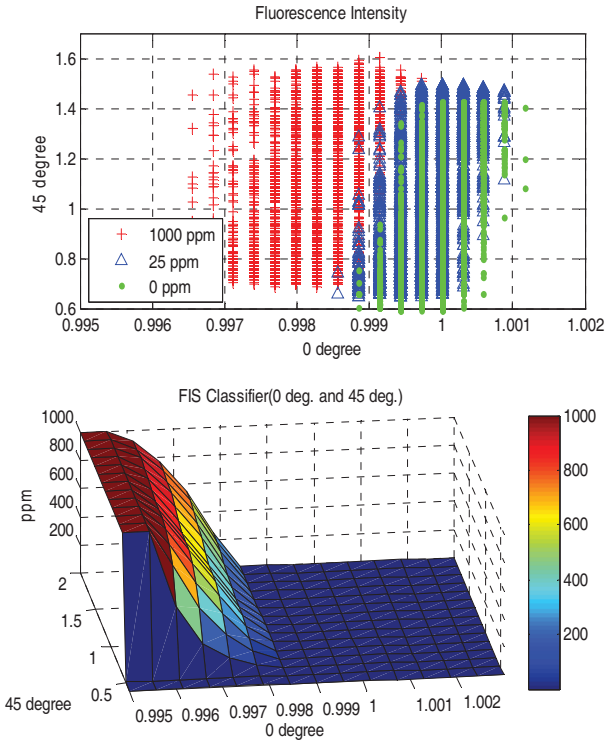
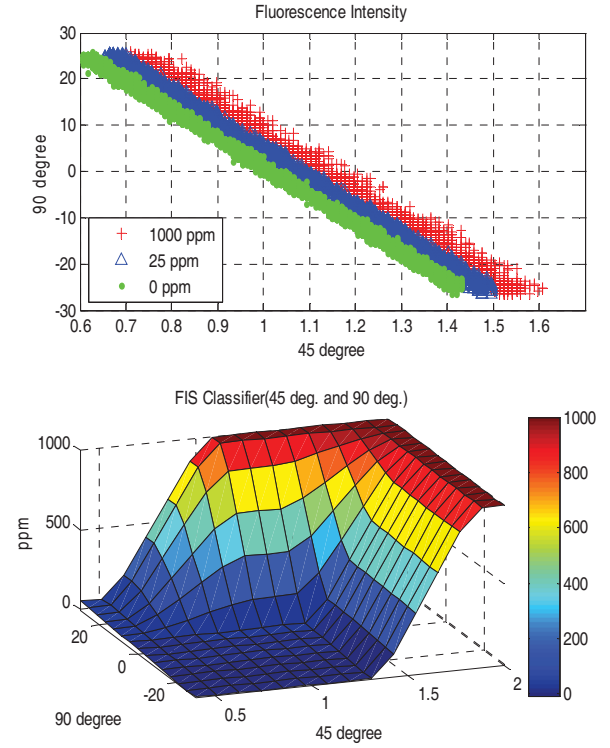


Figure 7: (a) Representation of TSK fuzzy reasoning, (b) Equivalent ANFIS Architecture, (c) Feature space partitioning of oil content.

ISBN: 978-989-95079-6-8



(a)



(b)

Figure 8: Intensity variations and FIS output surfaces; (a) Experimental results and classifier of 0° and 45°, (b) Experimental results and classifier of 45° and 90°.

3.2 Multiple Regression Analysis

Regression analysis is a statistical method for the investigation of relationships between variables [15]. Proposed rule-based fuzzy concentration map is similar to multiple linear regression analysis. Regression analysis is a statistical method for the investigation of relationships between variables. A dependent variable is called response variable or measurement, one or more independent variables are called explanatory variables or predictors in regression analysis. For example, a line in a two dimensional or two-variable space is written:

$$Y = a + bX \tag{4}$$

where a denote term of a constant or intercept, b is a slope as the regression coefficient or B coefficient.

The multiple regression procedures will be estimated in (5).

$$Y = a + b_1 X_1 + b_2 X_2 + \cdots + b_n X_n \tag{5}$$

The purpose of multiple regressions is to learn more about the relationship between several independent or predictor variables and a dependent or criterion variable [16]. The computational problem that needs to be solved in multiple regression analysis is to fit a straight line to number of points. When there is more than one independent variable, the regression line cannot be visualized in the two dimensional space, but can be computed. With n explanatory variables, multiple regression analysis will estimate the equation of a hyper-plane in n-space such that the sum of squared errors has been minimized.

Given experimental data, it shows that the more oil content increases, the more transmitted intensities are decreased, and scattered intensities are increased. Multiple regression analysis provides an update on relation of intensity in (6).

$$\begin{aligned} \rho_f &= b_1 \rho_t + b_2 \rho_c \\ &= (1-\lambda)\rho_t + \lambda \rho_c, \qquad \lambda = 1 - e^{-\eta \bar{\rho}} \end{aligned} \tag{6}$$

where ρ_t, ρ_c denote the independent variables of sub predicted oil content of ANFIS output, ρ_f denotes the final estimated, λ is the fusion effect ratio, η is the gradient scale, and $\bar{\rho}$ is the mean of two predicted oil content.

4 ANFIS Results and Discussion

This paper has attempted to make online measurement system of oil concentration with insufficient samples. It was used samples of 25 *ppm* and 1000 *ppm*, and analyzed contrast between samples with marked oil concentration. Proposed rule-based fuzzy concentration map method was used to convert sensing signal to *ppm*. The generated FIS output surface is shown in Fig. 8. When the measurement voltages are compared to the inference of oil content, this analysis method seems reasonable. After analysis method was established, insufficient content ranges of oil sample were calcu-

ISBN: 978-989-95079-6-8

lated as shown in Table 1. While The output of ANFIS(0° and 45°) was no change in good condition, FIS(45° and 90°) was sensitive. And the more *ppm* increases, the more ANFIS(45° and 90°) increases than ANFIS(0° and 45°). Considered in this relation, λ and η of multiple regression analysis parameters should be updated.

Table 1: Predicted *ppm*(*mg/l*) based on ANFIS.

$[A, B]_x = A \le x \le B$, condition λ=0.5

0 deg.	45 deg.	90 deg.	ANFIS(*ppm*) (0˚&45˚)	ANFIS(*ppm*) (45˚&90˚)	***ppm***
[0.998, 1.001]	[0.6, 1.4]	-20	[0.951, 91.184]	[4.492, 87.651]	[2.721, 89.417]
		-10	[0.951, 91.184]	[17.571, 99.364]	[9.261, 95.274]
		0	[0.951, 91.184]	[28.592, 109.550]	[14.771, 100.367]
		10	[0.951, 91.184]	[52.605, 314.557]	[26.778, 202.871]
		20	[0.951, 91.184]	[106.429, 806.196]	[53.690, 448.690]

5 Conclusion

In this paper, we have developed online content measurement system using transmitted and scattered intensity. We attempt to outline an alternative analysis of measurement of oil content using ANFIS. Two part of fluorescence intensities can be used oil content information. Given experimental data-set, it shows that the more oil content increases, the more fluorescence absorbance and scattering of hydrocarbon (-CH) are increased. So transmitted intensity is decreased, and scattered intensities are increased. Proposed rule-based fuzzy concentration map offers the possibility of conversion sensor signal into *ppm*. Because proposed method uses hybrid learning algorithm, the proposed method can provide an input-output mapping. Proposed analysis method and sensing system provide a stepping stone for developing basic technique for designing oil content meter.

Acknowledgment

This work was supported by the Grant of the Korean Ministry of Education, Science and Technology (The Regional Core Research Program/Institute of Logistics Information Technology).

References

[1] MARPOL, How to do it, Manual of the Practical Implication of Ratifying and Implementing MARPOL 73/78, *International Maritime Organization*, London, 1993.

[2] Clarke. R.B, Marine pollution, *Oxford Science 2nd edition*, 1989.

[3] Mason, C.F., Biology of fresh water pollution, *Longman Harlow 2nd edition*, 1991.

[4] Zhu B., Yuan H., Classification of oils an water using LIF spectrum parameters independent of the thickness of oil film, *SICE '95. Proceedings of the 34th SICE Annual Conference*, pp. 1539-1541, Jul. 1995.

[5] Smith, P.R., Green, D.A., Naimimohasses, R., Some investigations on neural processing of scattered light in water quality assessment, *IEE Proceedings Vision, Image and Signal Processing*, vol. 141, no. 4, pp. 261-266, Aug. 1994.

[6] Eisenlauer, J., Horn, D., Fibre-optic online flocculant dose control in water treatment operations, *Colloids and Surfaces 25*, pp. 111-129, 1987.

[7] Snel, D., Pitt, G.D., Oil content monitoring (practical considerations), *International conference proceedings on Optical techniques and process control*, pp.27-41, 1983.

[8] Smith, P.R., Green, D.A., Naimimohasses, R., Thomason, H., Aritificial neural network processing of scattered light measurements in the detection of immiscible water pollutants, *3rd international conference on Artificial neural networks*, pp. 282-285, 1993.

[9] Haavard A., Trond S., Reidar B. S., An in-line NIR/endoscope technique for observations in real hydrocarbon multiphase systems, *Flow Measurement and Instrumentation*, vol. 16, pp. 289-293, 2005.

[10] Xin B., Liankui D., Partial least squares with outlier detection in spectral analysis: A tool to predict gasoline properties, *Fuel*, vol. 88, pp. 1216-1222, 2009.

[11] Cristina Yubero, M. Carmen Garcia, M. Dolores Calzada, Using a halogen lamp to calibrate an optical system for UV-VIS radiation detection, *Optica Applicata*, vol. 38, no. 2, pp. 353-363, 2008.

[12] National Instruments, Field wiring and noise consideration for analog signals, *NI developer zone*, Feb. 2006.

[13] Jyh-Shing R.J., ANFIS: Adaptive-Network-Based Fuzzy Inference System, *IEEE Trans. on Systems, Man and Cybernetics*, vol. 23, no. 3, 1993.

[14] Kwang H. Lee, First Course on Fuzzy Theory and Applications, *Springer*, ISBN-10: 3540229884, 2004.

[15] Rawlings, J.O., Pantula, S.G., Dickey, D., Applied Regression Analysis, *Springer Verlag*, ISBN-10: 0387984542, 1998.

[16] Timothy Z. Keith, Multiple Regression and Beyond, *Prentice Hall*, ISBN-10: 0205326447, 2006.

Features stereo matching based on fuzzy logic

Grzegorz Nieradka, Bohdan Butkiewicz
Warsaw University of Technology
00-665 Warsaw, Nowowiejska 15/19 str, Poland
Email: G.Nieradka@ise.pw.edu.pl, B.Butkiewicz@ise.pw.edu.pl

Abstract— *This paper presents an entire scheme for the estimation of a sparse disparity map from stereo pair images. In contrary to a dense disparity map, for which disparity values are calculated for each pixel of image, the sparse disparity map is determined only for some distinguished set of pixels from an image. The pixels belong to this set are called features. Therefore, in the first stage the algorithm for determining sparse disparity map has to be able to detect and specify some pixels as the feature pixels. In this article a method for specifying features is introduced. The core of the presented method relies on a fuzzy edges detector. The algorithm of fuzzy edges detection, as well as a manner enabling determination of the features' set, are introduced. The disparity is calculated at the fuzzy domain based on a similarity measure which is the correlation of fuzzy sets. The results of the obtained disparity maps for some benchmark stereo pairs, and the comparison with the well known Marr-Poggio-Grimson algorithm designed for sparse disparity map estimation are presented.*

Keywords— feature points, stereo matching, disparity map, edges detection, fuzzy logic

1 Introduction

Stereo vision is a vivid researched technique of machine vision area which allows to obtain a three dimensional image of a scene observed. Its importance arises from its main characteristics, hence, it is a passive method and the equipment which allows to use it is rather cheap. The estimation of disparity map is one of the central problems of stereo vision application.

In spite of thirty years of research there is no good enough method which always gives a satisfactory solution. The stereo vision and especially the stereo matching research could be divided into two main streams [1, 2]. One of them are algorithms which try to match all pixels from images. As a result of this kind of algorithms the so called dense disparity map is obtained. [3]. The "dense" means that the disparity values are estimated for each pixel of the images. These algorithms, however suffer from the calculation being time consuming and the matching in uniform regions of images being almost impossible. But doubtless the advantage of this group of methods is better model reconstruction possibility. This paper does not consider the methods which produce the dense disparity map.

An alternative approach towards finding a solution to the stereo vision problem is a feature matching strategy. Although, the obtained disparity map has established values in much less amount of pixels, in contrary to the former group of methods, the result can be used in many real tasks as navigation of autonomous robots, manipulation of industrial robots or view morphing. Moreover, the disparity values could be interpolated in the areas where the matching is not made. The disadvantage of this approach is an insufficient amount of disparity values for visually good reconstruction of observed scene. On the other hand, the methods based on features reduce the searching space from full image pixels set to the pixels which were recognized and assigned as the features. It attracts reduction of calculation time and in a natural way, it causes eliminating the ambiguity that may occur inside the uniform regions of images, because in these regions there are usually no feature points.

Each clearly distinct pixel from image can be selected as the features considered for matching task. However, the most popular features taken under consideration are corners [4, 5] and edges [6]. The edges matching techniques are more popular for the reason that the disparity map estimated with dependence only on corners is too sparse. Matching the edges of the images gives a lot of more dense disparity map therefore matching edges is considered in this paper.

In another group of feature based approaches image segmentation is done as the first step. Later, they try to match the segmented regions [7]. This group of methods is not considered here, and the comparison of results would not give meaningful results because the ability of estimation disparity map depends mainly on the segmentation step, which is a totally different issue in machine vision research.

In order to improve results of features stereo matching there is a big group of methods which aim at matching piecewise-linear edge segments [8, 9]. The matching of edge segments has advantages that the error of isolated pixels in edges has very little influence on the position and orientation of the edge segments. There are different modifications in the descriptions of edge segments [10] and the edges could be matched as edge chains [11]. This group of methods is out of our interest because they do not depend on an edge detector and have a different philosophy behind them than matching edges directly.

The well known algorithm for matching features is the Marr-Poggio-Grimson algorithm, MPG in short [12, 13]. This algorithm is an implementation of the human stereo vision theory that was developed by Marr and Poggio [14]. Because the MPG algorithm works mainly with reliance on the edge detection it has been used here as the reference algorithm. All the details of this algorithm are given in the Sec. 5.1.

The paper proceeds in three stages: the first gives a formal statement of the stereo matching problem; the second part contains an algorithm based on fuzzy logic for edges detection and a manner for establishing feature points as well as a method for the calculation of a disparity map; the third part of the paper presents evaluation results of the introduced method and the comparison with the MPG algorithm.

ISBN: 978-989-95079-6-8

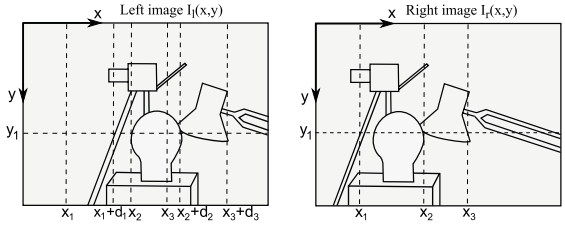


Figure 1: Depiction of the stereo matching problem

2 Problem statement

The stereo vision problem relies on establishing the correspondence between physically the same points in two different images representing the same scene. However, these two images have to be taken under some geometric restrictions in order to ensure they constitute a proper stereo pair [15, 16]. If two images constitute a proper stereo pair then there is a horizontal shift between these two images called *disparity*. Finding this shift is the main aim of almost every stereo matching algorithm. This task is illustrated in Fig. 1 which shows, that the same physical points (x_1, x_2, x_3 in the Fig. 1) in the left and the right images have different abscissa in the image coordinate system. The matching task is usually posed as the problem of finding points in the right image I_r corresponding to the selected points in the left image I_l. The value of abscissa shift of the determined point, called *disparity* is here refereed to as d. The task is to find the proper values of disparity d which correspond to the same physical points.

Usually, the epipolar constraint [15] restricts the search to a horizontal line of images. A further simplification is attained by limiting the maximum possible value of disparity d, for which a good estimate can exist. In this paper we restrict our considerations to a parallel rectified pair of cameras. This means that the searching is done for each line of images and the lines of the left and right images correspond to each other.

3 Features detector

Edge detection is a well known problem for the machine vision community. There are many more or less popular algorithms in the literature considering on image processing tasks [17, 18]. In this paper we used some new kind of an edge detector. This edge detector is based on fuzzy relation and with its characteristic feature being rather fat edges obtained as the detection results [19]. The thickness of the detected edges could be a disadvantage in some applications but in the matching problem has become an advantage, because the edges have more information about themselves and about their neighborhood.

3.1 Edges detection

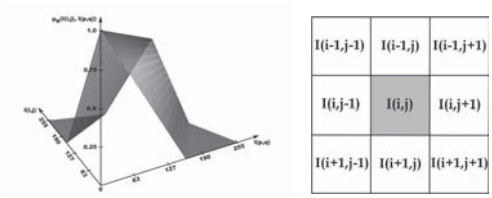
The principle behind the presented algorithm is the fact that for the pixel which belongs to an edge, the neighboring pixels have different values. Therefore if the edge is strong then in the determined small neighborhood of the pixel under consideration there are numerous pixels with the big difference of value.

The distinction between these pixels is made by using the fuzzy relation [20]. As it is pointed in [19] a lot different shapes of fuzzy relation can be used. In this paper we used the fuzzy relation in the triangular form (1). The shape of the triangular similarity relation is shown in Fig. 2(a) and its strength depends mainly on α value.

Formally the triangular fuzzy relation between the pixels $I(i,j)$ and $I(k,l))$ is expressed:

$$\mu_{RT}\left(I\left(i,j\right),I\left(k,l\right),\alpha\right)=\begin{cases}1-\frac{|I(i,j)-I(k,l)|}{\alpha} & if \quad |I\left(i,j\right)-I\left(k,l\right)|<\alpha\\ 0 & if \quad |I\left(i,j\right)-I\left(k,l\right)|\geq\alpha\end{cases} \tag{1}$$

In the edge detection stage the distinctions between one distinguished pixel $I(i,j)$ and its neighbors are calculated. If the differences are big, there are likely big changes in image, and there could be an edge present. The algorithm for edge detection is described as follows:



(a) (b)

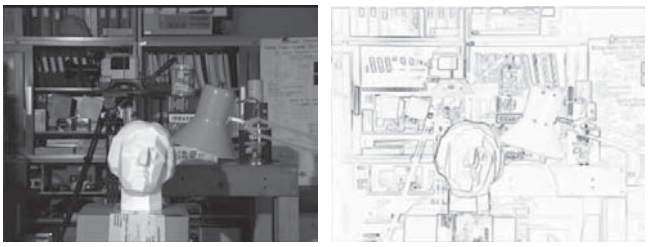
Figure 2: The shape of triangular fuzzy relation used for the edge detection (a) and the neighborhood (b) of distinguished pixel.

1. The determination of the differences between the distinguished pixel and its small established neighborhood. This situation of the distinguished pixel and its neighboring pixels is shown in Fig. 2(b). In this case we use the window size eight but many different sizes of this neighborhood can be used [19].
   The vector of homogenity description h_v is defined. The length of homogenity vector h_v is equal to the size of the neighborhood. For each pixel of the image the values of the vector h_v are calculated by (2):
   $$h_v=\bigcup_{k,l=-\rho}^{\rho}\mu_{RT}\left[I\left(i,j\right),I\left(i+k,j+l\right)\right] \tag{2}$$
2. The final response of the edge detector is defined as the sum of the homogenity vector values:
   $$h_e\left(i,j\right)=\sum h_v \tag{3}$$

If the region is uniform, the response of the edge detector would be big because all of the pixels have big similarity close to one. If the region is diverse, and if a strong edge is present in the region, the response of the detector is smaller. The results of the detector working for two images are shown in Fig. 3 and 4. We reverse the colors of the edge images in order to get the better presentation in black in white background. Originally the colors for the edges are brighter, the bigger response of the edge detector while the background is black.

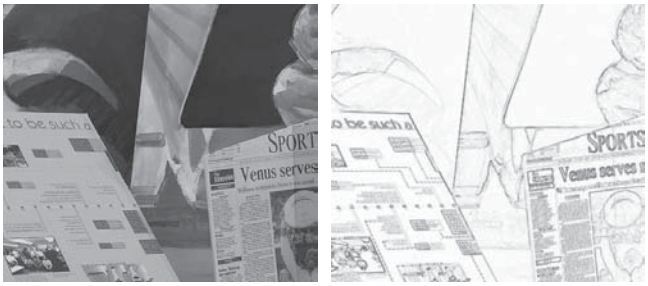
Moreover, the values of the edge detector response called "edginess" could be considered as values of some membership function. The bigger the response, the bigger the pixel's membership function value.

ISBN: 978-989-95079-6-8



(a) (b)

Figure 3: The left of stereo pair image called "Tsukuba" (a) and the result of detected edges (b).



(a) (b)

Figure 4: The left image of an another stereo pair called "Venus" (a) and the result of edges detection (b).

3.2 Edges thinning

The edges detected by the presented edges detection procedure are rather fat therefore we need to further reduce the number of points which are subject to match-making. In order to obtain the reduction of the edge points, a thinning algorithm is executed at the edge images. The thinning is done in three independent steps.

The full thinning algorithm proceeds in the following way:

1. *Thresholding:* The first step of the edge thinning is attenuation of small values by using the global threshold P_{th} with the value given by:

$$P_{th} = a_{th} \times E_{mean} \tag{4}$$

where a_{th} is an arbitrary chosen empirical threshold value. In this work it was set at the value equals 1.25. The term E_{mean} determines the average value of the edges image:

$$E_{mean} = \frac{1}{M \cdot N} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} e\left(i,j\right) \tag{5}$$

where $e\left(i,j\right)$ is the pixel value of the edge image.

2. *Horizontal non-maximum suppression:* This step has similar idea to the non-maximum suppression algorithm. But in this step, it is made only in the horizontal direction of the image, as carried out:

$$p_e\left(i,j\right) = \begin{cases} p_e\left(i,j\right) & \textbf{if} \quad p_e\left(i,j\right) > p\left(i-1,j\right) \\ & \textbf{and if} \quad p_e\left(i,j\right) > p\left(i+1,j\right) \\ 0 & \textbf{if} \quad else \end{cases} \tag{6}$$

3. *Vertical non-maximum suppression:* The last step is practically the same as the previous step, but it is done in the vertical direction of the image:

$$p_e\left(i,j\right) = \begin{cases} p_e\left(i,j\right) & \textbf{if} \quad p_e\left(i,j\right) > p\left(i,j-1\right) \\ & \textbf{and if} \quad p_e\left(i,j\right) > p\left(i,j+1\right) \\ 0 & \textbf{if} \quad else \end{cases} \tag{7}$$

As the result the images with the edges of one pixel thickness are obtained. In these images the positions of the remaining pixels determine the locations of the features which are to be matched.

The examples of the results obtained by the edge thinning algorithm and the locations of the feature points for the images considered earlier (see Fig. 3 and Fig. 4) are shown in Fig. 5.



(a) (b)

Figure 5: The feature points after edge thinning applied for the images in Fig. 3 (a) and Fig. 4 (b) respectively.

4 Feature points matching

A typical algorithm for finding a stereo matching solution works for an arbitrary range of disparity values from d_{min} to the d_{max}. There are two windows established, one of which, called a reference window is set at a stable position in one (usually left) of the images constituting a stereo pair and the other called a search window moves along the horizontal line in the other image (usually right) within the established disparity range. For each position of the search window some kind of matching measure is calculated. The value of d for which position of the search window the value of measure attains its supreme is taken as the proper value.

The presented algorithm works in the mentioned scheme, that is reference and search windows, but during the matching stage the feature image and the edges images are used. The feature image determines the position of the feature points received after using the edge thinning. It is used for establishing the position of the reference window in the reference edge image exactly at location of the feature point. After establishing the position of the reference window the matching is carried out by moving the search window in the other edge image.

The measure of matching is calculated based on the edge images. For each point which have determined as the feature point in the feature images, the calculation of matching score in the edge images is done. As it was mentioned earlier the values of the edge detector can be interpreted as the values of the membership function which gives the degree of pixel belonging to the "edginess" variable. For this reason the calculation of matching score is done in a fuzzy fashion.

ISBN: 978-989-95079-6-8

The matching measure used in this algorithm is the correlation between fuzzy sets [21]. Let us assume that there is a sequence of paired data $((\mu_A(x_1), \mu_B(x_1)), \ldots, (\mu_A(x_n), \mu_B(x_n)))$ which corresponds to the grades of the membership functions of fuzzy sets A and B defined on X. The correlation coefficient, $f_{A,B}$, between the fuzzy sets A and B is calculated as:

$$r_{A,B} = \frac{\sum_{i=1}^{n} (\mu_A(x_i) - \bar{\mu}_A)(\mu_B(x_i) - \bar{\mu}_B) / (n-1)}{S_A \cdot S_B} \quad (8)$$

where $\bar{\mu}_A$ and $\bar{\mu}_B$ denote the average membership grades of the fuzzy sets A and B according to the dependence:

$$\bar{\mu}_A = \frac{\sum_{i=1}^{n} \mu_A(x_i)}{n} \quad (9)$$

and S_A and S_B are the standard deviations of fuzzy sets A and B calculated as: $S_A = \sqrt{S_A^2}$, where S_A^2 represents the degree of variation of the membership function determined by:

$$S_A^2 = \frac{\sum_{i=1}^{n} (\mu_A(x_i) - \bar{\mu}_A)^2}{n-1} \quad (10)$$

The measure of correlation (8) is simply adapted to the measure of correlation between "edginess" in the reference window and "edginess" in the search window. As in a typical matching algorithm the score is calculated for each shift from the chosen range and the value d_k for which the correlation attains its maximal value is taken as the proper disparity value.

5 Results evaluation and comparison

For the testing purposes we used some benchmark stereo pairs made public accessible [3] from the Middlebury Stereo Vision Page [22]. The big advantage of these stereo pairs is that they have the true disparity map available. This allows to make a reasonable comparison of different algorithm in reference to the true result which should be obtained.

5.1 Reference feature matching algorithm

As the reference algorithm the Marr-Poggio-Grimson (MPG) has been selected. This algorithm is very popular, easy to implementation and gives good results. The implementation which is used here was based on the scheme described in [12, 13].

The MPG algorithm works in the coarse to fine scheme. In the description of this algorithm the word "channel" is used and it refers to image filtering with a specific filter window with established size w. The coarse channel means the filtering with the bigger size of filter window and the finest channel means that the image was filtered by using filter with smaller size.

The typical scheme of the MPG algorithm consists of six steps. The steps are outlined below:

1. *LOG filtering:* Started from the coarsest channel the left and right image of stereo pair is convolved with LOG filter. The LOG is a smoothed second derivative of the image signal and assumes the following form:

$$\nabla^2 (G(x,y)) = \left(\frac{x^2+y^2}{\sigma^2}\right) \exp\left(\frac{-(x^2+y^2)}{2\sigma^2}\right) \quad (11)$$

where ∇^2 is the Laplacian $\nabla^2 = \left(\frac{\partial^2}{\partial x^2}\right) + \left(\frac{\partial^2}{\partial y^2}\right)$ and $G(x,y))$ is the Gaussian function, which acts as a low-pass-filter to the image:

$$G(x,y) = \sigma^2 \exp\left(\frac{-(x^2+y^2)}{2\sigma^2}\right) \quad (12)$$

where the width of the channel "w" is related to σ as follows: $w = \sqrt{2}\sigma$

2. *Zero crossing extraction:* In the filtered image the detection of zero crossing is done by scanning the image horizontally for adjacent elements of opposite sign or for three horizontally adjacent elements, where the middle one is zero. Also the sign of zero crossing is searched. The located position of zero crossing with their signs are remembered.

3. *Matching:* For each scan line in the left image is centered reference window at the founded zero crossing point as the possible candidate to matching. In the right image the search window is moved in established disparity range. If in the right image is a zero crossing point on the search window and has the same sign as the left point, then this zero-crossing point produces a match. The disparity, difference between locations of zero crossing points in the left and right image, is stored in a dynamic buffer.

   Based on the matching process, the left zero crossing is marked as:

   - *unique match*, if only one right image zero crossing is matched with the left one
   - *multiple matches*, if more than one match is found
   - *no match*, if no match is found

4. *Disambiguation:* In this step the disparity map is checked for possible double matches. This kind of ambiguity is resolved by checking the disparities within the same region of the representation at the previous channel. If there is no coarser channel, or there is no a disparity value within this region in the coarser channel, or the disparity is not consistent with coarser level disparity, then the disparity is discarded. The disparity can be approximated by taking the average of the multiple match.

5. *Loop:* When the final map disparity for the current channel has been completed, the process return to the convolution to the next finer representation.

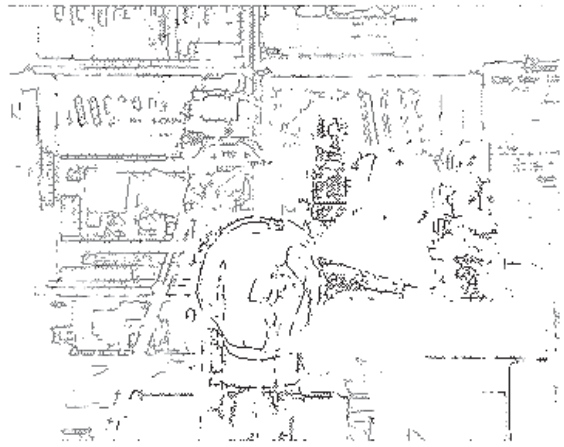
6. *Consistency:* After calculation of disparity in each channel has been completed, there is one final test done. Each disparity value at the finest channel is tested for consistency by checking its value with values at the coarser channel. If the values are inconsistent the matching is eliminated.

In this paper we used, as in the Grimson [12] implementation, four channels with size of 5, 9, 13 and 17 pixels respectively.

ISBN: 978-989-95079-6-8

5.2 *Results evaluation*

The algorithm proposed here and the MPG algorithm were applied for two different stereo pairs. The Fig. 6 presents the estimated disparity maps for "Tsukuba" stereo pair. Similar as in the edge images the colors were reversed for better visual presentation. If the color at the disparity map is darker the original disparity value has bigger value, and the element has bigger disparity value. The second image Fig. 7 presents the disparity maps obtained for the "Venus" stereo pair images.



(a) The result of the proposed algorithm.



(b) The result of the MPG algorithm.

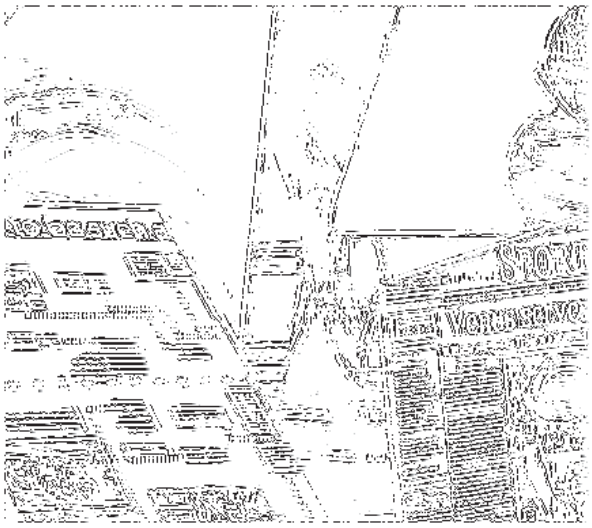
Figure 6: Disparity maps for the stereo pair "Tsukuba" (a) estimated by the proposed fuzzy algorithm and the MPG (b).

Visually these results are not so different. In the "Tsukuba" stereo pair can be seen that MPG algorithm gives some different matching but the comparison is very hard to done. Just comparison of numerical statistic given in the Tab.1 gives better view at obtained results. It can be seen that the amount of points classified as the features points is similar in both algorithms but the proposed algorithm based on fuzzy edge detection gives much more true matches. During generation of the results enclosed in Tab. 1 a point was classified as the proper matching point if the difference between its disparity and the true disparity taken from the true disparity map was equal zero or one. In this case the small mistake equal one was allowed. If the difference between estimated disparity value and the true value was bigger than one the matching point was classified as wrongly matched point.

In order to check the characteristics of these algorithms in some more deep manner the second experiment was done. In this case their behavior in the presence of noise in images was investigated. For this reason some amount of Gaussian noise was added to the both images consist of stereo pair. The amount of added noise is determined in the logarithmic scale and was established at $30dB$ value. The value denotes the amount of noise and is calculated by using expression:



(a) The result of the proposed algorithm



(b) The result of the MPG algorithm

Figure 7: Disparity maps for the stereo pair "Venus" (a) estimated by the proposed fuzzy algorithm and the MPG (b).

$$SNR = 10 \cdot \log_{10}\left(\frac{P_{image}}{P_{noise}}\right) \tag{13}$$

For discrete image I with size $M \times N$ and intensities $I(i,j)$ of pixels the power P_{image} is expressed as following:

$$P_{image} = \frac{1}{M \cdot N}\sum_{i,j=1}^{M,N} I^2(i,j) \tag{14}$$

For the purpose to produce the image with assigned SNR to the image the Gaussian noise was added iteratively and the power of noise was calculated for the noise image:

$$I_{noise} = I_{distorted} - I_{original} \tag{15}$$

ISBN: 978-989-95079-6-8

Table 1: The quantitative results of comparison for matching algorithms.

Image	Points	Right	Bad	Right %	Bad %
Results obtained by the fuzzy matching					
Tsukuba	11579	10392	1187	89.8	10.2
Venus	18588	16479	2109	88.7	11.3
In the presence of noise in image; $SNR = 30dB$					
Tsukuba	11898	2629	9269	22.1	77.9
Venus	19269	3313	15956	17.2	82.8

Results obtained by the Marr-Poggio-Grimson algorithm					
Tsukuba	11579	9549	2030	82.4	17.6
Venus	19143	12449	6694	65.1	34.9
In the presence of noise in image; $SNR = 30dB$					
Tsukuba	12345	582	11763	4.7	95.3
Venus	22705	564	22141	2.5	97.5

where the power $I_{original}$ had been calculated (13) before adding noise and the P_{noise} was calculated for the noised image. The amount of noise had been increasing to the moment for which the SNR attained the value equal to $30dB$.

As it is shown in Tab.1 both algorithms have very small resistance at the noise. The number of features does not decrease but the the amount of the proper matching is decreasing very significantly. But the proposed fuzzy approach still give better results.

6 Conclusion

In this paper the full scheme for features stereo matching based on fuzzy edge detection has been presented. The application of fuzzy logic allowed to obtain results better than the good and stable MPG algorithm.

However, both of these algorithms are very sensitive to the noise present in the images under matching consideration. It seems that the most sensitive stage is the edge detection. This aspect of these algorithms should be researched more deeply because the noise is very important in the real applications of stereo vision. In the real world numerous kind of distortions should be considered and this topic should be examined more carefully.

The obtained results are probably also image depended. In the case where the image is more complicated and there are many more edges in an image, as in "Tsukuba" stereo pair, the number of properly matched points is always bigger. In the "Venus" stereo pair, in which long straight edges are present, both algorithms found much more points classified as the feature points but the matching in both cases was worse. In that kind of images probably the algorithms based on segments edge matching would give better results. But when the image encloses a big number of short edges algorithms presented here should have better properties. This is probably a clue for further research but it is out of scope of this paper at this moment.

References

[1] M.Z. Brown, D. Burschka, and G.D. Hager. Advances in computational stereo. *IEEE Trans. on Pattern Analys. and Machine Intell.*, 25(8):993–1008, 2003.

[2] U.R. Dhond. Structure from stereo - a review. *IEEE Transc. on Systems, Man and Cybernetics*, 19(6):1489–1510, 1989.

[3] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Computer Vision*, 47(1/2/3):7–42, 2002.

[4] F. Tiesheng, N. Bing, W. Qingsong, W. Tao, and Q. Dapeng. Novel stereo matching method on multi-scale harris corner points. *IEEE Proc. of the Int. Symposium on Comp. Science and Computat. Techn.*, pages 167–170, 2008.

[5] D. Zhao, D. Liu, and Y. Yang. A new stereo matching method based on sub-pixel corner detection. In *IEEE Proc. of the Int. Conf. on Comp. Science and Soft. Eng.*, pages 899–902, 2008.

[6] X. Su and T.M. Khoshgoftaar. A progressive edge-based stereo correspodence method. In *Proc. of the Int. Symposioum on Visual Comput. (ISVC)*, volume 1, pages 248–257, 2007.

[7] K. Kyung-Yoo, K. Sung-Yeol, and H. Yo-Sung. Stereo matching using global and local segmentation. In *Proc. of the Int. Technical Conf. on Circuts/Systems, Comp. and Communicat. (ITC-CSCC'07)*, pages 1225–1226, 2007.

[8] H. Bay, V. Ferrari, and L. Van Gool. Wide-baseline stereo matching with line segments. In *Proc. of the IEEE Comp. Society Conf. on Comp. Vision and Pattern Recogn. (CVPR'05)*, 2005.

[9] A. Luo, W. Tao, and H. Burkhardt. A new multilevel line-base stereo vision algorithm based on fuzzy techniques. In *Proc. of the Int. Conf. on. Pattern Recogn. (ICPR'96)*, volume 1, page 383, 1996.

[10] X. Shenghua, L. Jiping, P. Jianping, and W. Yong. Straight line extraction for stereoimage feature matching. In *IEEE Proc. of the Congress on Image and Signal Processing*, 2008.

[11] B. Tang, D. Ait-Boudaoud, J. Matuszewski, and L. Shark. An efficient feature based matching for stereo images. In *Proc. of the Geometric Modeling and Imaging - New Trends (GMAI'06)*. IEEE, 2006.

[12] W. Grimson. Computational experiments with a feature based stereo algorithm. *IEEE Transc. on Pattern Analysis and Machine Intell.*, 7(1):17–34, 1985.

[13] D.J. Braunegg. Stereo feature matching in disparity space. *IEEE*, pages 796–803, 1990.

[14] D. Marr and T. Poggio. A computational theory of human stereo vision. *Proc. of the Royal Society of London*, B(204):301–328, 1979.

[15] O.D. Faugeras. *Three-dimensional computer vision: a geometric viewpoint*. MIT Press, Cambridge, 1993.

[16] E. Trucco E. and A. Verri. *Introductory techniques for 3-D computer vision*. Prentice Hall, New Jersey, 1998.

[17] W.K. Pratt. *Digital image processing*. John Wiley & Sons, Inc., third edition edition, 2001.

[18] R. Nevatia. *Machine perception*. Prentice-Hall, Inc., 1982.

[19] G. Nieradka. Method for edge detection in images using fuzzy relation. In *Proceedings of the SPIE*, volume 6937, pages 69373E–69373E–8, 2008.

[20] G.J. Klir and B. Yuan. *Fuzzy sets and fuzzy logic. Theory and applications.* Prentice Hall PTR, 1995. Upper Saddle River, New Jersey.

[21] D.-A. Chiang and N.P. Lin. Correlation of fuzzy sets. *Fuzzy sets and systems*, 102(2):221–226, 1999.

[22] Middlebury Stereo Vision Page. Available from: `http://vision.middlebury.edu/stereo`. [cited: 20.01.2009].

ISBN: 978-989-95079-6-8

Non-additive robust ordinal regression with Choquet integral, bipolar and level dependent Choquet integrals

Silvia Angilella Salvatore Greco Benedetto Matarazzo

Department of Economics and Quantitative Methods,
Faculty of Economics, University of Catania,
Corso Italia 55, I-95129 Catania, Italy.
Email: angisil@unict.it, salgreco@unict.it, matarazz@unict.it

Abstract— *Choquet integral has proved to be an effective aggregation model in multiple criteria decision analysis when interactions between criteria have to be taken into consideration. Recently, some generalizations of Choquet integral have been proposed to take into account more complex forms of interaction. This is the case of the bipolar Choquet integral and of the level dependent Choquet integral. To apply Choquet integral and its generalizations in decision problems it is necessary to determine one capacity permitting to represent the preferences of the Decision Maker (DM). In general the capacities are determined on the basis of some exemplary decisions supplied by the DM. It has been observed that effectively there is not only one capacity compatible with the DM's preferences, but rather a whole set of capacities. The determination of the whole set of compatible capacities and the consequent definition of proper preference relations is the domain of the non-additive robust ordinal regression. The authors have already proposed a methodology for non-additive robust ordinal regression when dealing with classical Choquet integral in ranking or choice decision problems. In this presentation, we want to give the basis of a general methodology for non-additive robust ordinal regression for Choquet integral and its generalizations (therefore also the bipolar Choquet integral and the level dependent Choquet integral) in the whole spectrum of decision problems (i.e. not only ranking and choice, but also multicriteria classification).*

Keywords— Choquet integral; Bi-Capacity; Bipolar Choquet integral; Level dependent Choquet integral; Non-additive robust ordinal regression.

1 Introduction

In the field of Multiple Criteria Decision Aid (MCDA), the main purpose of the Multi-Attribute Utility Theory (MAUT) [1] is to represent the preferences of the Decision Maker (DM) on a set of alternatives, taking in account different and conflicting points of view, called *criteria*, to get an overall utility function of every alternative.

The principal assumption underlying MAUT is the independence of criteria, not well suited to many real decision problems in which some interactions between criteria should be considered. In this last direction, Choquet integral [2] has proved to be an effective aggregation model in multiple criteria decision analysis when interactions between criteria have to be taken into consideration. Recently, according to some studies in psychology [3], many real decision problems are often based on *affect*. It seems natural to consider the criteria evaluations on a scale going from negative (bad) to positive (good) values, with a central neutral value. Consequently, the criteria are evaluated on a *bipolar* scale, *i.e.* some complex interactions among criteria arise depending on their good or positive values. To handle bipolar scales of criteria, the notion of capacity has been extended to that of bi-capacity by Grabisch and Labreuche in [4] and independently, to the bipolar capacity by Greco, Matarazzo and Słowiński in [5]. As a result, the Choquet integral has been generalized with the bipolar Choquet integral [6].

In this work we also consider a recent generalization of the Choquet integral: the level dependent Choquet integral [7]. The level dependent Choquet integral handles either unipolar scales or bipolar scales and takes into account the fact that importance of criteria depends also on the level of their evaluation.

In this paper, we propose a general framework for *non-additive robust ordinal regression* (see [8]) permitting to determine the capacities of the Choquet integral and of its above generalizations being compatible with the DM's preferences. On the basis of these sets of compatible capacities, we can define appropriate preference relations permitting to give a recommendation in the whole spectrum of decision problems (*i.e.* not only ranking and choice, but also multicriteria classification). Moreover, we propose to help the DM by identifying one capacity being the most representative among the many compatible capacities for the decision problem at hand.

The paper is organized as follows. In Section 2, we present the basic concepts relative to the Choquet integral and its generalizations, the bi-polar Choquet integral and the level dependent Choquet integral. In Section 3, the non-additive robust ordinal regression is extended to such generalizations. Section 4 contains some conclusions.

2 Choquet integral and its generalizations

2.1 Preliminary notation

In a multiple criteria decision analysis, let $X = X_1 \times X_2 \times \cdots X_n$ with $X_1, \cdots, X_i, \cdots, X_n \subseteq \mathbb{R}^n$ be the possible values taken by n criteria describing a finite set X of m alternatives. We denote every alternative $\mathbf{x} \in X$ by the evaluation vector $\mathbf{x} = (x_1, \cdots, x_i, \cdots, x_n) \in X$ and the index set of criteria by $N = \{1, \cdots, i, \cdots n\}$.

2.2 Choquet integral

A *fuzzy measure* (called also capacity) on N is a set function

$$\mu : 2^N \rightarrow [0, 1]$$

with $\mu(\emptyset) = 0$, $\mu(N) = 1$ (*boundary conditions*) and $\forall\ A \subseteq B \subseteq N,\ \mu(A) \leq \mu(B)$ (*monotonicity condition*).

ISBN: 978-989-95079-6-8

In the framework of multicriteria decision problems, the value $\mu(A)$ on the set of criteria A can be interpreted as the importance weight given by the DM to the set of criteria A.

A fuzzy measure is *additive* if $\mu(A \cup B) = \mu(A) + \mu(B)$, for any $A, B \subseteq N$ such that $A \cap B = \emptyset$.

In case of additive fuzzy measures, $\mu(A)$ is simply obtained by $\mu(A) = \sum_{i \in A} \mu(\{i\}), \forall\ A \subseteq N$. In the other cases, we have to define a value $\mu(A)$ for every subset A of N, obtaining 2^n coefficients values.

Given $\mathbf{x} \in X \subseteq \mathbb{R}^n_+$ and μ being a fuzzy measure on N, then the *Choquet integral* [2] is defined by:

$$\begin{array}{l} \mathcal{C}_\mu(\mathbf{x}) = \sum_{i=1}^n \left[x_{(i)} - x_{(i-1)}\right] \mu\left(A_i\right) = \\ = \sum_{i=1}^n x_{(i)} \Big(\mu\left(A_i\right) - \mu\left(A_{i+1}\right) \Big) \end{array}$$

where $_{(\cdot)}$ stands for a permutation of the indices of criteria such that:

$$x_{(1)} \leq x_{(2)} \leq x_{(3)} \leq ... \leq x_{(n)}, \qquad (1)$$

with $A_i = \{(i), ..., (n)\}$ where $A_{n+1} = \{\emptyset\}$ $(i = 1, .., n)$ and $x_{(0)} = 0$.

2.3 Bipolar Choquet integral

Let $\mathcal{S}(N) = \{(C, D)\ :\ C \subseteq N,\ D \subseteq N,\ C \cap D = \emptyset\}$ be the set of pairs of subsets of N.

A *bi-polar capacity*, defined in [5], is a function

$$\breve{\mu}\ :\ \mathcal{S}(N) \rightarrow [0, 1] \times [0, 1]$$

such that,

1. $\breve{\mu}(A, \emptyset) = (a, 0)$ and $\breve{\mu}(\emptyset, B) = (0, b)$, with $A, B \in \mathcal{S}(N)$ and $a, b \in [0, 1]$;
2. $\breve{\mu}(N, \emptyset) = (1, 0)$ and $\breve{\mu}(\emptyset, N) = (0, 1)$;
3. For each $(C, D), (E, F) \in \mathcal{S}(N)$, such that $C \supseteq E$ and $D \subseteq F$, we have $\breve{\mu}(C, D) = (c, d)$ and $\breve{\mu}(E, F) = (e, f)$, $c, d, e, f \in [0, 1]$, with $c \geq e$ and $d \leq f$.

The properties 1) and 2) are the *boundary conditions*, while the property 3) is the *monotonicity condition*.

Given $(C, D) \in \mathcal{S}(N)$ with $\breve{\mu}(C, D) = (c, d)$, we use the following notation, $\breve{\mu}^+(C, D) = c$ and $\breve{\mu}^-(C, D) = d$.

A *bi-capacity*, defined in [4], is a function, $\hat{\mu}\ :\ \mathcal{S}(N) \rightarrow [-1, 1]$ such that,

1. $\hat{\mu}(\emptyset, \emptyset) = 0$ and $\hat{\mu}(N, \emptyset) = 1$ and $\hat{\mu}(\emptyset, N) = -1$ *(boundary conditions)*;
2. If $C \supseteq E$ and $D \subseteq F$, then $\hat{\mu}(C, D) \geq \hat{\mu}(E, F)$ *(monotonicity conditions)*.

From each bi-polar capacity, a corresponding bi-capacity is obtained by

$$\hat{\mu}(C, D) = \breve{\mu}^+(C, D) - \breve{\mu}^-(C, D),\ \ \forall\ (C, D) \in \mathcal{S}(N).$$

Let be $_{(\cdot)}$ a permutation of the elements of N such that,

$$|x_{(1)}| \leq |x_{(2)}| \leq \ldots \leq |x_{(i)}| \leq \ldots \leq |x_{(n)}|,$$

and $|x_{(0)}| = 0$.

Let be the following two subsets of N, $A_i^+ = \{j \in N\ :\ x_j \geq |x_{(i)}|\}$ and $A_i^- = \{j \in N\ :\ x_j < 0,\ -x_j \geq |x_{(i)}|\}$.

The bi-polar Choquet integral of the positive part is defined as,

$$\mathcal{C}^+(\mathbf{x}, \breve{\mu}) = \sum_{i \in N} \Big(|x_{(i)}|\ - |x_{(i-1)}| \Big) \breve{\mu}^+(A_i^+, A_i^-).$$

Analogously, *the bi-polar Choquet integral* of the negative part is defined as,

$$\mathcal{C}^-(\mathbf{x}, \breve{\mu}) = \sum_{i \in N} \Big(|x_{(i)}| - |x_{(i-1)}| \Big) \breve{\mu}^-(A_i^+, A_i^-).$$

The bi-polar Choquet integral is defined (see [5]) as:

$$\mathcal{BC}(\mathbf{x}, \hat{\mu}) = \mathcal{C}^+(\mathbf{x}, \breve{\mu}) - \mathcal{C}^-(\mathbf{x}, \breve{\mu}).$$

$\mathcal{BC}(\mathbf{x}, \hat{\mu})$ can be also formulated in terms of bi-capacities (see [6]), as follows,

$$\mathcal{BC}(\mathbf{x}, \hat{\mu}) = \sum_{i \in N} \Big(|x_{(i)}| - |x_{(i-1)}| \Big) \hat{\mu}(A_i^+, A_i^-).$$

2.4 Some particular submodels

In the bipolar decision making setting, one of the main drawbacks is the huge number of parameters to be elicitated from the DM in order to define the bi-capacity or the bipolar capacity. In fact, since $\mathcal{S}(N)$ is isomorphic to the set of functions from $\mathbb{N}$ to $\{-1, 0, 1\}$, then $|\mathcal{S}(N)| = 3^n$. Let us remark that in this contest the non-additive ordinal regression is particularly useful, because it does not require the elicitation of all the parameters, but it is mainly based on some holistic preferences on a reference set of alternatives from which the whole set of bi-capacities or bipolar capacities compatible with the DM's preferences. Nevertheless, dealing with a model with smaller number of parameters is always useful and for this reason, in the following sections we recall some well-known bipolar submodels, that are interesting from the point of view of applications and a decomposition of the bi-polar capacities, introduced in [9], more meaningful from a DM's point of view.

2.4.1 Decomposable bi-polar measures

We define a $2-$order decomposable bi-polar measure (see [9]) such that

$$\begin{array}{rcl} \mu^+(C, D) & = & \sum_{i \in C} a^+(\{i\}, \emptyset) + \sum_{\{i,j\} \subseteq C} a^+(\{i, j\}, \emptyset) + \\ & & + \sum_{i \in C,\ j \in D} a^+(\{i\}, \{j\}) \end{array}$$

$$\begin{array}{rcl} \mu^-(C, D) & = & \sum_{j \in D} a^-(\emptyset, \{j\}) + \sum_{\{i,j\} \subseteq D} a^-(\emptyset, \{i, j\}) + \\ & & \sum_{i \in C,\ j \in D} a^-(\{i\}, \{j\}) \end{array}$$

ISBN: 978-989-95079-6-8

The above decomposition of the bi-polar capacity has a more manageable and meaningful interpretation according to the DM's preferences.

In fact, $a^{\pm}(\cdot)$ can be interpreted in the following way:

- $a^{+}(\{i\},\emptyset)$, represents the power of the criterion i by itself; this value is always positive.
- $a^{+}(\{i,j\},\emptyset)$, represents the interaction between i and j, when their values are both positive; when its value is zero there is no interaction; on the contrary, when the value is positive there is a synergy effect when putting together i and j; a negative value means that the two criteria are redundant.
- $a^{+}(\{i\},\{j\})$, represents the power of the criterion j against the criterion i, when the criterion i has a positive value and j has a negative value; this provokes always a reduction or no effect on the value of μ^{+} since this value is always non-positive.

Analogous interpretation can be applied to the value of $a^{-}(\emptyset,\{j\})$, $a^{-}(\emptyset,\{i,j\})$, and $a^{-}(\{i\},\{j\})$.

In what follows, for the sake of simplicity, we will use a_i^{+}, a_{ij}^{+}, $a_{i|j}^{+}$, instead of $a^{+}(\{i\},\emptyset)$, $a^{+}(\{i,j\},\emptyset)$, and, $a^{+}(\{i\},\{j\})$, respectively; and a_j^{-}, a_{ij}^{-}, $a_{i|j}^{-}$, instead of $a^{-}(\emptyset,\{j\})$, $a^{-}(\emptyset,\{i,j\})$, and $a^{-}(\{i\},\{j\})$, respectively.

2.5 The level dependent Choquet integral

Let us also recall a recent further generalization of the bi-capacity (see [7]): the generalized capacity. Such measures take in account the fact that importance of criteria depends also on the level of their evaluation.

In particular, we define a *generalized capacity* a function $\mu^{\mathcal{G}}: 2^N \times \mathbb{R}_+ \to [0,1]$ such that

1. for all $t \in \mathbb{R}_+$ and $A, B \subseteq N$, $\mu^{\mathcal{G}}(A,t) \leq \mu^{\mathcal{G}}(B,t)$;
2. for all $t \in \mathbb{R}_+$, $\mu^{\mathcal{G}}(\emptyset,t) = 0$ and $\mu^{\mathcal{G}}(N,t) = 1$;

We define the generalized Choquet integral of $\mathbf{x} = [x_1,\cdots,x_n] \in \mathbb{R}_+^n$, with respect to the generalized capacity $\mu^{\mathcal{G}}$ as follows:

$$\mathcal{GC}(\mathbf{x},\mu^{\mathcal{G}}) = \int_0^{+\infty} \mu^{\mathcal{G}}(A(\mathbf{x},t),t)\,dt$$

where $A(\mathbf{x},t) = \{i \in N\colon x_i \geq t\}$.

Let us remark that the generalized Choquet integral can always be written as:

$$\mathcal{GC}(\mathbf{x},\mu^{\mathcal{G}}) = \sum_{i=1}^{n} \int_{x_{(i-1)}}^{x_{(i)}} \mu^{\mathcal{G}}(A(\mathbf{x},t),t)\,dt.$$

The level dependent Choquet integral can be defined also with respect to a generalized bi-capacity.

In particular, we define a *generalized bi-capacity* a function $\mu^{\mathcal{G}}: \mathcal{S}(N) \times \mathbb{R} \to \mathbb{R}$ such that

1. for all $t \in \mathbb{R}$ and $(A,B),(C,D) \in \mathcal{S}(N)$, $A \subseteq C$, $B \supseteq D$, $\mu^{\mathcal{G}}(A,B,t) \leq \mu^{\mathcal{G}}(C,D,t)$;
2. for all $t \in \mathbb{R}$, $\mu^{\mathcal{G}}(\emptyset,N,t) = 0$ and $\mu^{\mathcal{G}}(N,\emptyset,t) = 1$;

We define the generalized bipolar Choquet integral of $\mathbf{x} = [x_1,\cdots,x_n] \in \mathbb{R}^n$, with respect to the generalized bi-capacity $\mu^{\mathcal{G}}$ as follows:

$$\mathcal{GC}(\mathbf{x},\mu^{\mathcal{G}}) = \int_0^{+\infty} \mu^{\mathcal{G}}(A^{+}(\mathbf{x},t),A^{-}(\mathbf{x}t),t)\,dt$$

where $A^{+}(\mathbf{x},t) = \{i \in N\colon x_i \geq t\}$ and $A^{-}(\mathbf{x},t) = \{i \in N\colon -x_i \geq t\}$.

2.5.1 Interval level dependent capacity

The generalized Choquet integral is quite difficult to calculate and consequently, to be applied since a capacity $\mu^{\mathcal{G}}(A,t)$ is needed for each level $t \in \mathbb{R}$. For this reason in [7], a manageable class of generalized capacities are proposed, namely the *interval level dependent capacities*. For the sake of simplicity, but without loss of the generality, in the following we consider $\mathbf{x} \in [0,1]^n$.

A generalized capacity $\mu^{\mathcal{G}}$ is defined *interval level dependent capacity* if there exist

1. $a_0, a_1, a_2, , a_{m-1}, a_m \in [0,1]$, with m a positive integer such that
$$0 = a_0 < a_1 < a_2 < \cdots < a_{m-1} < a_m = 1$$
2. m capacities $\mu_1,\cdots,\mu_m$ on N.

such that, for all $A \subseteq N$ and all $t \in [0,1]$, $\mu^{\mathcal{G}}(A,t) = \mu_j(A)$ if $t \in]a_{j-1},a_j[$, $j = 1,\cdots,m$. In this case, $\mu^{\mathcal{G}}$ is an interval level dependent capacity relative to the breakpoints $a_0, a_1, a_2, , a_{m-1}, a_m \in]0,1[$ and to the capacities $\mu_1,\cdots,\mu_m$.

Finally, we recall an useful theorem, proposed in [7], which permits to split $\mathcal{GC}(\mathbf{x},\mu^{\mathcal{G}})$ in the sum of a finite number of classical Choquet integrals, more precisely one Choquet integral for each interval $]a_{j-1},a_j[$, $j = 1,\cdots,m$.

Theorem 1 *If $\mu^{\mathcal{G}}$ is the interval level dependent capacity relative to the breakpoints $a_0, a_1, a_2, , a_{m-1}, a_m \in]0,1[$ and to the capacities $\mu_1,\cdots,\mu_m$ then for each $\mathbf{x} \in [0,1]^n$*

$$\mathcal{GC}(\mathbf{x},\mu^{\mathcal{G}}) = \sum_{j=1}^{n} \mathcal{C}_{\mu}(\mathbf{x}^j,\mu_j)$$

where $\mathbf{x}^j \in [0,1]^n$ is the vector having its elements x_i^j , $i = 1,\cdots,n$, defined as follows:

$$x_i^j = \begin{cases} 0 & \text{if} \quad x_i < a_{j-1} \\ x_i - a_{j-1} & \text{if} \quad a_{j-1} \leq x_i \leq a_j \\ a_j - a_{j-1} & \text{if} \quad x_i \geq a_j. \end{cases}$$

3 Non-additive robust ordinal regression

3.1 Sorting problems

Within the multicriteria aggregation-disaggregation framework, *ordinal regression* aims at inducing the parameters of a decision model, for example those of a utility function, which have to represent some holistic preference comparisons of the DM. Usually, among the many utility functions representing the DM's preference information, only one utility function is

ISBN: 978-989-95079-6-8

selected (for example, in the UTA method [10]). In this context we also remember some UTA like-methods within the Choquet integral framework, proposed in [11]. Since such a choice is arbitrary to some extent, recently *additive robust ordinal regression* has been proposed with the purpose of taking into account all the sets of parameters *compatible* with the DM's preference information (for more details, see the multicriteria methodologies UTA$^{\text{GMS}}$ and GRIP proposed, respectively, in [12] and [13]). Let us remark that the principles of the *additive robust ordinal regression* have been applied also in sorting problems (see the UTADIS$^{\text{GMS}}$ method [14]).

Until now, robust ordinal regression has been implemented to additive utility functions under the assumption of criteria independence. In [15], the authors have proposed a *non-additive robust ordinal regression* on a set of alternatives X, whose utility is evaluated in terms of the Choquet integral which permits to represent the interaction among criteria, modeled by the fuzzy measures, parameterizing their approach.

In [15], besides holistic pairwise preference comparisons of alternatives from a subset of reference alternatives $X' \subseteq X$, the DM is also requested to express the intensity of preference on pairs of alternatives from X' and to supply pairwise comparisons on the importance of criteria, and the sign and intensity of interaction among pairs of criteria.

In the following, we recall the binary preference relations on the set of reference alternatives defined in [15].

Let us suppose that the preference of the DM is given by a partial pre-order $\succsim$ on $X' \subseteq X$.

The preference relation $\succsim$ can be decomposed into its symmetric part $\sim$ and into its asymmetric part $\succ$, whose semantics are, respectively:

$$\mathbf{x} \sim \mathbf{y} \;\Leftrightarrow\; \mathbf{x} \text{ is indifferent to } \mathbf{y},$$
$$\mathbf{x} \succ \mathbf{y} \;\Leftrightarrow\; \mathbf{x} \text{ is preferred to } \mathbf{y}, \text{ with } \mathbf{x}, \mathbf{y} \in X'.$$

The relation on the intensity of preference on pairs alternatives is represented by a partial pre-order $\succsim^*$ on $X' \times X'$, whose semantics is: for $\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{t} \in X'$

$$(\mathbf{x}, \mathbf{y}) \succsim^* (\mathbf{z}, \mathbf{t}) \;\Leftrightarrow\; \mathbf{x} \text{ is preferred to } \mathbf{y}$$
$$\text{at least as much as } \mathbf{z} \text{ is preferred to } \mathbf{t}.$$

The following system of linear constraints synthesizes the DM's preference information expressed in the approach proposed in [15].

$$\begin{cases} \mathbf{x} \succsim \mathbf{y} \;\Leftrightarrow\; \mathcal{C}_\mu(\mathbf{x}) \geq \mathcal{C}_\mu(\mathbf{y}), \text{ with } \mathbf{x}, \mathbf{y} \in X', \\ (\mathbf{x}, \mathbf{y}) \succsim^* (\mathbf{z}, \mathbf{t}) \;\Leftrightarrow\; \mathcal{C}_\mu(\mathbf{x}) - \mathcal{C}_\mu(\mathbf{y}) \geq \mathcal{C}_\mu(\mathbf{z}) - \mathcal{C}(\mathbf{t}), \\ \text{with } \mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{t} \in X', \\ \vdots \\ \text{Constraints on the importance and interaction of criteria} \\ \vdots \\ \text{Boundary, monotonicity conditions} \end{cases}$$

The output of the approach defines a set of fuzzy measures (capacities) μ defined *compatible* with the DM's preference information if the Choquet integral, calculated with respect to it, restores the DM's ranking on X', *i.e.*

$$\mathbf{x} \succsim \mathbf{y} \;\Leftrightarrow\; \mathcal{C}_\mu(\mathbf{x}) \geq \mathcal{C}_\mu(\mathbf{y}) \;\; \forall \mathbf{x}, \mathbf{y} \in X'.$$

Moreover, using linear programming, our decision model establishes two preference relations:

- for any $\mathbf{r}, \mathbf{s} \in X$, the *necessary* weak preference relation $\succsim^N$, if for all *compatible* fuzzy measures the utility of $\mathbf{r}$ is not smaller than the utility of $\mathbf{s}$, *i.e.* $\mathbf{r} \succsim^N \mathbf{s} \;\Leftrightarrow\; \mathcal{C}_\mu(\mathbf{r}) \geq \mathcal{C}_\mu(\mathbf{s})$,
- for any $\mathbf{r}, \mathbf{s} \in X$, the *possible* weak preference relation $\succsim^P$, if for at least one *compatible* fuzzy measure the utility of $\mathbf{r}$ is not smaller than the utility of $\mathbf{s}$, *i.e.* $\mathbf{r} \succsim^P \mathbf{s} \;\Leftrightarrow\; \mathcal{C}_\mu(\mathbf{r}) \geq \mathcal{C}_\mu(\mathbf{s})$.

Since as it is shown in literature, bi-capacities are the right tools to represent many decision bipolar behaviors, in this work we suggest to extend the *non-additive robust ordinal regression*, described in [15], to the bipolar decision setting and to the level dependent Choquet integral.

However, the greater flexibility of decision strategies with the criteria on a bipolar scale and level dependent Choquet integral is offset by the huge number of parameters to be elicitated by the DM and the not easy interpretation of the bi-capacities and level dependent capacity for the DM.

Let us remark that in our approach the DM is not compelled to give preference information on all criteria.

Moreover, not all criteria are compulsory on a bipolar scale; some criteria could be on the usual unipolar scale (see for a result in this topic the concept of partially symmetric bi-capacities, introduced in [16]).

The set of constraints relative to the bi-polar Choquet integral is as follows:

2) $$\begin{cases} \mathbf{x} \succsim \mathbf{y} \;\Leftrightarrow\; \mathcal{BC}(\mathbf{x}, \hat{\mu}) \geq \mathcal{BC}(\mathbf{y}, \hat{\mu}), \text{ with } \mathbf{x}, \mathbf{y} \in X', \\ (\mathbf{x}, \mathbf{y}) \succsim^* (\mathbf{z}, \mathbf{t}) \;\Leftrightarrow \\ \mathcal{BC}(\mathbf{x}, \hat{\mu}) - \mathcal{BC}(\mathbf{y}, \hat{\mu}) \geq \mathcal{BC}(\mathbf{z}, \hat{\mu}) - \mathcal{BC}(\mathbf{t}, \mu), \\ \text{with } \mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{t} \in X', \\ \vdots \\ \text{Constraints on the bipolar measures,} \\ \vdots \\ \text{Boundary, monotonicity conditions.} \end{cases}$$

The first set constraints are interpreted in the same way of the first set of constraints of system **1)**, with the only difference that every alternative is evaluated by the bi-polar Choquet integral.

Concerning the constraints on the bipolar measures, we suggest to adopt the decomposition explained in subsection 2.4.1. Such decomposition is more easy for the DM to be interpreted and to be elicitated.

The set of constraints relative to the level dependent Choquet integral is as follows:

3) $$\begin{cases} \mathbf{x} \succsim \mathbf{y} \;\Leftrightarrow \\ \mathcal{GC}(\mathbf{x}, \mu^{\mathcal{G}}) \geq \mathcal{GC}(\mathbf{y}, \mu^{\mathcal{G}}), \text{ with } \mathbf{x}, \mathbf{y} \in X', \\ (\mathbf{x}, \mathbf{y}) \succsim^* (\mathbf{z}, \mathbf{t}) \;\Leftrightarrow \\ \mathcal{GC}(\mathbf{x}, \mu^{\mathcal{G}}) - \mathcal{GC}(\mathbf{y}, \mu^{\mathcal{G}}) \geq \mathcal{GC}(\mathbf{z}, \mu^{\mathcal{G}}) - \mathcal{GC}(\mathbf{t}, \mu^{\mathcal{G}}), \\ \text{with } \mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{t} \in X', \\ \vdots \\ \text{Constraints on the level dependent capacity.} \end{cases}$$

ISBN: 978-989-95079-6-8

In this case, we adopt the interval level dependent capacity defined in Section 2.5.1, since it is more easy for the DM to express some ordinal constraints on such capacities. For example, if $X \subseteq [0,1]^n$ and we split the interval $]0,1[$ in two equal subintervals $I_1 =]0, \frac{1}{2}[$, $I_2 =]\frac{1}{2}, 1[$, DM's preference statement could be that criterion i is more important if the level of evaluation of every alternative is in I_2 than if it is in I_1.

We denote the system of constraints (types **1)**, **2)** and **3)**) on the reference alternatives X' by $E_\varepsilon(X')$.

Since linear programming is not able to handle strict inequalities in $E_\varepsilon(X')$, we put the constraints in the form of weak inequalities, by adding a small arbitrary positive value ε (see [17] for a result on this topic).

$$\textbf{4)} \begin{cases} U(\mathbf{x}) \geq U(\mathbf{y}) + \varepsilon, \text{ with } \mathbf{x}, \mathbf{y} \in X', \\ U(\mathbf{x}) - U(\mathbf{y}) \geq U(\mathbf{z}) - U(\mathbf{t}) + \varepsilon, \\ \text{with } \mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{t} \in X', \\ \vdots \\ \text{Constraints on the bipolar measures,} \\ \vdots \\ \text{Boundary, monotonicity conditions,} \end{cases}$$

where U stands for the aggregation operator than can be the bipolar Choquet integral or the level dependent Choquet integral. A similar approach can be used with the bipolar level dependent Choquet integral.

Once the decision making setting on the set of reference alternatives X' is chosen by the DM (types **1)**, **2)** and **3)**), two different optimization problems arise to establish a *necessary* and *possible* preference relation for any $\mathbf{r}, \mathbf{s} \in X$:

$$\begin{array}{ll} \max & \varepsilon \\ \text{s.t.} & E_\varepsilon^{X'} \quad \text{plus the constraint } U(\mathbf{s}) \geq U(\mathbf{r}) + \varepsilon \end{array} \tag{2}$$

and

$$\begin{array}{ll} \max & \varepsilon \\ \text{s.t.} & E_\varepsilon^{X'} \quad \text{plus the constraint } U(\mathbf{r}) \geq U(\mathbf{s}). \end{array} \tag{3}$$

If the problem (2) finds a solution with $\varepsilon \leq 0$, then $U(\mathbf{r}) \geq U(\mathbf{s})$ for all compatible sets of fuzzy measures, that implies $\mathbf{r} \succsim^N \mathbf{s}$ with $\mathbf{r}, \mathbf{s} \in X$.

On the contrary, if a positive ε solves the linear program indicated in (3), then there exists at least one *compatible* fuzzy measure such that $U(\mathbf{r}) \geq U(\mathbf{s})$, that implies $\mathbf{r} \succsim^P \mathbf{s}$ with $\mathbf{r}, \mathbf{s} \in X$.

3.2 The most representative value function

To consider the whole set of fuzzy measures compatible with the preference expressed by the DM reduces arbitrariness in the decision process. However to take into account one specific fuzzy measure can help the DM in understanding the decision process. As already proposed for robust ordinal regression with additive value functions [18], the most representative fuzzy measure is that one which better represents the necessary ranking maximizing the difference of evaluations between alternatives for which there is a preference in the necessary ranking. As secondary objective, one can consider minimizing the difference of evaluations between actions for which there is not a preference in the necessary ranking.

This comprehensive "most representative" value function can be determined through the following procedure:

1. Determine the necessary and the possible rankings in the considered set of actions.
2. For all pairs of alternatives $(\mathbf{x}, \mathbf{y})$, such that $\mathbf{x}$ is necessarily preferred to $\mathbf{x}$, add the following constraints to the linear programming constraints of types **1)**, **2)** or **3)**: $U(\mathbf{x}) \geq U(\mathbf{y}) + \varepsilon$.
3. Maximize the objective function ε.
4. Add the constraint $\varepsilon = \varepsilon^*$, with $\varepsilon^* = \max \varepsilon$ in the previous point, to the linear programming constraints of robust ordinal regression of types **1)**, **2)** and **3)**.
5. For all pairs of actions $(\mathbf{x}, \mathbf{y})$, such that neither $\mathbf{x}$ is necessarily preferred to $\mathbf{y}$ nor $\mathbf{y}$ is necessarily preferred to $\mathbf{x}$, add the following constraints to the linear programming constraints of of types **1)**, **2)** and **3)** and to the constraints considered in above point 4): $U(\mathbf{x}) - U(\mathbf{y}) \leq \delta$ and $U(\mathbf{y}) - U(\mathbf{x}) \leq \delta$.
6. Minimize the objective function δ.

3.3 Sorting problems

In this section, we briefly illustrate the principles of nonadditive robust ordinal regression to sorting problems in case of nonadditive value function represented by Choquet integral or some of its generalizations.

In MCDA, the sorting problem consists in the assignment of m alternatives of a finite set X into $h = 1, \cdots, k$ homogeneous classes $C_1, \cdots C_h, \cdots C_k$, which are increasingly ordered with respect to preference, i.e. all the elements in the class C_h have a better evaluation than the elements in class C_{h-1}. Let $b_1, b_2, \cdots, b_h$ be a set of thresholds relative to the h classes. We have that $\mathbf{x}$ belongs to class h if $U(\mathbf{x}) \geq b_{h-1}$ and $U(\mathbf{x}) < b_h$, where U stands for the aggregation operator than can be the Choquet integral or one of its generalizations. In this case, the DM's preference information will consist in the assignment of some reference alternatives to some classes. Taking into account the Choquet integral, the proposed approach determines the sets of pairs $(\mu, \mathbf{b})$, where μ is a capacity and $\mathbf{b} = [b_1, b_2, \cdots, b_h]$, compatible with the DM's assignment of the reference alternatives [19]. An alternative $\mathbf{x} \in X$ is said that *possibly belongs to class* C_h if there is at least one pair $(\mu, \mathbf{b})$ for which $\mathcal{C}_\mu(\mathbf{x}) \geq b_{h-1}$ and $\mathcal{C}_\mu(\mathbf{x}) < b_h$. One can deal analogously with the above mentioned extensions of Choquet integral.

As proposed in [20] with respect to robust ordinal regression for sorting problems based on additive value function through the UTADIS$^{\text{GMS}}$ method [14], also in this case we can help the DM with the concept of the "most representative" value function.

The idea is to select among compatible fuzzy measure that one which better highlights the possible sorting which is considered the most stable part of the robust sorting obtained by UTADIS$^{\text{GMS}}$. Thus it is selected the fuzzy measure that maximizes the difference of evaluations between alternatives for

ISBN: 978-989-95079-6-8

which the intervals of possible sorting are disjoints. As secondary objective, one can consider maximize the minimal difference between values of actions **x** and **y**, such that for any compatible fuzzy measure, **x** is assigned to a class not worse than the class of **y**, and for at least one compatible value function, **x** is assigned to a class which is better than the class of **y**. In case there is still more than one such fuzzy measure, the most representative fuzzy measure minimizes the maximal difference between values of alternatives being in the same class for all compatible fuzzy measures, and between values of alternatives for which the order of classes is not univocal.

4 Conclusions

Nonadditive value functions and bipolar decision making are one of the underpinning directions of research within MCDA, as many studies have shown that in a multicriteria decision problem the criteria under consideration could be on a bipolar scale. In such context, we have proposed a multicriteria methodology taking inspiration from some recent approaches based on the principle of the *robust ordinal regression*: the UTA$^{\text{GMS}}$ [12] and GRIP [13] in the context of choice and ranking problems, the UTADIS$^{\text{GMS}}$ method relative to sorting problems [14]. On this basis, with the aim of representing interactions between criteria, we proposed the *non-additive robust ordinal regression* which consists in the extension of the idea of robust ordinal regression to non-additive decision. More precisely, the Choquet integral and its generalizations have been adopted as utility function in different decision problems such as ranking, choice and sorting (for the specific case of Choquet integral applied to ranking and choice problems see [15]). The non-additive robust ordinal regression seems very useful because it permits to take into account the whole set of capacities compatible with the DM's preferences which are expressed through very simple questions such as:

- "is the representative alternative a better than the representative alternative b"?
- "to what class the representative alternative a belongs"?

References

[1] R.L. Keeney and H. Raiffa. *Decision with Multiple Objectives - Preferences and value Tradeoffs*. Wiley, New York, 1976.

[2] G. Choquet. *Theory of capacities*. Annales de l'Institut Fourier, 1953.

[3] P. Slovic, E. Peters M. Finucane, and D. G. MacGregor. *Heuristics and biases: the psychology on intuitive judgement*, chapter The affect heuristic, pages 397–420. Cambridge University Press, 2002.

[4] M. Grabisch and C. Labreuche. Bi-capacities–I: definition, Möbius transform and interaction. *Fuzzy Sets and Systems*, 151:211–236, 2005.

[5] S. Greco, R. Słowiński, and B. Matarazzo. Bipolar Sugeno and Choquet integrals. pages 191–196, Varenna, Italy, September 2002. EUROFUSE Workshop on Informations Systems.

[6] M. Grabisch and C. Labreuche. Bi-capacities–II: the Choquet integral. *Fuzzy Sets and Systems*, 151:237–259, 2005.

[7] S. Greco, S. Giove, and B. Matarazzo. The Choquet integral with respect to a level dependent capacity. *Fuzzy Sets and Systems*, submitted, 2008.

[8] S.Greco, R. Słowiński, J. Figueira, and V. Mousseau. *New Trends in Multiple Criteria Decision Analysis*, chapter Robust ordinal regression. Springer Science + Business Media, Inc, in press.

[9] J. Figueira and S. Greco. Dealing with interactivity beween bi-polar multiple criteria preferences in outranking methods. Manuscript, 2003.

[10] E. Jacquet-Lagrèze and Y. Siskos. Assessing a set of additive utility functions for multicriteria decision-making, the uta method. *European Journal of Operational Research*, 10:151–164, 1982.

[11] S. Angilella, F. Lamantia., S. Greco, and B. Matarazzo. Assessing non-additive utility for multicriteria decision aid. *European Journal of Operational Research*, 158:734–744, 2004.

[12] S. Greco, V. Mousseau, and R. Słowiński. Ordinal regression revisited: Multiple criteria ranking with a set of additive value functions. *European Journal of Operational Research*, 191:416–436, 2008.

[13] J. Figueira, S. Greco, and R. Słowiński. Building a set of additive value functions representing a reference preorder and intensities of preference: Grip method. *European Journal of Operational Research*, 195 (2):460–486, 2008.

[14] S. Greco, V. Mousseau, and R. Słowiński. Multiple criteria sorting with a set of additive value functions. *European Journal of Operational Research*, submitted.

[15] S. Angilella, S. Greco, and B. Matarazzo. Non-additive robust ordinal regression: a multiple criteria decision model based on the Choquet integral. *European Journal of Operational Research*, doi:10.1016/j.ejor.2009.02.023, 2009.

[16] M. Grabisch and C. Labreuche. Partially unipolar bi-capacities in MCDM. In: SCIS-ISIS, Yokohama, Japan, 21-24 September 2004.

[17] J.L. Marichal and M. Roubens. Determination of weights of interacting criteria from a reference set. *European Journal of Operational Research*, 124:641–650, 2000.

[18] J. Figueira, S. Greco, and R. Słowiński. Identifying the "most representative " value function among all compatible value functions in the grip method. 68$^{\text{th}}$ Meeting of the European Working Group on Multiple Criteria Decision Aiding, Chania, October 2-3, 2008.

[19] S. Angilella, S. Greco, and B. Matarazzo. Sorting decisions with interacting criteria. A.M.A.S.E.S. conference, Trento, Italy, September 1-4, 2008.

[20] S. Greco, M. Kadziński, and R. Słowiński. The most representative value functionin robust multiple criteria sorting. 69$^{\text{th}}$ Meeting of the European Working Group on Multiple Criteria Decision Aiding, Bruxelles, April 2-3, 2009 and submitted to Computers & Operations Research.

ISBN: 978-989-95079-6-8

Towards a unified logical framework of fuzzy implications to compare fuzzy sets

Hoel Le Capitaine Carl Frélicot

MIA Laboratory, University of La Rochelle
La Rochelle, France
Email: {hoel.le_capitaine,carl.frelicot}@univ-lr.fr

Abstract*— *In fuzzy set theory, comparison of fuzzy sets plays an important role. Among the several ways to compare fuzzy sets, we address the logical theoretic approach using fuzzy implications. We propose a general framework allowing to generate many measures of comparison: inclusion, similarity and distance, and study their properties. Since the literature on the use fuzzy implications for defining such measures is abundant, we also attempt to relate this work to existing research.

***Keywords*—** aggregation operators, distance, fuzzy implication, inclusion, similarity

1 Introduction

In fuzzy set theory, inclusion (or subsethood) and similarity measures between fuzzy sets are basic concepts. Among the different approaches to compare fuzzy sets, one finds three categories. The first one is based on measuring a distance between two real functions and do not refer to a specific interpretation, e.g. the Minkowski r-distances. The second category involves set-theoretic operations for fuzzy sets (fuzzy intersection, union, cardinality) [1, 2]. This paper addresses the third category which relies on logical theory, mainly approached by using fuzzy implications. Following the early paper by Bandler and Kohout [3] introducing the use of implications to measure inclusion of fuzzy sets, many authors have proposed new measures satisfying some specific axioms [4, 5, 6].

This paper will be organized as follows. We first recall some basic definitions and theorems on aggregation operators, triangular norms and fuzzy implications that will be used in the sequel in section 2. Next, in section 3 we propose a general framework of comparison measures of fuzzy sets (i.e. inclusion, similarity and distance) based on logical considerations, i.e. using fuzzy implications. A general description of the overall aggregation method is also provided, characterizing each measure belonging to this category. In section 4, we present the existing approaches that could be linked to this general formulation. We finally conclude and mention some application domains we have in mind in section 5.

2 Preliminaries and notations

Aggregating numbers plays an important role in decision-making systems. Values to be aggregated are generally defined on a finite real interval or on ordinal scales. In this paper, we assume with no loss of generality that they come from the unit interval. If not, a simple transformation can be found to make this assumption true. Given n numbers, an aggregation operator is a mapping $\mathcal{A} : [0,1]^n \to [0,1]$, satisfying boundary conditions

$$\mathcal{A}(0,\cdots,0) = 0 \quad \text{and} \quad \mathcal{A}(1,\cdots,1) = 1. \tag{1}$$

and monotonocity

$$\begin{aligned} \forall n \in \mathbb{N}, x_1 \leq y_1, \cdots, x_n \leq y_n \quad &\text{implies} \\ \mathcal{A}(x_1,\cdots,x_n) \leq \mathcal{A}(y_1,\cdots,y_n). \end{aligned} \tag{2}$$

Adding properties like idempotency, continuity, associativity lead to others definitions. In the literature, one finds many aggregation operators, e.g.: triangular norms (t-norms for short), OWA (*Ordered Weighted Averaging*) operators, γ-operators, or fuzzy integrals. They belong to several categories, depending on the way the values are aggregated: conjunctives, disjunctives, compensatory, and weighted operators. Briefly, an aggregation operator is said to be *conjunctive* if its output value is lower than the minimum of the input values, *compensatory* if the output value lies between the minimum and the maximum input value, and *disjunctive* if its output is greater than the maximum value of the inputs, refer to [7] for a large survey on aggregation operators.

Theorem 1. *If $\mathcal{A}$ is a strictly monotonic compensatory aggregation operator, then*

$$\mathcal{A}(x_1,\ldots,x_n) = 1 \Leftrightarrow x_1 = \cdots = x_n = 1 \tag{3}$$

and

$$\mathcal{A}(x_1,\ldots,x_n) = 0 \Leftrightarrow x_1 = \cdots = x_n = 0 \tag{4}$$

Proof.

(3) ($\Leftarrow$) obvious with Eq. (1).
($\Rightarrow$) by contraposition. We show that $\exists i$ such that $x_i < 1$ implies $\mathcal{A}(x_1,\cdots,x_n) < 1$. If $x_i < 1$ for some i, then, since $x_j \leq 1$ for all j, by strict monotonicity of $\mathcal{A}$, we get $\mathcal{A}(x_1,\cdots,x_n) < \mathcal{A}(1,\cdots,1) = 1$, which ends the contraposition.

(4) ($\Leftarrow$) obvious with Eq. (1).
($\Rightarrow$) by contraposition. We show that $\exists i$ such that $x_i > 0$ implies $\mathcal{A}(x_1,\cdots,x_n) > 0$. If $x_i > 0$ for some i, then, since $x_j \geq 0$ for all j, by strict monotonicity of $\mathcal{A}$, we get $\mathcal{A}(x_1,\cdots,x_n) > \mathcal{A}(0,\cdots,0) = 0$, which ends the contraposition.

□

Example 1. *The family of mean operators are strictly monotonic compensatory aggregation operators.*

ISBN: 978-989-95079-6-8

Remark 1. *From Theorem 1, we immediately see that a disjunctive aggregation operator cannot be strictly monotonic.* $\forall j\ x_j \leq 1$, *and* $\exists i,\ x_i < 1$, $\exists k,\ x_k = 1$. *If* $\mathcal{A}$ *is strictly monotonic,* $\mathcal{A}(x_1, \cdots, x_n) < \mathcal{A}(1, \cdots, 1) = 1$. *Since* $\mathcal{A}$ *is disjunctive, i.e.* $\mathcal{A}(x_1, \cdots, x_n) \geq \max(x_1, \cdots, x_n)$, *we have* $1 = \max(x_1, \cdots, x_n) \leq \mathcal{A}(x_1, \cdots, x_n) < 1$, *which is a contradiction.*

A t-norm is an increasing, associative and commutative mapping $\top : [0,1]^2 \to [0,1]$ satisfying the boundary condition $\top(x,1) = x$ for all $x \in [0,1]$. The most popular continuous t-norms are:

- Standard: $\top_S(x,y) = \min(x,y)$
- Algebraic (Product): $\top_A(x,y) = x\,y$
- Łukasiewicz: $\top_L(x,y) = \max(x+y-1, 0)$

Alternatively, the dual operators with respect to a strict negation are called triangular conorms (t-conorms for short). A t-conorm is an increasing, associative and commutative mapping $\perp : [0,1]^2 \to [0,1]$ satisfying the boundary condition $\perp(x,0) = x$ for all $x \in [0,1]$. The most popular continuous t-conorms are:

- Standard: $\perp_S(x,y) = \max(x,y)$
- Algebraic (Product): $\perp_A(x,y) = x + y - x\,y$
- Łukasiewicz: $\perp_L(x,y) = \min(x+y, 1)$

Various parametrical families have been introduced, e.g. the Hamacher family defined by: given $\gamma \in [0, +\infty[$,

- $x \top_H y = \frac{x\,y}{\gamma + (1-\gamma)(x+y-x\,y)}$
- $x \perp_H y = \frac{x+y-x\,y-(1-\gamma)\,x\,y}{1-(1-\gamma)\,x\,y}$.

Such families result in an infinite number of t-norm couples, including non parametrical ones e.g. the Algebraic couple for $\gamma = 1$.

A general problem in fuzzy logic is to handle conditional statements *if* x, *then* y where x and y are fuzzy predicates. A widely used method consists in managing them by mappings $I : [0,1] \times [0,1] \to [0,1]$ such that the truth value of I depends on the initial propositions x and y. We generally speak about an *implication function* if I is non-increasing in the first variable, non-decreasing in the second variable and $I(0,0) = I(1,1) = 1$, and $I(1,0) = 0$, see [8] for a survey on fuzzy implication functions and [9] for a large overview on the use of parametrical implications in fuzzy inference systems. The four most usual implications are:

1. *S-implications*, defined by:
$$I_\perp(x,y) = \perp(x^c, y) \tag{5}$$
where $(.)^c$ is the usual complementation $x^c = 1 - x$. This implication is an immediate generalization of the usual boolean implication $x \to y \equiv x^c \vee y$.

2. *R-(for Residual) implications*, defined by:
$$I_\top(x,y) = \sup_t \{t \in [0,1] | \top(x,t) \leq y\}. \tag{6}$$
Note that if $\top$ is a left-continuous t-norm, the *supremum* operation can be substituted by the *maximum* one.

3. *QL-(for Quantum mechanic Logic) implications*, defined by:
$$I_{QL}(x,y) = \perp(x^c, \top(x,y)). \tag{7}$$

4. *D-implications*, defined by:
$$I_D(x,y) = \perp(\top(x^c, y^c), y) \tag{8}$$
which are the contraposition of QL-implications with respect to the complementation.

In the sequel, we will denote by:

- $X = \{x_1, \cdots, x_n\}$ the (supposed finite) universe of discourse,
- $\mathcal{C}(X)$ and $\mathcal{F}(X)$ the sets of all crisps and fuzzy sets in X, respectively,
- $f_A(x)$, $\forall x \in X$, the membership function of a fuzzy set A over X,
- $[\frac{1}{2}]$ the constant fuzzy set defined by $[\frac{1}{2}](x) = \frac{1}{2}$ for any $x \in X$.

3 General framework

There are several ways to compare fuzzy values or fuzzy quantities. The first one is based on a broad class of measures of equality based on a distance measure which is specified for membership functions of fuzzy sets. This approach takes its roots from studies on how to measure the distance between two real functions and do not refer to any specific interpretation. The general form of a Minkowski r-metric is usually taken and leads to well known distance functions (Hamming, Euclidean, Chebyshev):

$$d_r(A,B) = \left(\sum_{x \in X} |f_A(x) - f_B(x)|^r \right)^{1/r} \tag{9}$$

A second way to compare fuzzy values comes from some basic set-theoretic considerations where union, intersection and complementation are defined for fuzzy sets. Cardinal and possibility based measures belong to this category. In this paper, we will focus our attention on the third way: the logical framework. We present the design a unified logical framework to compare fuzzy sets, which includes usual measures depending the aggregation operators we use. The use of parametrical t-norms and t-conorms to define the implications will enable to obtain parametrical measures of comparison, where the parameters can be set according to user needs.

3.1 Inclusion measures

An inclusion measure is a relation between two fuzzy sets which indicates to which extent one fuzzy set is contained in another one. Since its original definition by $A \subset B$ iff $f_A(x) \leq f_B(x)$, for all x in X, which was crisp assessment, Bandler and Kohout enlarged this point of view by giving a degree of subsetness [3], more in the spirit of the fuzzy theory. Inclusion measures are generally defined by a set of axioms [4] and by using fuzzy implication operators [3, 10].

Definition 1. *A mapping* $\mathcal{I} : \mathcal{F}(X) \times \mathcal{F}(X) \to [0,1]$ *is called an inclusion (or subsethood) measure if it satisfies*

ISBN: 978-989-95079-6-8

(P1) $\mathcal{I}(A,B) = 1$ *iff* $A \subseteq B$, $\quad \forall A, B \in \mathcal{F}(X)$.

(P2) *if* $[\frac{1}{2}] \subseteq A$, *then* $\mathcal{I}(A, A^c) = 0$ *iff* $A = X$.

(P3) $\forall A, B, C \in \mathcal{F}(X)$, *if* $A \subseteq B \subseteq C$, *then* $\mathcal{I}(C,A) \leq \mathcal{I}(B,A)$, *and if* $A \subseteq B$, $\mathcal{I}(C,A) \leq \mathcal{I}(C,B)$.

Note that if a measure $\mathcal{I}$ satisfies only the two last properties, it is called a weak inclusion measure.

Theorem 2. *Let $I_\top$ be a residual implication function. Given X and arbitrary $A, B \in \mathcal{F}(X)$, let*

$$\mathcal{I}(A,B) = \mathop{\mathcal{A}}_{i=1}^{n} I_\top \left(f_A(x_i), f_B(x_i)\right) \tag{10}$$

for all x_i in X, where $\mathcal{A}$ is a conjunctive or a strictly monotonic compensatory aggregation operator satisfying Eqs.(1) and (2). Then $\mathcal{I}$ is an inclusion measure.

Proof.

(P1) ($\Leftarrow$) if $A \subseteq B$, then $f_A(x_i) \leq f_B(x_i)$ for all $x_i \in X$. Since if $x \leq y$ then $I_\top(x,y) = 1$, we have $I_\top(f_A(x_i), f_B(x_i)) = 1$ for all $x_i \in X$, which gives $\mathcal{I}(A,B) = 1$ by boundary condition on $\mathcal{A}$.
($\Rightarrow$) two cases are considered
- if $\mathcal{A}$ is disjunctive, then we obviously have $f_A(x_i) \leq f_B(x_i)$ for all $x_i \in X$, since the minimum value of all implications on X is 1
- if $\mathcal{A}$ is a strictly monotonic compensatory aggregation operator, then by using Theorem 1, we have $I_\top\left(f_A(x_i), f_B(x_i)\right) = 1$ for all $x_i \in X$, giving $A \subseteq B$ by the confinement principle: $x \leq y$ *iff* $I(x,y) = 1$.

(P2) ($\Leftarrow$) if $A = X$, then $I_\top(f_X(x_i), \emptyset(x_i)) = 0$ for all $x_i \in X$. By boundary condition on $\mathcal{A}$, it follows that $\mathcal{I}(A, A^c) = 0$.
($\Rightarrow$) if $\mathcal{I}(A, A^c) = 0$, then $I_\top(f_A(x_i), 1 - f_A(x_i)) = 0$ for all $x_i \in X$ by Theorem 1. Assuming that $A \neq X$, then $\exists i$, x_i such that
$\frac{1}{2} \leq f_A(x_i) < 1$, i.e. $0 < 1 - f_A(x_i) \leq \frac{1}{2}$.
By non-increasingness in the first variable of $I_\top$, we have $I_\top(f_A(x_i), 1 - f_A(x_i)) \geq I_\top(1, 1 - f_A(x_i)) = 1 - f_A(x_i) \neq 0$, since $I_\top$ satisfies the border principle: $I(1,x) = x, \forall x \in [0,1]$. This is a contradiction with $I_\top(f_A(x_i), 1 - f_A(x_i)) = 0$, so that $A = X$.

(P3) if $A \subseteq B \subseteq C$, $f_A(x_i) \leq f_B(x_i) \leq f_C(x_i)$, for all $x_i \in X$. By non-increasingness in the first variable and non-decreasingness in the second variable, it follows that $I_\top(f_C(x_i), f_A(x_i) \leq I_\top(f_B(x_i), f_A(x_i))$ and $I_\top(f_C(x_i), f_A(x_i)) \leq I_\top(f_C(x_i), f_B(x_i))$ for all i. By monotonicity of $\mathcal{A}$, we have $\mathcal{I}(C,A) \leq \mathcal{I}(B,A)$ and $\mathcal{I}(C,A) \leq \mathcal{I}(C,B)$, which concludes the proof.

□

A necessary condition to define a strong inclusion measure $\mathcal{I}$ is that the implication I holds the confinement principle and the border principle. It is easy to show that the four usual implications $I_\top$, $I_\perp$, I_{QL} and I_D satisfy the latter, but only $I_\top$ satisfies the former. Therefore S, QL and $D-$implications define weak inclusion measures while $R-$implications define strong inclusion measures provided the operator $\mathcal{A}$ is not disjunctive, see Remark 1.

3.2 Similarity and distance measures

Definition 2. *A mapping $\mathcal{S} : \mathcal{F}(X) \times \mathcal{F}(X) \to [0,1]$ is called a similarity measure if it satisfies*

(P1) $\mathcal{S}(A,B) = \mathcal{S}(B,A), \quad \forall A, B \in \mathcal{F}(X)$.

(P2) $\mathcal{S}(A,A) = 1$, $\forall A \in \mathcal{F}(X)$.

(P3) $\mathcal{S}(D, D^c) = 0, \quad \forall D \in \mathcal{C}(X)$.

(P4) $\forall A, B, C \in \mathcal{F}(X)$, *if* $A \subseteq B \subseteq C$, *then* $\mathcal{S}(A,C) \leq \mathcal{S}(A,B) \wedge \mathcal{S}(B,C)$
or, equivalently
$\forall A, B, C, D \in \mathcal{F}(X)$, *if* $A \subseteq B \subseteq C \subseteq D$, *then* $\mathcal{S}(A,D) \leq \mathcal{S}(B,C)$

However, the symmetry property (P1) is still subject to experimental investigation: if $\mathcal{S}(x,y)$ is the answer to the question *how is x similar to y?*, then, when making comparisons, subjects focus more on the feature x than on y. This corresponds to the notion of saliency [1] of x and y: if y is more salient than x, then x is more similar to y than vice versa, which is experimentally confirmed. Accordingly to [11], for a certain element of the universe of discourse X, a degree of equality of two fuzzy elements x and y can be defined by using implications as follows:

$$(x \equiv y) = \frac{1}{2}\big((x \to y) \wedge (y \to x) + (x^c \to y^c) \wedge (y^c \to x^c)\big) \tag{11}$$

where $\wedge$ stands for minimum, $\to$ is an R-implication.
Then applying Eq. (6) with $x \geq y$, gives

$$(x \equiv y) = \frac{1}{2}\big((x \to y) + (y^c \to x^c)\big) \tag{12}$$

since 1 is the neutral element of t-norms.
Furthermore, due to the opposition law: $I(x,y) = I(y^c, x^c)$, we obtain

$$(x \equiv y) = (x \to y) \tag{13}$$

A convenient way to define a similarity measure is to quantify to which extent two fuzzy membership degrees are similar, so that it is closely related to the problem of matching fuzzy quantities, or fuzzy sets similarity. So we propose to use fuzzy implication functions as similarity measures by the following theorem:

Theorem 3. *Let $I_\top$ be a residual implication function. For arbitrary $A, B \in \mathcal{F}(X)$, let*

$$\mathcal{S}(A,B) = \mathop{\mathcal{A}}_{i=1}^{n} I_\top \left(f_{(1)}(x_i), f_{(2)}(x_i)\right) \tag{14}$$

for all x_i in X, where $f_{(.)}$ is a permutation of f_A and f_B such that $f_{(1)}(x_i) = (f_A \cup f_B)(x_i)$, $f_{(2)}(x_i) = (f_A \cap f_B)(x_i)$, and $\mathcal{A}$ an aggregation operator satisfying Eqs. (1) and (2). Then $\mathcal{S}$ is a similarity measure.

Proof.

(P1) we have $I_\top(x,x) = 1$, for any $x \in [0,1]$. By boundary conditions on $\mathcal{A}$, see Eq. (1), we obtain $\mathcal{S}(A,A) = 1$.

ISBN: 978-989-95079-6-8

(P2) by commutativity of union and intersection of fuzzy sets, we have

$$\begin{aligned}\mathcal{S}(A,B) &= \mathop{\mathcal{A}}_{i=1}^{n} I_\top\left((f_A \cup f_B)(x_i), (f_A \cap f_B)(x_i)\right)\\ &= \mathop{\mathcal{A}}_{i=1}^{n} I_\top\left((f_B \cup f_A)(x_i), (f_B \cap f_A)(x_i)\right)\\ &= \mathcal{S}(B,A)\end{aligned}$$

(P3) by definition, $I(1,0)=0$. By boundary conditions on $\mathcal{A}$, see Eq. (1), we obtain $\mathcal{S}(D,D^c)=0$.

(P4) since $A \subseteq B \subseteq C \subseteq D$, we have for all $x_i \in X$

$$f_D(x_i) \geq f_C(x_i) \tag{15}$$
$$f_B(x_i) \geq f_A(x_i) \tag{16}$$

By non-increasingness in the first variable and non-decreasingness in the second variable of $I_\top$, we obtain for all $x_i \in X$

$I_\top(f_D(x_i), f_A(x_i)) \leq I_\top(f_C(x_i), f_A(x_i))$ by Eq. (15)

$I_\top(f_C(x_i), f_A(x_i)) \leq I_\top(f_C(x_i), f_B(x_i))$ by Eq. (16)

Last, monotonicity of $\mathcal{A}$, see Eq. (2), gives $\mathcal{S}(A,D) \leq \mathcal{S}(B,C)$ which concludes the proof.

□

In contrast to inclusion measures, no restriction is imposed to the $\mathcal{A}$ operator, it can be freely chosen provided Eqs.(1) and (2) are satisfied. On another hand, here again, the confinement principle is necessary to obtain property (P1) of similarity measures. Since $\mathcal{S}(A,B)$ is reflexive and symmetrical, i.e. $\mathcal{S}(A,A)=1$ and $\mathcal{S}(A,B)=\mathcal{S}(B,A)$ hold for any $A,B \in \mathcal{F}(X)$, $\mathcal{S}$ is a proximity relation on $\mathcal{F}(X)$.

Definition 3. *A mapping $\mathcal{D}: \mathcal{F}(X)\times\mathcal{F}(X) \to [0,1]$ is called a distance measure if $\mathcal{D}$ has the following properties*

(P1) $\mathcal{D}(A,B) = \mathcal{D}(B,A), \quad \forall A,B \in \mathcal{F}(X)$.

(P2) $\mathcal{D}(A,A) = 0, \forall A \in \mathcal{F}(X)$.

(P3) $\mathcal{D}(D,D^c) = 1, \quad \forall D \in \mathcal{C}(X)$.

(P4) $\forall A,B,C \in \mathcal{F}(X)$, *if* $A \subseteq B \subseteq C$, *then* $\mathcal{D}(A,B) \leq \mathcal{D}(A,C)$ *and* $\mathcal{D}(B,C) \leq \mathcal{D}(A,C)$

Proposition 1. *For arbitrary $A,B \in \mathcal{F}(X)$, and $\mathcal{S}(A,B)$ a similarity measure defined by Eq. (14), then the mapping $\mathcal{D}: \mathcal{F}(X)\times\mathcal{F}(X) \to [0,1]$ defined by $\mathcal{D}(A,B) = \mathcal{S}^c(A,B)$ is a distance measure between A and B.*

3.3 Examples

Some examples of inclusion and similarity measures of the literature, and the new Hamacher inclusion and similarity measures, all obtained through the proposed general logical framework are given in Table 1. Moreover, this table shows how it is easy to check wether an inclusion measure is strong (non disjunctive $\mathcal{A}$ and $I_\top$) or weak (non disjunctive $\mathcal{A}$), see Definition 1, as well as the proper definition of a similarity measure (any $\mathcal{A}$ and $I_\top$), see Theorem 3.

For illustration purpose, Fig. 1 shows some examples of fuzzy similarity measures as well as the influence of the parameter γ for Hamacher residual implications, which are defined by:

$$I_{\top_{H_\gamma}}(x,y) = \begin{cases} 1 & \text{if } y \geq x \\ \frac{y(\gamma+x-\gamma x)}{y(\gamma+x-\gamma x)+x-y} & \text{if } y \leq x \end{cases} \tag{17}$$

where $\gamma \in [0,+\infty[$. The different plots show the similarity value of a given fuzzy set $A = \{0.4/x_1, 0.3/x_2\}$ to all the possible two-dimensional fuzzy sets B for various $I_\top$. As one could expect, the closer to x_1 or x_2 the higher the similarity. One can also note that different $I_\top$ lead to different shapes for isosurfaces of the similarity.

4 Related works

Let us relate, as far as possible, the proposed framework to some works issued from the abundant literature on the use of fuzzy implications for defining inclusion measures. In their seminal paper [3], Bandler and Kohout propose to use implication functions in order to quantify the inclusion of each element in another, and then aggregate these individual measures by a conjunctive operator: the minimum. Some years later, Hirota and Pedrycz propose to use implications for matching fuzzy quantities [11]. They aggregate the different implication truth values over X by a Choquet integral, computing the fuzzy measure with the help of a family of fuzzy sets taken as fuzzy prototypes. They also propose an entropy measure based on this matching measure, which gives an *impression concerning the uncertainty of matching*. In [14], Kosko criticizes the original definition of fuzzy set containment: B contains A if and only if $f_A(x) \leq f_B(x)$ for all x in X by Zadeh, pointing out that *if this inequality holds for all but just a few x, we can still consider A to be a subset of B to some degree*. So he proposes a second definition based on fuzzy intersection of two fuzzy sets which can be seen as the conditional probability $P(B|A)$ under certain circumstances. Furthermore, he defines the fuzziness of a fuzzy set A as the inclusion measure of $A \cup A^c$ in $A \cap A^c$, which satisfies the axioms of fuzzy entropy of De Luca and Termini [15]. Inclusion and similarity measures from a general set-theoretic point of view, coming from the proposition of Tversky [1], are described by Bouchon-Meunier *et al.* in [2]. The authors introduce a general framework for similitude, satisfiability and inclusion measures, also study the aggregation of measures of comparison, but take only two examples for the aggregation operator $\mathcal{A}$: a t-norm and the OWA operator. In this paper, we propose a study of the properties of $\mathcal{A}$ in its more general meaning. The same remark applies to the work by Young [4] proposing an axiomatization of inclusion measures and their connection to implication operators since she restricts herself to both the minimum and the arithmetical mean for $\mathcal{A}$. Furthermore, this work do not give necessary conditions on implications for the definition of strong or weak inclusion measure whereas our's does (Definition 1). Wang [13] presents two similarity measures which can be viewed as particular cases of the framework we propose. He enlarges his definition to the similarity of fuzzy elements belonging to various fuzzy sets. Again, the framework we propose allows to obtain a similarity between fuzzy elements since we compute implications for each element. Starting from Kosko and Young observations, Botana [16] presents a set of new measures derived

ISBN: 978-989-95079-6-8

Table 1: Inclusion and similarity measures of the literature, and the new Hamacher inclusion and similarity measures, all obtained through the proposed general logical framework.

Inclusion Measure $\mathcal{I}$	**Aggregation Operator** $\mathcal{A}$	**Implication** I
$\mathcal{I}(A,B) = \min_{x\in X}(\min(1, 1 - f_A(x) + f_B(x)))$ as defined in [12]	min	$I_{\top_L}$
$\mathcal{I}(A,B) = \frac{1}{n}\sum_{x\in X}\min(1, 1 - f_A(x) + f_B(x))$	arithmetical mean	$I_{\top_L}$
$\mathcal{I}(A,B) = \frac{1}{n}\sum_{x\in X}\max(1 - f_A(x), f_B(x))$	arithmetical mean	$I_{\perp_S}$
$\mathcal{I}(A,B) = \frac{1}{n}\sum_{x\in X} 1 - f_A(x) + f_A(x)\, f_B(x)$ as defined in [4]	arithmetical mean	$I_{\perp_A}$
$\mathcal{I}(A,B) = \frac{1}{n}\sum_{x\in X}\frac{f_B(x)(\gamma + f_A(x) - \gamma f_A(x))}{f_B(x)(\gamma + f_A(x) - \gamma f_A(x)) + f_A(x) - f_B(x)}$	arithmetical mean	$I_{\top_{H_\gamma}}$
Similarity Measure $\mathcal{S}$	**Aggregation Operator** $\mathcal{A}$	**Implication** I
$\mathcal{S}(A,B) = \frac{1}{n}\sum_{x\in X}\frac{\min(f_A(x), f_B(x))}{\max(f_A(x), f_B(x))}$ as defined in [13]	arithmetical mean	$I_{\top_A}$
$\mathcal{S}(A,B) = \frac{1}{n}\sum_{x\in X} 1 - \lvert f_A(x) - f_B(x)\rvert$ as defined in [13]	arithmetical mean	$I_{\top_L}$
$\mathcal{S}(A,B) = \max_{x\in X}\min(f_A(x), f_B(x))$	max	$I_{\top_S}$
$\mathcal{S}(A,B) = 1 - \max_{x\in X}\lvert f_A(x) - f_B(x)\rvert$	min	$I_{\top_L}$
$\mathcal{S}(A,B) = \frac{1}{n}\sum_{x\in X}\frac{f_{(2)}(x)(\gamma + f_{(1)}(x) - \gamma f_{(1)}(x))}{f_{(2)}(x)(\gamma + f_{(1)}(x) - \gamma f_{(1)}(x)) + f_{(1)}(x) - f_{(2)}(x)}$	arithmetical mean	$I_{\top_{H_\gamma}}$

from fuzzy implications. He studies whether the introduced inclusion measures satisfy Young's axioms [4] and/or those in [12] when using Wu, Goguen, modified Goguen, Gödel and Standard strict implications. He also gives the formulation of the corresponding entropy in the sense of [15, 14]. Another approach to aggregation operators is proposed in [17] consisting in combining implication truth values through respectively disjunctive and conjunctive functions for an optimistic and pessimistic aggregation. Fan et al. discuss the links between inclusion, entropy and fuzzy implications, and propose some new axioms for these measures [5]. Burillo et al. present a family of implication operators derived from the Łukasiewicz implication in order to define a family of inclusion grade operators [6] using the minimum operator for $\mathcal{A}$. In [18], Kehagias and Konstantinidou introduce L-fuzzy valued inclusion, similarity and distance measures, i.e. mappings $\mathcal{I}, \mathcal{S}, \mathcal{D} : \mathcal{F}(X) \times \mathcal{F}(X) \to [0,1]^n$, but restrict the output to crisp values. As pointed out by the authors, the vector output can lead to a difficult interpretation of the result. Furthermore, as vectors are partially ordered they are even harder to compare, they require a new measure to compare outputs. The framework we propose can provide a vector output since an implication on each element of X is computed. By contrast to [18], it would be a *fuzzy L-fuzzy valued* measure. More recently [19, 20], a distance between fuzzy operators, fuzzy implication functions in particular, is proposed. It leads to normalized tensor-norms which allow to define a similarity of fuzzy sets, and present an heuristic to choose the most suitable implication function to a fuzzy inference system. Zhang and Zhang propose an hybrid inclusion measure in [10] and use it to define similarity and distance measures of fuzzy sets. They restrict themselves to the weighted mean for $\mathcal{A}$, so it can be viewed as a special case of the work we propose, but contrarily to Young the involved implications satisfy the confinement and the border principle. Let us finally mention the work by Fono et al. [21] where as many difference operations as many implications are used to define measures of comparison of fuzzy sets in the set-theoretic framework, but we remind that such measures are out of the scope of this paper.

5 Conclusion and perspectives

In this paper, we propose a unified logical framework to compare fuzzy sets as well as fuzzy elements. Within this framework, new measures of inclusion, similarity and distance can be easily derived. These measures depend on any fuzzy implication I, provided it satisfies the necessary conditions we give (the border and the confinement principles) and any aggregation operator $\mathcal{A}$ (for similarity and distance) or any conjunctive and strictly monotonic compensatory aggregation operator $\mathcal{A}$ (for inclusion). Choosing specific $(\mathcal{A}, I)$ enables to retrieve most of the measures of the literature.
We hope that results of this work would be of great help to set comparison functions in many fields: fuzzy mathematical morphology [6], cluster validity [5], as well as other domains e.g. image retrieval or feature selection.

References

[1] A. Tversky. Features of similarity. *Psychological review*, 84(4):327–352, 1977.

[2] B. Bouchon-Meunier, M. Rifqi, and S. Bothorel. Towards general measures of comparison of objects. *Fuzzy Sets and Systems*, 84(2):143–153, 1996.

[3] W. Bandler and L. Kohout. Fuzzy power sets and fuzzy implication operators. *Fuzzy Sets and Systems*, 4:13–30, 1980.

[4] V. R. Young. Fuzzy subsethood. *Fuzzy Sets and Systems*, 77(3):371–384, 1996.

ISBN: 978-989-95079-6-8

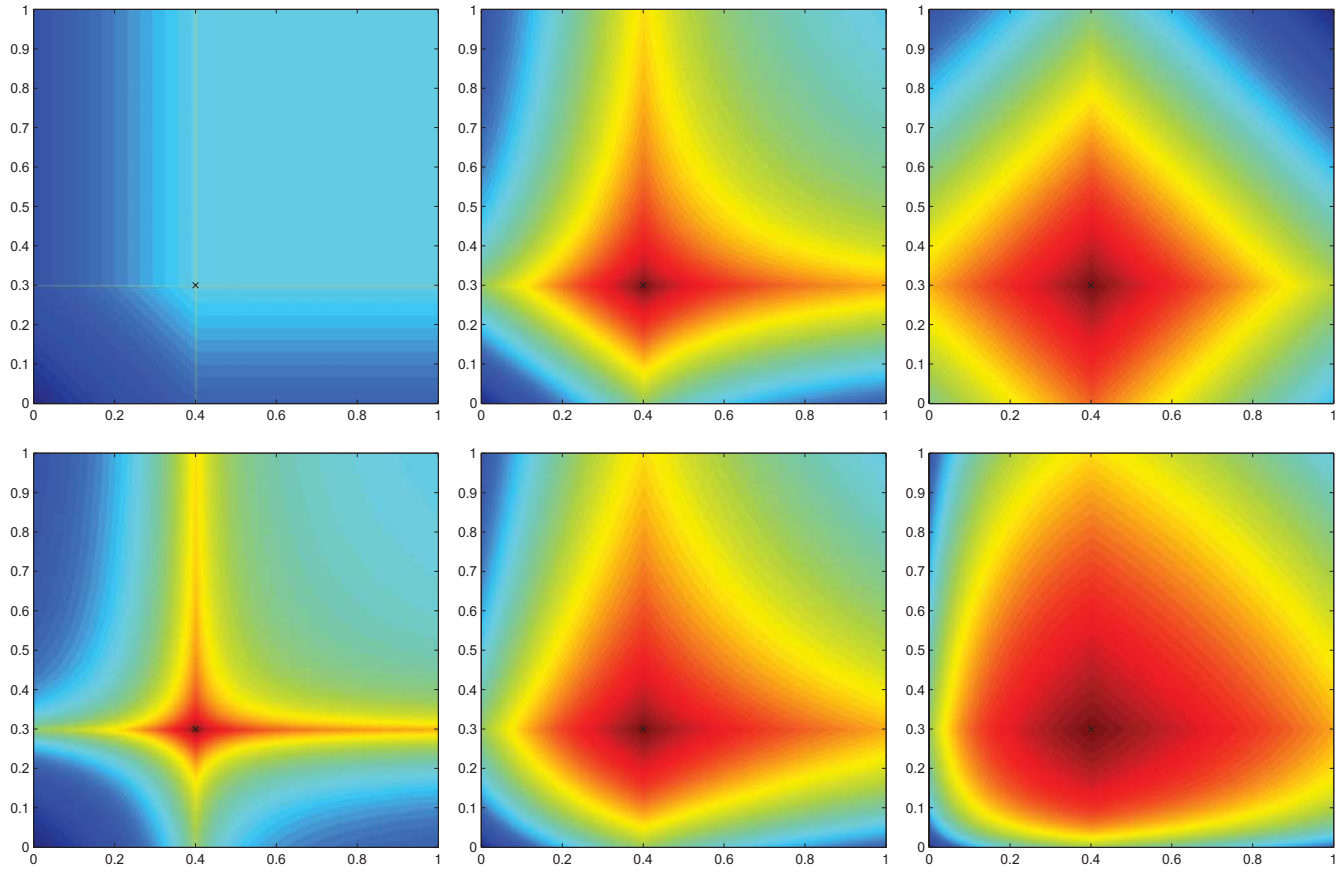


Figure 1: Examples of similarity measures of $A = \{0.4/x_1, 0.3/x_2\}$ and all fuzzy sets $\mathcal{F}(X)$, $n = 2$ where high and low values correspond to red and blue colors respectively. First row: $\mathcal{A}$ is the arithmetical mean, and we use $I_{\top_S}$, $I_{\top_A}$ and $I_{\top_L}$, respectively from left to right. Second row: $\mathcal{A}$ is the arithmetical mean and we used the Hamacher implication $I_{\top_{H_\gamma}}$, where $\gamma = 0, 2, 5$ from left to right.

[5] J. Fan, W. Xie, and J. Pei. Subsethood measure: new definitions. *Fuzzy Sets and Systems*, 106(2):201–209, 1999.

[6] P. Burillo, N. Frago, and R. Fuentes. Inclusion grade and fuzzy implication operators. *Fuzzy Sets and Systems*, 114(3):143–153, 2000.

[7] T. Calvo, A. Kolesarova, M. Komornikova, and R. Mesiar. *Aggregation Operators: New Trends and Applications*, volume 97 of *Studies in Fuzziness and Soft Computing*, chapter Aggregation Operators: Properties, Classes and Construction Methods, pages 1–104. Physica Verlag, 2002.

[8] M. Mas, M. Monserrat, J. Torrens, and E. Trillas. A survey on fuzzy implication functions. *IEEE Transactions on Fuzzy Systems*, 15(6):1107–1121, 2007.

[9] T. Whalen. Parameterized r-implications. *Fuzzy Sets and Systems*, 134(2):231–281, 2003.

[10] H-Y. Zhang and W-X. Zhang. Hybrid monotonic inclusion measure and its use in measuring similarity and distance between fuzzy sets. *Fuzzy Sets and Systems*, 160(1):107–118, 2009.

[11] K. Hirota and W. Pedrycz. Matching fuzzy quantities. *IEEE Transactions on Systems, Man, and Cybernetics*, 21(6):1580–1586, 1991.

[12] D. Sinha and E. R. Dougherty. Fuzzification of set inclusion : theory and applications. *Fuzzy Sets and Systems*, 55(1):15–42, 1993.

[13] W-J. Wang. New similarity measures on fuzzy sets and on elements. *Fuzzy Sets and Systems*, 85:305–309, 1997.

[14] B. Kosko. *Neural networks and fuzzy systems*. Prentice Hall, Englewood Cliffs, NJ, 1992.

[15] A De Luca and S Termini. A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory. *Information Control*, 20:301–312, 1972.

[16] F. Botana. Deriving fuzzy subsethood measures from violations of the implication between elements. In *11th Int. Conf. on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*, volume LNCS 1415, pages 234–243. Springer, 1998.

[17] P. Fonck, J. Fodor, and M. Roubens. An application of aggregation procedures to the definition of measures of similarity between fuzzy sets. *Fuzzy Sets and Systems*, 97(1):67–74, 1998.

[18] A. Kehagias and M. Konstantinidou. L-fuzzy valued inclusion measure, l-fuzzy similarity and l-fuzzy distance. *Fuzzy Sets and Systems*, 136(3):313–332, 2003.

[19] B.K. Papadopoulos, G. Trasanides, and A.G. Hatzimichailidis. Optimization method for the selection of the appropriate fuzzy implication. *Journal of Optimization Theory and Applications*, 134(1):135–141, 2007.

[20] V. Balopoulos, A. Hatzimichailidis, and B. Papadopoulos. Distance and similarity measures for fuzzy operators. *Information Sciences*, 177:2336–2348, 2007.

[21] L. Fono, H. Gwet, and B. Bouchon-Meunier. Fuzzy implication operators for difference operations for fuzzy sets and cardinality-based measures of comparison. *European Journal of Operational Research*, 183(1):314–326, 2007.

ISBN: 978-989-95079-6-8

Preservation of properties of interval-valued fuzzy relations

B. Pękala
Institute of Mathematics, University of Rzeszów
Rejtana 16a, 35-310 Rzeszów, Poland
Email: bpekala@univ.rzeszow.pl

Abstract— *The goal of this paper is to consider properties of the composition of interval-valued fuzzy relations which were introduced by L.A. Zadeh in 1975. Fuzzy set theory turned out to be a useful tool to describe situations in which the data are imprecise or vague. Interval-valued fuzzy set theory is a generalization of fuzzy set theory which was introduced also by Zadeh in 1965. This paper generalizes some properties of interval matrices considered by Pękala (2007) on these of interval-valued fuzzy relations.*

Keywords— Fuzzy relations, interval-valued fuzzy relations, properties of interval-valued fuzzy relations

1 Introduction

The idea of a fuzzy relation was defined in [28]. An extension of fuzzy set theory is interval-valued fuzzy set theory. Any interval-valued fuzzy set is defined by an interval-valued membership function: a mapping from the given universe to the set of all closed subintervals of [0,1] (it means that information is incomplete). In this paper we study properties of the composition of interval-valued fuzzy relations. Consideration of diverse properties of the composition is interesting not only from a theoretical point of view but also for the applications, since the composition of interval-valued fuzzy relations has proved to be useful in several fields, see for example, [20] (performance evaluation), [27] (genetic algorithm), [19] (approximate reasoning) or in other (see [1, 16]). In Section 2, we recall elementary properties of the composition of interval-valued fuzzy relations. Next, we consider preservation of properties of interval-valued fuzzy relations by the composition and lattice operations.

We give the following definition. Let X, Y, Z be crisp finite non-empty sets.

Definition 1 (cf. [29, 26]). Let $Int([0,1])$ be the set of all closed subintervals of $[0,1]$. An interval-valued fuzzy relation R in a universe X, Y is a mapping $R : X \times Y \to Int([0,1])$ such that $R(x,y) = [\underline{R}(x,y), \overline{R}(x,y)] \in Int([0,1])$, for all pairs $(x,y) \in (X \times Y)$. The class of interval-valued fuzzy relations in a universe $X \times Y$ will be denoted by $IVFR(X \times Y)$ or $IVFR(X)$ for $X = Y$.

Interval-valued fuzzy relations reflect the idea that membership grades are often not precise and the intervals represent such uncertainty.

The boundary elements in $IVFR(X \times Y)$ are $\mathbf{1} = [1,1]$ and $\mathbf{0} = [0,0]$. The relation R^{-1} between Y and X is defined by $R^{-1}(y,x) = R(x,y)$ for all $(x,y) \in (X \times Y)$ which we will call the inverse relation of R.

Let us look at immediate properties for interval-valued fuzzy relations: Let $P, R \in IVFR(X \times Y)$.

Then for every $(x,y) \in (X \times Y)$ we can define

$$P(x,y) \leq R(x,y) \Leftrightarrow \underline{P}(x,y) \leq \underline{R}(x,y), \overline{P}(x,y) \leq \overline{R}(x,y),$$

$$(P \vee R)(x,y) = [\max(\underline{P}(x,y), \underline{R}(x,y)), \max(\overline{P}(x,y), \overline{R}(x,y))],$$

$$(P \wedge R)(x,y) = [\min(\underline{P}(x,y), \underline{R}(x,y)), \min(\overline{P}(x,y), \overline{R}(x,y))],$$

where operations $\vee$ and $\wedge$ are the supremum and the infimum in $IVFR(X \times Y)$, respectively. For arbitrary set $T \neq \emptyset$ similarly we use

$$(\bigvee_{t \in T} R_t)(x,y) = [\bigvee_{t \in T} \underline{R_t}(x,y), \bigvee_{t \in T} \overline{R_t}(x,y)],$$

$$(\bigwedge_{t \in T} R_t)(x,y) = [\bigwedge_{t \in T} \underline{R_t}(x,y), \bigwedge_{t \in T} \overline{R_t}(x,y)].$$

We know that $([0,1], \max, \min)$ is a complete, distributive lattice, and therefore $(IVFR(X \times Y), \vee, \wedge)$ is also a complete, distributive lattice. So it is a particular case of lattices considered by Goguen in [18].

Interval-valued fuzzy relations (sets) are equivalent to some other extensions of fuzzy relations (sets) (see [12]). Among others, interval-valued fuzzy relations are isomorphic to Atanassov's fuzzy relations. This fact was noticed by several authors [5, 11, 12]. An Atanassov's fuzzy relation is a pair of fuzzy relations, namely a membership and a nonmembership functions, which represent positive and negative aspects of the given information. This objects introduced by Atanassov and originally called intuitionistic fuzzy relations were recently suggested to be called Atanassov's intuitionistic fuzzy relations or just bipolar fuzzy relations [15]. An Atanassov's fuzzy set theory is also widely applied in solving real-life problems. An example of such application is the optimization in Atanassov's intuitionistic fuzzy environment (an extension of fuzzy optimization and an application of bipolar fuzzy sets) where by applying this concept it is possible to reformulate the optimization problem by using degrees of rejection of constraints and values of the objective which are non-admissible. This concept allows one to define a degree of rejection which cannot be simply a complement of the degree of acceptance [2]. The idea of a positive and negative information was confirmed by psychological investigations [9]. Moreover, multiattribute decision making using Atanassov's intuitionistic fuzzy sets is possible (see [22, 23]).

Definition 2 (cf. [3]). Let $X \neq \emptyset$, $\mathfrak{R},\ \mathfrak{R}^d : X \times Y \to [0,1]$ be fuzzy relations fulfilling the condition

$$\mathfrak{R}(x,y) + \mathfrak{R}^d(x,y) \leqslant 1, \quad (x,y) \in (X \times Y).$$

A pair $\rho = (\mathfrak{R}, \mathfrak{R}^d)$ is called an Atanassov's intuitionistic fuzzy relation. The family of all Atanassov's intuitionistic fuzzy relations described in a given sets X, Y is denoted by $AIFR(X \times Y)$.

ISBN: 978-989-95079-6-8

Basic operations on Atanassov's intuitionistic fuzzy relations $\rho = (\mathfrak{R}, \mathfrak{R}^d)$, $\sigma = (\mathfrak{S}, \mathfrak{S}^d)$ are defined in the following way

$$\rho \cup \sigma = (\max(\mathfrak{R}, \mathfrak{S}), \min(\mathfrak{R}^d, \mathfrak{S}^d)),$$

$$\rho \cap \sigma = (\min(\mathfrak{R}, \mathfrak{S}), \max(\mathfrak{R}^d, \mathfrak{S}^d)),$$

$$\rho \leqslant \sigma \ \Leftrightarrow \ (\mathfrak{R} \leq \mathfrak{S}, \ \mathfrak{S}^d \leq \mathfrak{R}^d).$$

The isomorphism which proves the equivalence between Atanassov's intuitionistic fuzzy relations and interval-valued fuzzy relations is the following

Theorem 1 (cf. [11]). *The mapping $\psi : IVFR(X \times Y) \to AIFR(X \times Y)$, such that $R \to \rho$ is an isomorphism between the lattices $(IVFR(X \times Y), \vee, \wedge)$ and $(AIFR(X \times Y), \cup, \cap)$, where $R(x,y) = [\underline{R}(x,y), \overline{R}(x,y)]$, $R \in IVFR(X \times Y)$ and $\rho(x,y) = (\underline{R}(x,y), 1 - \overline{R}(x,y))$, $(x,y) \in (X \times Y)$.*

2 Composition of interval-valued fuzzy relations

Now, we consider the composition of interval-valued fuzzy relations.

Definition 3 (cf. [8]). Let $* : [0,1]^2 \to [0,1]$,
$P \in IVFR(X \times Y)$, $R \in IVFR(Y \times Z)$.
By the $\sup -*$ composition of relations P and R we call the relation $P \circ R \in IVFR(X \times Z)$,

$$(P \circ R)(x,z) = [(\underline{P} \circ \underline{R})(x,z), (\overline{P} \circ \overline{R})(x,z)], \quad (1)$$

where

$$(\underline{P} \circ \underline{R})(x,z) = \bigvee_{y \in Y} (\underline{P}(x,y) * \underline{R}(y,z)),$$

$$(\overline{P} \circ \overline{R})(x,z) = \bigvee_{y \in Y} (\overline{P}(x,y) * \overline{R}(y,z))$$

and $(\underline{P} \circ \underline{R})(x,z) \leq (\overline{P} \circ \overline{R})(x,z)$.

Other types of compositions may be also considered. For example, the composition of interval-valued fuzzy relations with supremum and $*$ being a triangular norm or a triangular conorm is examined in [8].

Definition 4 (cf. [24]). A triangular norm T (conorm S) is an increasing, commutative, associative operation
$T(S) : [0,1]^2 \to [0,1]$ with a neutral element 1 (0).

In [8] it was proved, for finite sets X, Y, Z, that compositions mentioned there are associative if and only if the first operation in composition is $\vee$ or $\wedge$ and the second is an arbitrary triangular norm or conorm. This is why the choice of the first operation in the composition (1) is reasonable.

For our further considerations we need the following properties

Definition 5 (cf. [18]). Let $* : [0,1]^2 \to [0,1]$. Operation $*$ is infinitely sup-distributive, if

$$\bigvee_{t \in T} (x_t * y) = (\bigvee_{t \in T} x_t) * y, \quad \bigvee_{t \in T} (y * x_t) = y * (\bigvee_{t \in T} x_t).$$

Definition 6 (cf. [6]). Operation $* : [0,1]^2 \to [0,1]$ is isotone if it fulfils the condition:

$$\underset{x,y,z \in [0,1]}{\forall} x \leq y \Rightarrow x * z \leq y * z, z * x \leq z * y.$$

By generalization of the results of the papers [14, 25] and using [13] we obtain

Lemma 1. *If $*$ is isotonic, then $\sup -*$ composition is also isotonic.*

Proof. Let $*$ be left side isotonic.
For $P, R \in IVFR(X \times Y), Q \in IVFR(Y \times Z)$

$$P \leq R \Leftrightarrow [\underline{P}, \overline{P}] \leq [\underline{R}, \overline{R}] \Leftrightarrow$$

$$\underset{x,y \in X \times Y}{\forall} \underline{P}(x,y) \leq \underline{R}(x,y), \overline{P}(x,y) \leq \overline{R}(x,y)$$

by isotonicity of the $*$ and supremum we have for
$x \in X, y \in Y, z \in Z$

$$\bigvee_{y \in X} (\underline{P}(x,y) * \underline{Q}(y,z)) \leq \bigvee_{y \in X} (\underline{R}(x,y) * \underline{Q}(y,z)),$$

$$\bigvee_{y \in X} (\overline{P}(x,y) * \overline{Q}(y,z) \leq \bigvee_{y \in X} (\overline{R}(x,y) * \overline{Q}(y,z)) \Leftrightarrow$$

$$\underline{P} \circ \underline{Q} \leq \underline{R} \circ \underline{Q}, \overline{P} \circ \overline{Q} \leq \overline{R} \circ \overline{Q} \Leftrightarrow P \circ Q \leq R \circ Q.$$

The right side isotonicity of $\circ$ may be proved similarly. □

Lemma 2. *Let the operation $*$ have a zero element z. If $*$ has a neutral element e, then $\sup -*$ composition has a neutral element $\mathbf{S}_e = [S_e, S_e]$,*

$$S_e(x,y) = \begin{cases} e & if \ x = y \\ z & if \ x \neq y \end{cases} \quad for \ x \in X, y \in Y.$$

Proof. Let $*$ operation have a zero element z and a neutral element e.

$$(\mathbf{S}_e \circ R)(x,y) =$$

$$[\bigvee_{z \in X} (S_e(x,z) * \underline{R}(z,y)), \bigvee_{z \in X} (S_e(x,z) * \overline{R}(z,y))] =$$

$$[(e * \underline{R}(x,y)) \vee \bigvee_{x \neq z} (z * \underline{R}(z,y)), (e * \overline{R}(x,y)) \vee \bigvee_{x \neq z} (z * \overline{R}(z,y))] =$$

$$[e * \underline{R}(x,y), e * \overline{R}(x,y)] = [\underline{R}(x,y), \overline{R}(x,y)] = R(x,y).$$

The proof for $R \circ \mathbf{S}_e = R$ is similar. □

If $*$ is an isotonic operation, then we may prove

$$\bigvee_{t \in T} (P_t \circ R) \leq (\bigvee_{t \in T} P_t) \circ R, \ \bigwedge_{t \in T} (P_t \circ R) \geq (\bigwedge_{t \in T} P_t) \circ R. \quad (2)$$

Now we examine the problem of sup-distributivity. Authors [8] in Theorem 6 present condition for sup-distributivity by $*$ equal to any t-norm or t-conorm and by finite non-empty X, Y. By generalization we obtain

Lemma 3. *If an operation $*$ is infinitely sup-distributive, then $\sup -*$ composition is also infinitely sup-distributive i.e., for $P_t \in IVFR(X \times Y), R \in IVFR(Y \times Z)$*

$$\bigvee_{t \in T} (P_t \circ R) = (\bigvee_{t \in T} P_t) \circ R. \quad (3)$$

ISBN: 978-989-95079-6-8

Proof. Let $*$ be sup-distributive, $x \in X, y \in Y, z \in Z$. Then

$$((\bigvee_{t\in T} P_t) \circ R)(x,z) =$$

$$[\bigvee_{y\in Y}((\bigvee_{t\in T}\underline{P_t})(x,y)*\underline{R}(y,z)), \bigvee_{y\in Y}((\bigvee_{t\in T}\overline{P_t})(x,y)*\overline{R}(y,z))] =$$

$$[\bigvee_{y\in Y}(\bigvee_{t\in T}(\underline{P_t}(x,y)*\underline{R}(y,z))), \bigvee_{y\in Y}(\bigvee_{t\in T}(\overline{P_t}(x,y)*\overline{R}(y,z)))] =$$

$$[\bigvee_{t\in T}(\bigvee_{y\in Y}(\underline{P_t}(x,y)*\underline{R}(y,z))), \bigvee_{t\in T}(\bigvee_{y\in Y}(\overline{P_t}(x,y)*\overline{R}(y,z)))] =$$

$$[\bigvee_{t\in T}(\underline{P_t}\circ\underline{R})(x,z), \bigvee_{t\in T}(\overline{P_t}\circ\overline{R})(x,z)] = \bigvee_{t\in T}(P_t\circ R)(x,z).$$

□

Here we discuss the most important property of binary operations, i.e. the associativity. The associativity of $*$ is not sufficient for the associativity of the $\sup -*$ composition. Some results of this problem we see in ([8], Theorem 9) for t-norms or t-conorms on finite sets. As a result the question about the associativity of the composition of interval-valued fuzzy relations we get for arbitrary sets X, Y, Z, U the following lemmas.

Lemma 4. *If an operation $*$ is associative and infinitely sup-distributive, then* $\sup -*$ *composition is associative.*

Proof. Let operation $*$ be associative and infinitely sup-distributive, $P \in IVFR(X\times Y), R \in IVFR(Y\times Z)$, $Q \in IVFR(Z\times U)$ and $x \in X, y \in Y, z \in Z, u \in U$.

$$((P\circ R)\circ Q)(x,u) = [((\underline{P}\circ\underline{R})\circ\underline{Q})(x,u), ((\overline{P}\circ\overline{R})\circ\overline{Q})(x,u)] =$$

$$[\bigvee_z(\underline{P}\circ\underline{R})(x,z)*\underline{Q}(z,u), \bigvee_z(\overline{P}\circ\overline{R})(x,z)*\overline{Q}(z,u)] =$$

$$[\bigvee_z(\bigvee_y(\underline{P}(x,y)*\underline{R}(y,z))*\underline{Q}(z,u)),$$

$$\bigvee_z(\bigvee_y(\overline{P}(x,y)*\overline{R}(y,z))*\overline{Q}(z,u))] =$$

$$[\bigvee_z\bigvee_y(\underline{P}(x,y)*(\underline{R}(y,z)*\underline{Q}(z,u))),$$

$$\bigvee_z\bigvee_y(\overline{P}(x,y)*(\overline{R}(y,z)*\overline{Q}(z,u)))] =$$

$$[\bigvee_y(\underline{P}(x,y)*\bigvee_z(\underline{R}(y,z)*\underline{Q}(z,u))),$$

$$\bigvee_y(\overline{P}(x,y)*\bigvee_z(\overline{R}(y,z)*\overline{Q}(z,u)))] =$$

$$[(\underline{P}\circ(\underline{R}\circ\underline{Q}))(x,u), (\overline{P}\circ(\overline{R}\circ\overline{Q}))(x,u)] = (P\circ(R\circ Q))(x,u).$$

□

As a direct consequence of the above lemmas we observed that set of all interval-valued fuzzy relations with the composition (1) create a semigroup.

Proposition 1. *If $*$ is associative, infinitely sup-distributive operation with a zero element z=0 and a neutral element e=1, then $(IVFR(X),\circ)$ is an ordered semigroup with the identity $\boldsymbol{I} = [I, I]$.*

In the sequel we denote by D the set of all binary operations $* : [0,1]^2 \to [0,1]$ which are associative and infinitely sup-distributive (these conditions imply that $*$ is isotonic). As a result a special case of $*$ may be a left-continous triangular norm or conorm. If $* \in D$, then in a semigroup $(IVFR(X),\circ)$ we can consider the powers of its elements, i.e. relations R^n for $R \in IVFR(X)$, $n \in \mathbb{N}$. By analogy to [21] we define

Definition 7. By the powers of a relation $R \in IVFR(X)$ we call interval-valued fuzzy relations

$$R^1 = R,\ R^{m+1} = R^m \circ R, \text{ where } m = 1, 2, \ldots .$$

By the upper closure $R^\vee$ and the lower closure $R^\wedge$ of the relation R we call, respectively

$$R^\vee = \bigvee_{k=1}^{\infty} R^k,\ \ R^\wedge = \bigwedge_{k=1}^{\infty} R^k, \text{ where } R^k = [\underline{R}^k, \overline{R}^k]. \quad (4)$$

Proposition 2. *If $* \in D$ and $P, R \in IVFR(X)$, then*

$$\bigvee_{n\in\mathbb{N}}(P\vee R)^n \geq P^n \vee R^n,\ \ (P\wedge R)^n \leq P^n \wedge R^n, \quad (5)$$

$$(P\vee R)^\vee \geq P^\vee \vee R^\vee,\ \ (P\wedge R)^\vee \leq P^\vee \wedge R^\vee, \quad (6)$$

$$(P\vee R)^\wedge \geq P^\wedge \vee R^\wedge,\ \ (P\wedge R)^\wedge \leq P^\wedge \wedge R^\wedge. \quad (7)$$

Proof. Let $n \in \mathbb{N}$. From the isotonicity of $*$ and the Lemma 1 we know that the $\sup -*$ composition is also isotonic. As a result we obtain the isotonicity for powers. Then by

$$P\vee R \geq P,\ P\vee R \geq R$$

$$(\underline{P}\vee\underline{R} \geq \underline{P},\ \underline{P}\vee\underline{R} \geq \underline{R},\ \overline{P}\vee\overline{R} \geq \overline{P},\ \overline{P}\vee\overline{R} \geq \overline{R})$$

we have

$$(\underline{P}\vee\underline{R})^n \geq \underline{P}^n,\ (\underline{P}\vee\underline{R})^n \geq \underline{R}^n \text{ and}$$

$$(\overline{P}\vee\overline{R})^n \geq \overline{P}^n,\ (\overline{P}\vee\overline{R})^n \geq \overline{R}^n,$$

so we obtain

$$(P\vee R)^n \geq P^n,\ (P\vee R)^n \geq R^n \Rightarrow (P\vee R)^n \geq P^n \vee R^n.$$

Similarly, we can prove $(P\wedge R)^n \leq P^n \wedge R^n$. Moreover, supremum and infimum are isotonic, so from the condition (5) also closures have this property
$P^\vee \leq (P\vee R)^\vee$, $R^\vee \leq (P\vee R)^\vee \Rightarrow P^\vee \vee R^\vee \leq (P\vee R)^\vee$,
$(P\wedge R)^\vee \leq P^\vee$, $(P\wedge R)^\vee \leq R^\vee \Rightarrow (P\wedge R)^\vee \leq P^\vee \wedge R^\vee$.
Similarly, we may prove the inequalities in (7). □

Proposition 3. *Let $* \in D$ and $P, R \in IVFR(X)$.*
If $P\circ R = R\circ P$, then

$$\bigvee_{n\in\mathbb{N}}(P\circ R)^n = P^n \circ R^n, \quad (8)$$

$$(P\circ R)^\vee \leq P^\vee \circ R^\vee,\ \ (P\circ R)^\wedge \geq P^\wedge \circ R^\wedge. \quad (9)$$

ISBN: 978-989-95079-6-8

Proof. The given equality in (8) may be proved by the mathematical induction and associativity of $*$ and commutativity of powers, which is implied by commutativity of P and R. Then

$$(P \circ R)^{\vee} = \bigvee_{k=1}^{\infty} (P \circ R)^k = [\bigvee_{k=1}^{\infty} (\underline{P}^k \circ \underline{R}^k), \bigvee_{k=1}^{\infty} (\overline{P}^k \circ \overline{R}^k)],$$

by (3), (2) and isotonicity of the operation $\sup - *$ we have

$$[\bigvee_{k=1}^{\infty} (\underline{P}^k \circ \underline{R}^k), \bigvee_{k=1}^{\infty} (\overline{P}^k \circ \overline{R}^k)] \leq [\bigvee_{k=1}^{\infty} (\underline{P}^{\vee} \circ \underline{R}^k), \bigvee_{k=1}^{\infty} (\overline{P}^{\vee} \circ \overline{R}^k)]$$

$$\leq [\underline{P}^{\vee} \circ \bigvee_{k=1}^{\infty} \underline{R}^k, \overline{P}^{\vee} \circ \bigvee_{k=1}^{\infty} \overline{R}^k] = [\underline{P}^{\vee} \circ \underline{R}^{\vee}, \overline{P}^{\vee} \circ \overline{R}^{\vee}] = P^{\vee} \circ R^{\vee},$$

which proves the first condition in (9). The second condition in (9) one may be justified in a similar way. □

3 Properties of interval-valued fuzzy relations

Now, we will examine whether the given properties are preserved by the composition of interval-valued fuzzy relations. We see that many properties of the $*$ operation are transposed to the operation of $\sup - *$ composition but not all of them. Namely, if $* = \min$, then $\sup - \min$ composition is not commutative. Now, we examine very interesting properties, namely subidempotency and superidempotency. These properties are of the large interest for example in economy where they are applied in valuation of supply and demand. Similarly to definitions of properties of fuzzy relations in [21] we have

Definition 8. Let $R \in IVFR(X)$. The relation R is called idempotent, subidempotent (transitive) or superidempotent if $R^2 = R,\ R^2 \leq R,\ R^2 \geq R$, respectively.

For the Boolean matrices of dimension $n \times n$ the following computations hold true

Table 1: The subidempotent relations in the family of all relations.

n	all	subidempotent	%
2	16	13	81,25000
3	512	171	33,39844
4	65536	3994	6,09436
5	33554432	154303	0,45986

A similar situation takes place in any distributive and bounded lattice, it means that the percentage of the subidempotent matrices is rapidly decreasing with the growth of n. This is why the consideration and determination of the upper closure $R^{\vee}$ is very important.

Theorem 2. *Let $* \in D$ and $R \in IVFR(X)$. $R^{\vee}$ is the least subidempotent relation greater than or equal to R. Moreover, the relation R is subidempotent if and only if $R = R^{\vee}$.*

Proof. Let $R, S, Q \in IVFR(X)$. If R is subidempotent, then

$$\bigwedge_{n \in \mathbb{N}} R^n \leq R \text{ and } R \leq R^{\vee} = \bigvee_{n \in \mathbb{N}} R^n \leq R,$$

so we obtain $R^{\vee} = R$.
If $R^{\vee} = R$, then for $S = R^{\vee}$ we have

$$S^2 = \bigvee_{k=2}^{\infty} R^k \leq R^{\vee} = S$$

and $R^{\vee}$ is subidempotent.
We show that if there exists $Q = [\underline{Q}, \overline{Q}]$ such that

$$[\underline{R}, \overline{R}] \leq [\underline{Q}, \overline{Q}] \text{ and } [\underline{Q}^2, \overline{Q}^2] \leq [\underline{Q}, \overline{Q}],$$

then by isotonicity the $\sup - *$ composition (Lemma 1) we obtain

$$[\underline{R}^2, \overline{R}^2] \leq [\underline{R} \circ \underline{Q}, \overline{R} \circ \overline{Q}] \leq [\underline{Q}^2, \overline{Q}^2]$$

thus

$$[\underline{R}^k, \overline{R}^k] \leq [\underline{Q}^k, \overline{Q}^k] \leq [\underline{Q}, \overline{Q}] \text{ for } k \in \mathbb{N}.$$

So by isotonicity of the supremum $R^{\vee} \leq Q$. □

The closures and powers of interval-valued fuzzy relations also preserve some properties of such relations.

Theorem 3. *Let an operation $* \in D$ be commutative and $R \in IVFR(X)$. If R is subidempotent (superidempotent), then R^n, $R^{\wedge}$ ($R^{\vee}$) are subidempotent (superidempotent).*

Proof. Let $R^2 \leq R$. Then R^n are also subidempotent, $n \in \mathbb{N}$. By (2) and subidempotency of R we have

$$(R^{\wedge})^2 \leq (R^2)^{\wedge} \leq R^{\wedge},$$

so $R^{\wedge}$ is subidempotent. The property of superidempotency may be proved analogously. □

Theorem 4. *Let $T \neq \emptyset$ and $* \in D$ and $R_t \in IVFR(X)$, $t \in T$. If $(R_t)_{t \in T}$ is a family of subidempotent relations, then the relation $R = \bigwedge_{t \in T} R_t$ is subidempotent.*

Proof. Let $R_t^2 \leq R_t$, $t \in T$, then by (2)

$$R^2 = (\bigwedge_{t \in T} R_t)^2 = (\bigwedge_{s \in T} R_s) \circ (\bigwedge_{t \in T} R_t) \leq$$

$$\bigwedge_{t \in T} ((\bigwedge_{s \in T} R_s) \circ R_t) \leq \bigwedge_{s,t \in T} (R_s \circ R_t) \leq \bigwedge_{t \in T} R_t = R$$

i.e., the relation R is subidempotent. □

Theorem 5. *Let $* \in D$ and $P, R \in IVFR(X)$. If P, R are subidempotent, then $P \vee R$ is subidempotent if and only if $P \circ R \vee R \circ P \leq P \vee R$.*

Proof. Since $P^2 \leq P$, $R^2 \leq R$ so by (3) we obtain

$$(P \vee R)^2 = [(\underline{P} \vee \underline{R}) \circ (\underline{P} \vee \underline{R}), (\overline{P} \vee \overline{R}) \circ (\overline{P} \vee \overline{R})] =$$

$$[(\underline{P} \circ \underline{P}) \vee (\underline{R} \circ \underline{P}) \vee (\underline{P} \circ \underline{R}) \vee (\underline{R} \circ \underline{R}),$$
$$(\overline{P} \circ \overline{P}) \vee (\overline{R} \circ \overline{P}) \vee (\overline{P} \circ \overline{R}) \vee (\overline{R} \circ \overline{R})] \leq$$

$$[\underline{P} \vee (\underline{R} \circ \underline{P}) \vee (\underline{P} \circ \underline{R}) \vee \underline{R}, \overline{P} \vee (\overline{R} \circ \overline{P}) \vee (\overline{P} \circ \overline{R}) \vee \overline{R}] =$$

$$[\underline{P} \vee \underline{R} \vee (\underline{R} \circ \underline{P}) \vee (\underline{P} \circ \underline{R}), \overline{P} \vee \overline{R} \vee (\overline{R} \circ \overline{P}) \vee (\overline{P} \circ \overline{R})],$$

as a result

$$(P \vee R)^2 \leq P \vee R \Leftrightarrow P \circ R \vee R \circ P \leq P \vee R,$$

because

$$P \leq R \Leftrightarrow P \vee R = R.$$

□

ISBN: 978-989-95079-6-8

Theorem 6. *Let $* \in D$ and $P, R \in IVFR(X)$. If P, R are subidempotent (superidempotent) and $P \circ R = R \circ P$, then $P \circ R$ is subidempotent (superidempotent).*

Proof. If $\underline{P} \circ \underline{R} = \underline{R} \circ \underline{P}$ and $\overline{P} \circ \overline{R} = \overline{R} \circ \overline{P}$, $P^2 \leq P$, $R^2 \leq R$, then by the associativity and monotonicity of the $\sup -*$ composition we have

$$(P \circ R)^2 = [(\underline{P} \circ \underline{R})^2, (\overline{P} \circ \overline{R})^2] =$$

$$[\underline{P} \circ (\underline{R} \circ \underline{P}) \circ \underline{R}, \overline{P} \circ (\overline{R} \circ \overline{P}) \circ \overline{R}] =$$

$$[\underline{P} \circ \underline{P} \circ \underline{R} \circ \underline{R}, \overline{P} \circ \overline{P} \circ \overline{R} \circ \overline{R}] = [\underline{P}^2 \circ \underline{R}^2, \overline{P}^2 \circ \overline{R}^2] \leq$$

$$[\underline{P} \circ \underline{R}, \overline{P} \circ \overline{R}] = P \circ R.$$

As a result $P \circ R$ is subidempotent. The proof for superidempotency is similar. □

4 Conclusion

In this work we present only some problems connected with the preservation of interval-valued fuzzy relation properties by the $\sup -*$ composition and related to it operation. We can also consider preservation of other properties (for example, symmetry, asymmetry, antisymmetry, reflexivity, irreflexivity, connectedness) by the composition of interval-valued fuzzy relations and also by its powers and lattice operations.
We may consider dual composition to the one defined in (1). This is the $\inf -*'$ composition with the dual binary operation $*'$, where $x *' y = 1 - (1 - x) * (1 - y)$ for $x, y \in [0, 1]$. The properties of this composition may be deduced from the $\sup -*$ composition.
Moreover, we may study interval-valued t-norms and t-conorms as operations on *Int*([0,1]) which are important functions because they are useful in approximate reasoning, in medical diagnosis and information retrieval. For example, the authors of [7, 10] examine the construction of t-norms and t-conorms in the lattice $(Int[0, 1], \vee, \wedge)$ and analyze some properties of them.

References

[1] E. Agustench, H. Bustince and V. Mohedano. Method of least squares applied to the GMP with international-valued fuzzy sets. *Mathware & Soft Computing*, 6(2-3):267–276, 1999.

[2] P.P. Angelov, Optimization in an intuitionistic fuzzy environment, *Fuzzy Sets and Systems*, 86:299–306, 1997.

[3] K.T. Atanassov, Intuitionistic fuzzy sets, *Fuzzy Sets and Systems*, 20:87–96, 1986.

[4] K.T. Atanassov, *Intuitionistic fuzzy sets*, Physica-Verlag, New York, 1999.

[5] K.T. Atanassov and G. Gargov, Interval valued intuitionistic fuzzy sets, *Fuzzy Sets and Systems*, 31:343–349, 1989.

[6] G. Birkhoff, *Lattice theory*, AMS Coll. Publ. 25, Providence 1967.

[7] H. Bustince, E. Barrenechea and M. Pagola. Generation of interval-valued fuzzy and Atannasov's intuitionistic fuzzy connectives from fuzzy connectives and from K_α operators: laws for conjunction and disjunction, amplitude. *International Journal of Intelligent Systems*, 23:680-714, 2008.

[8] H. Bustince and P. Burillo. Mathematical analysis of interval-valued fuzzy relations: Application to approximate reasoning. *Fuzzy Sets and Systems*, 113:205-219, 2000.

[9] J.T. Cacioppo, W.L. Gardner and G.G. Berntson. Beyond bipolar conceptualization and measures: the case of attitudes and evaluative spaces. *Personality and Social Psychology Review*, 1:3–25, 1997.

[10] G. Deschrijver. A representation of t-norms in interval-valued L-fuzzy set theory. *Fuzzy Sets and Systems*, 159:1597-1618, 2008.

[11] G. Deschrijver and E.E. Kerre, On the relationship between some extensions of fuzzy set thory, *Fuzzy Sets and Systems*, 133(2):227–235, 2003.

[12] G. Deschrijver and E.E. Kerre, On the position of intuitionistic fuzzy set theory in the framework of theories modelling imprecision, *Information Sciences*, 177:1860–1866, 2007.

[13] J. Drewniak and K. Kula, Generalized compositions of fuzzy relations, *International Journal of Uncertainty, Fuzziness Knowledge-Based Systems*, 10:149-163, Suppl., 2002.

[14] J. Drewniak and B. Pękala. Properties of fuzzy relations powers. *Kybernetika*, 43(2):133-142, 2007.

[15] D. Dubois, S. Gottwald, P. Hajek, J. Kacprzyk and H. Prade, Terminological difficulties in fuzzy set theory - The case of "Intuitionistic Fuzzy Sets", *Fuzzy Sets and Systems*, 156:485–491, 2005.

[16] D. Dubois, H. Prade and S. Sessa. Recent literature. *Fuzzy Sets and Systems*, 159:857–865, 2008.

[17] D. Dubois and H. Prade. An introduction to bipolar representations of information and preference, *International Journal of Intelligent Systems*, 23(8):866–877, 2008.

[18] A. Goguen. L-fuzzy sets. *Journal of Mathematical Analysis and Applications*, 18:145–174, 1967.

[19] M.B. Gorzałczany. A method of inference in approximate reasoning based on interval-valued fuzzy sets. *Fuzzy Sets and Systems*, 21(1):1-17, 1987.

[20] Y.-Y. Guh, M.-S. Yang, R.-W. Po and E.S. Lee. Interval-valued fuzzy relation-based clustering with its application to performance evaluetion. *Computers and Mathematics with Applications*, 57:841–849, 2009.

[21] A. Kaufmann. *Introduction to the theory of fuzzy subsets*, Acad. Press, New York, 1975.

[22] D.-F. Li. Multiattribute decision making models and method using intuitionistic fuzzy sets, *Journal of Computer and System Sciences*, 70:73–85, 2005.

[23] L. Lin, X-H. Yuan and Z-Q. Xia. Multicriteria fuzzy decision-making methods based on intuitionistic fuzzy sets, *Journal of Computer and System Sciences*, 73:84–88, 2007.

[24] E.P. Klement, R. Mesiar and E. Pap. *Triangular norms*. Kluwer Acad. Publ., Dordrecht, 2000.

[25] B. Pękala. Operations on interval matrices, in: M. Kryszkiewicz et al. (eds.), *Rough Sets and Intelligent Systems Paradigms*, Springer-Verlag, Berlin, 613–620, 2007.

[26] R. Sambuc. Fonctions ϕ-floues. Application á l'aide au diagnostic en pathologie thyroidienne. Ph.D. Thesis, Université de Marseille, France, 1975.

[27] L. Sánchez and I. Couso. Advocating the use of imprecisely observed data in genetic fuzzy systems. *IEEE Transaction on Fuzzy Systems*, 15(4):551–562, 2007.

[28] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[29] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning-I, *Information Sciences*, 8:199–249, 1975.

ISBN: 978-989-95079-6-8

A fuzzy approach for the model of sliding window. An application to behaviour patterns mining

María Ros, Miguel Delgado, Amparo M. Vila[1]

1.Department of Computer Science and A.I. University of Granada, España
Email: {marosiz, mdelgado, vila}@decsai.ugr.es

Abstract— *Many traditional environment applications base their operations on using sensor. For instance, the Tagged World uses the information from sensors to identify user behaviour and to provide services. We present a method to limit an user behaviour through external user knowledge: a Temporal Window. However, an user behaviour has random imprecision by definition. Thus, we have adapted the method to manage the imprecision, Fuzzy Temporal Window, which includes Fuzzy Logic concepts. As an application, we present a method to extract behaviour patterns on time defined by a Fuzzy Temporal Window.*

Keywords— Behaviour, Fuzzy Windows, Quantified Sentences, Tagged World, Frequent Itemsets

1 Introduction

In the last years, Smart Computers have suffered a great evolution that is transforming the world. *Ubiquitous Computing*[1] uses them to make user daily life easier. Computers are included on daily life without disturbing normal user activities. Different devices are added in traditional environment from common PDA or Mobile Phone to sensors. Nowadays, the last trend in *Ubiquitous Computing* is to provide the environment with different mechanisms to get information.

Sensors are devices that are able to know what is happening in the environment. There are many kind of sensors from temperature sensor to mechanical one. We also find RFID sensors [2] that identify every object in the environment with an unique number by means of radio frequency. Placing RFID tags in an environment we get to construct an intelligent space in which everyone can enjoy services. An intelligent space obtains user position, user behaviour, and environment information around a user and so on by a sensor network [3]. Other example about these spaces can be found in [4]. We find an intelligent space that provides services in a variety of scenes, but hardly ever, without reasoning or inference. The development of these projects is the Tagged World concept.

In [5, 3], the Tagged World project is presented. This project is developed in the University of Ritsumeikan, Japan and consists on providing appropriate personalized services for each user, to make their life easier and safer by recognizing and reasoning the human behaviour. They use a wearable computer as a Pocket Assistant that compares an access log with patterns to recognize human activities. The system is based on a Bayesian network and obtains results with a probability value for every extracted behaviour.

Other alternative is proposed by Philipose and et al. [6]. They proposed a system to infer Activities of Daily Living in a Tagged World. They present a new paradigm for ADL inferencing leverage radio-frequency-identification technology, data mining and a probabilistic inference engine to recognize ADLs based on objects that people use. As sensor, they use RFID technology with other sensor streams to fill in the gaps. The system represents activities as linear sequences of activity stages, and annotates each stage with the involved objects and the probability of their involvement.

In [7][8] is proposed a system to identify correct behaviour using Data Mining Techniques. The system is divided into two main parts: inductive learning mechanism, which produces a behaviour database and a reasoning system for the recognition of sequences that uses this database. The first stage uses Frequent Itemsets, while the second one does Regular Grammar.

This paper is organized in six main sections. In Section 2 we present the formal problem to solve. In Section 3, Fuzzy Temporal Window concept is explained. Section 4 presents the method that manage problem uncertainty. In section 5, we show some empirical results, comparing crisp and fuzzy models. Finally, the conclusions and future works are reported in section 6.

2 Formal Problem

This section defines the formal problem: to obtain sequence patterns to identify user behaviour that are defined on a specific domain and context. Thus, we have to define some basic concepts which are the objectives of the system.

DEFINITION **2.1 (Action)** *An action, a, is an activity that happens using a specific object.*

However, it results very interesting to know when an action happens on time.

DEFINITION **2.2 (Action on time)** *An action on time, a, is an activity that happens over a specific object in a known time and it is denoted as a pattern $a = (h, l)$, where h is a fact and l is a temporal label that defines the time of the action.*

These definitions give us a basic element to work, but our aim is to find different actions that make a behaviour up.

DEFINITION **2.3 (Behaviour)** *Let $A = \{a_1, ..., a_n\}$ be set of possible user actions in some situations or domains. An user behaviour is a finite set of actions:*

$$\beta = \{\alpha_1, \alpha_2, , ...\alpha_{p(\beta)}\}$$

with $\alpha_j \in A \forall j$, and where α_j is performed before α_k iif $j \leq k$.

ISBN: 978-989-95079-6-8

In this definition, we do not pay attention on the time, although, in general, a behaviour happens in a specific moment on time.

DEFINITION **2.4 (Behaviour on time)** *Let $A = \{a_1, ..., a_n\}$ be set of possible user actions in some situations or domains. Let τ be the temporal line, then A user behaviour on time is a finite set of pair such as:*

$$b_I = \{(\alpha_1, t_1), ...(\alpha_{n(b)}, t_{n(b)})\}$$

where $\alpha_j \in A \forall j$, where α_j is performed before α_k iif $j \leq k$, $t_j \in [0, \tau]$ $\forall t_i < t_j$ if $i \geq j$, with $I = [t_1, t_{n(b)}]$

The System is based on a database about user actions. This database has been obtained from user observation, i.e., from actions that user has done. In rest of this document, we name this database as Observation Data Base (ODB).

The representation of ODB is a transactional database T, where every row is an observation over an user. Normally, we study his activity in a whole day; and every column is a possible action from A set, $A = \{a_1, ..., a_n\}$.

3 Fuzzy Temporal Windows. Extract behaviour patterns tool.

In previous section, we have presented behaviour sequence pattern problem. In [7][8] we study behaviours in a crisp way, however the human behaviour is not crisp. Human activity has random imprecision by definition.

3.1 Motivations

In [7][8] we present a process to obtain correct sequence patterns of actions from user behaviour. This method extracts from a particular ODB common actions. A general ODB contains actions that a user has realized for a whole day. For this reason, we do not know where a behaviour starts or ends. The proposed method needs an expert to indicate when a behaviour is performed. This proposal is not applicable in a real system, since every person has his/her specific activities and habits.

Therefore, we have to design a mechanism to identify the interval of the ODB that we have to study. In this point, we use a knowledge about the behaviour :generally, there exist some behaviours that are realized by user at the same time every day.

EXAMPLE **3.1** *Luis leaves home at 8:30 o'clock.*

We use this knowledge to situate an interval on the temporal line. The result is a subset from particular ODB to the studied behaviour.

In spite of this fact, nobody usually does an action at the exactly moment every day. They usually do actions roughly at the same moment. This raises a new problem: how we fix the interval's ranges to detect a specific activity. Ranges should consider random imprecision of the situations.

EXAMPLE **3.2** *Let us suppose that* Luis often leaves home at 8:30 o'clock*, then we could control actions that happen about 8:30 (from 8:20 to 8:40, for instance).*

We name the interval defined over temporal line as *Temporal Window*. It permits to get a subset from each ODB tuple (See figure 1). In this subset, we know that actions have not the same degree of importance, since actions are about the specific time are more important than actions far from the interval centre. So, we could assign a degree of importance at each action, in function of its situation in the interval. Then, we can define the interval as a fuzzy set and name it as *Fuzzy Temporal Window* (See figure 2).

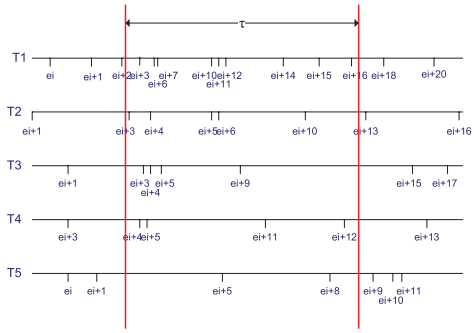


Figure 1: Temporal Window over a ODB

3.2 Model Formulation

In this section we present a formal representation to *Temporal Window* and *Fuzzy Temporal Window*.

DEFINITION **3.1 (Temporal Window, W)** *Let I be a interval from the temporal line τ, ODB the Observation Data Base and $t \in ODB$ a tuple from ODB. Let i_j be an action $i_j = (h_j, l_j)$, then a Temporal Window, W, for a specific behaviour, which happens in interval I of τ, is defined as a subset of t where*

$$\forall i_j \in W(t) \text{ then } i_j \in t \text{ and } l_j \in I$$

DEFINITION **3.2 (Fuzzy Temporal Window, FW)** *Let I be a interval from the temporal line τ, ODB the Observation Data Base and $t \in ODB$ a tuple from ODB. Let i_j be an action $i_j = (h_j, l_j)$ and a Temporal Window, W, to a specific behaviour. Let fs be a fuzzy set over τ, then it defines a* Fuzzy Temporal Window*, FW, as a Temporal Window where*

$$\forall i_j \in W(t)\ \mu_{FW}(i_j) = \mu_{fs}(l_j)$$

If we apply a Fuzzy Temporal Window over the ODB, we obtain an image where each item has its membership degree. As we have represented an ODB as a Transactional Database, we obtain a Transactional Database image too.

DEFINITION **3.3 (Fuzzy ODB applying W)** *Let ODB be an Observation DataBase and W a Fuzzy Temporal Window, it defines $\tilde{ODB}$ as a Fuzzy Observation Data Base constructs as W(ODB) such as*

$$\forall t \in ODB, W(t) \in \tilde{ODB}$$

DEFINITION **3.4 (Fuzzy Transactional Data Base)** *Let T be a Transactional Data Base and W a Fuzzy Temporal Window, it defines $\tilde{T}$ as a Fuzzy Transactional Data Base constructs as W(t) $\forall t \in T$.*

We represent the Fuzzy Transactional DataBase as a table where for every row and column we have membership degree corresponding to the Fuzzy Temporal Window.

ISBN: 978-989-95079-6-8

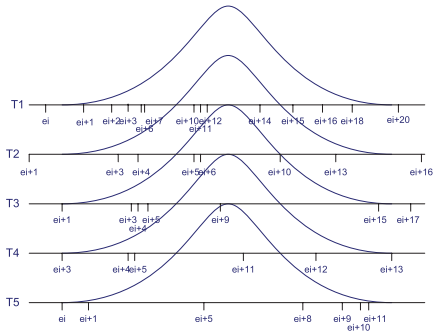


Figure 2: Fuzzy Temporal Window over a ODB

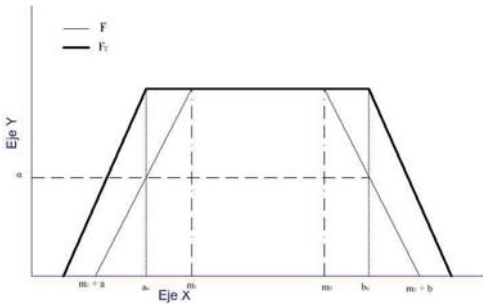


Figure 3: Transformations over a Trapezoidal Fuzzy Temporal Windows

3.3 Extending the Fuzzy Temporal Windows.

Until this point, we establish a Fuzzy Temporal Window with previous behaviour knowledge. We suppose this knowledge is completely true, although this statement always is not correct.

EXAMPLE **3.3** *Let us suppose that we want to check the behaviour* `Luis leaves home at 8:30 o'clock`. *However, we know that Luis sometimes arrives late at work. So, we can affirm that* `Luis leaves home at time almost always`.

In the example 3.3, we give a quantified adjective to actions of a behaviour to express certain. We employ a semantic approach based on evaluation of quantified sentences [9]. A quantified sentence is an expression of the form "Q of F are G", where F and G are two fuzzy subsets of a finite set X, and Q is a relative fuzzy quantifier. Some examples could be (The most times, Luis leaves home, 8:30), (Almost never, Luis leaves home, on time), (Almost all times, Luis leaves home, 8:30).

This knowledge is used to expand the Fuzzy Temporal Window size. We use the method presented in [10]. This method transforms the used function to define a specific behaviour according with the knowledge expressed by a quantified sentence. The proposed process obtains a new window F' in two steps:

1. Firstly, they truncate a fuzzy number, our Fuzzy Temporal Window, using the certainty degree α associated to the fuzzy value A. After this operation, we obtain a non normalized fuzzy set A^{α}.
2. Secondly, they normalize the fuzzy set. The authors assume that uncertainty is being translated into imprecision under the condition of the amount of information provided by the fuzzy number remains equal before and after normalization process.

The transformation function is defined as:

DEFINITION **3.5** *Let $A \in \tau$ be a fuzzy number such that $A = \{(m_1, m_2, a, b), \alpha_A\}$, where m_1, m_2, a, b are the values that defines a trapezoidal fuzzy number and α_A is the height of A. Let $\alpha \in (0, 1]$. We will denote $\triangle(\alpha_A, \alpha) = \triangle$ and define*

$$T_{\alpha}(A) = \left\{ \left(m_1, m_2, a + \frac{\triangle}{k}, b + \frac{\triangle}{k} \right), \alpha \right\} \quad (1)$$

for those α in which the transformation makes sense.

To use the definition 3.5, we need to define two basic parameters:

- k.- scale parameter
- α.- value for doing the transformation

In figure 3 we represent the followed process over a trapezoidal window.

In this point, we introduced the quantified sentences over a Fuzzy Temporal Windows, using their evaluation to know α value. When we have this value, we transform A set in its α version (A^{α}), which is the basis of the final transformation A^{T}. Example 3.4 shows the followed process. We use a trapezoidal Fuzzy Temporal Window for the sake of clarity.

EXAMPLE **3.4** *Let us suppose that we want to check the behaviour* `Luis hardly ever leaves home at 8:30 o'clock`. *Thus, our Fuzzy Temporal Windows, represented as a lr-number as $F = (m_1, m_2, a, b)$ where m_1, m_2 is the higher points of a trapezoidal fuzzy number, and a, b are the time to establish the lower points.*

$$F = (8:25, 8:35, 0:05, 0:05) \quad (2)$$

Firstly, we have to evaluate a quantified sentence to find out the α-value. There exists a lot of ways to evaluate it. Here we have selected the basic way, but not the best: Zadeh. The expression of the Zadeh Cardinal is $\alpha_A = \frac{P(A)}{|X|}$ where $P(A) = sum_{x \in X} A(x)$ and $|X|$ is the X set cardinal.

So, $\alpha_A = \frac{15}{20} = 0.75$. The evaluation of the quantified sentence is $Z_Q(A) = Q(\alpha_A) = 0.53$.

Next, we do the transformation of our windows using α value.

$$F = \{(8:25, 8:35, 0:05, 0:05), 1\} \rightarrow \quad (3)$$

$$F^{\alpha} = \left\{ \begin{pmatrix} 8:25 - 0:05(1-\alpha), 8:35 - \\ 0:05(1-\alpha), 0:05\alpha, 0:05\alpha \end{pmatrix}, \alpha \right\} \rightarrow \quad (4)$$

$$F^{T} = \left\{ \begin{pmatrix} 8:25 - 0:05(1-\alpha), 8:35 - 0:05 \\ (1-\alpha), 0:05\alpha - \frac{1-\alpha}{k\alpha}, 0:05\alpha - \frac{1-\alpha}{k\alpha} \end{pmatrix}, \alpha \right\} \quad (5)$$

Replacing α value with $Z_Q(A)$, i.e., $\alpha = 0.53$ we expand our Fuzzy Temporal Windows to

$$F = (8:23, 8:37, 0:08, 0:08) \quad (6)$$

ISBN: 978-989-95079-6-8

4 Fuzzy method to extract behaviour patterns. Obtaining sequence patterns by alpha-cuts

Once we have information, we should get common actions and valid sequences for every behaviour. However, we can not apply the method explained in [7] [8] directly, because we have to manage membership degree corresponded to the Fuzzy Temporal Window.

Here, we cannot present our Crisp model but we have to clarify some basic concepts of it that are need to can understand the process.

4.1 Previous concepts

In this section, we want to specify some important concepts: common actions and valid sequences.

1. Common actions: We define the common actions as a sequence of events that occurs more often in the observed knowledge. We can identify the common actions with the concept of Frequent Itemset: every frequent itemset corresponds to a particular common behaviour [8]. However, an itemset is a set and, so, it has not order relationship among its elements. However, a basic condition for a behaviour is the existence of this relationship. Since, we have to control which sequences of these itemsets are valid.

2. Valid/Correct sequences We define the valid sequences as the permutations of the set whose order relationship among their elements appear in the knowledge of the *ODB*. We use the permutation concept and an designed algorithm to determine if a sequence is valid or not, since not all permutation will be accepted as a pattern sequence.

4.2 Changes to manage the random imprecision

In this subsection, we present the different changes to manage the random imprecision. Thus, the objective is to develop a system that obtains stimulation from user who realizes some actions sequentially. The problem consists on extracting the sequence patterns to specific behaviour when we have a Fuzzy ODB $O\tilde{D}B$, represented as $\tilde{T}$, and a Fuzzy Temporal Window W defined from user knowledge.

Firstly, we have to transform a fuzzy problem to a crisp problem to apply the method explained in [7] [8]. We could use α-cut set [11]. By the α-cut, we would have a new crisp image of $\tilde{T}$, T^{α} where every value is in $\{0, 1\}$. Then, we apply classical method to T^{α}, obtaining frequent itemsets and sequence patterns to specific α value (I^{α} and P^{α}, respectively). After extracting all sequence patterns, we create a fuzzy set **P** by sequence patterns which have obtained to every α-value applying the Representation Theorem.

As we use frequent itemsets, we need to represent extracted frequent itemsets as an unique fuzzy set. Thus, we have to ensure *consistent restriction* between every α frequent itemset I^{α_i}. For sequence patterns representation is the same. The proof of these statements can be studied in [8].

4.3 An illustrative example

In this example we want to show the operation of the method. We start from ODB which is collected from the touched object by the user in the daily activity[1].

Let W a Fuzzy Temporal Window defines by the following fuzzy set over temporal line τ:

$$W = \left\{ \begin{array}{c} \frac{0}{8:20}, \frac{0,2}{8:21}, \frac{0,4}{8:22}, \frac{0,6}{8:23}, \frac{0,8}{8:24}, \frac{1}{8:25}, \frac{1}{8:26} \\ \frac{1}{8:27}, \frac{1}{8:28}, \frac{1}{8:29}, \frac{1}{8:30}, \frac{1}{8:31}, \frac{1}{8:32}, \frac{1}{8:33}, \frac{1}{8:34} \\ \frac{1}{8:35}, \frac{0,8}{8:36}, \frac{0,6}{8:37}, \frac{0,4}{8:38}, \frac{0,2}{8:39}, \frac{0}{8:40} \end{array} \right\} \quad (7)$$

Then, we apply W over T and obtain a Fuzzy T $\tilde{T}$ where every time of actions is replaced with the membership degree value in W. We apply the α-cut for α values α_1 = 0.4, α_2 = 0.6, α_3 = 0.8, α_4 = 1.0. We obtain for each α values the T^{α}, the result of apply the α-cut in $\tilde{T}$. After these operations, we have transformed the fuzzy problem to a crisp problem. Now, we extract frequent itemsets and sequence patterns with the model explained in [7] and [8].

We have executed the Apriori Algorithm with two support values: *minsup = 0.8 and minsup = 0.9*. The results are showed in tables 1 and table 2.

Since we have the common actions to specific behaviour, the next stage consists of obtaining the valid sequences patterns and the final representation as a fuzzy set applying the Identity Principle. The valid patterns are showed in table 3 and table 4. And the final pattern representation in equation 8 to minsup = 0.8 and equation 9 to minsup = 0.9.

Table 1: Frequent itemset to minsup = 0.8

I^{α_1}	{Shoes, Bag, Keys, ODoor, MobilePhone} {Keys, Keys2, ODoor, ODoor2, MobilePhone}
I^{α_2}	{Shoes, Bag, Keys, ODoor, MobilePhone} {Keys, Keys2, ODoor, ODoor2, MobilePhone}
I^{α_3}	{Shoes, Bag, Keys, ODoor, MobilePhone}
I^{α_4}	{Shoes, Bag, Keys, ODoor, MobilePhone}

Table 2: Frequent itemset to minsup = 0.9

I^{α_1}	{Keys, ODoor, MobilePhone}
I^{α_2}	{Keys, ODoor, MobilePhone}
I^{α_3}	{Keys, ODoor}
I^{α_4}	{ODoor}

Table 3: Valid patterns to minsup = 0.8

P^{α_1}	{Shoes, Bag, Keys, MobilePhone, ODoor}	p_1
	{Shoes, Keys, MobilePhone, Bag, ODoor}	p_2
	{Shoes, MobilePhone, Bag, Keys, ODoor}	p_3
	{Keys, MobilePhone, ODoor, Keys2, ODoor2}	p_4
	{MobilePhone, Keys, ODoor, Keys2, ODoor2}	p_5
P^{α_2}	IDEM	
P^{α_3}	{Shoes, Bag, Keys, MobilePhone, ODoor}	p_1
	{Shoes, Keys, MobilePhone, Bag, ODoor}	p_2
	{Shoes, MobilePhone, Bag, Keys, ODoor}	p_3
P^{α_4}	IDEM	

$$\tilde{P} = \left\{ \frac{p_1}{\alpha_4}, \frac{p_2}{\alpha_4}, \frac{p_3}{\alpha_4}, \frac{p_4}{\alpha_2}, \frac{p_5}{\alpha_2} \right\} = \left\{ \frac{p_1}{1}, \frac{p_2}{0.4}, \frac{p_3}{1}, \frac{p_4}{1}, \frac{p_5}{0.6} \right\} \quad (8)$$

[1] In this paper, we do not show the Data Base due to the lack of space.

ISBN: 978-989-95079-6-8

Table 4: Valid patterns to minsup = 0.9

P^{α_1}	{Keys, MobilePhone, ODoor}	p_1
	{MobilePhone, Keys, ODoor}	p_2
P^{α_2}	IDEM	
P^{α_3}	{Keys, ODoor}	p_3
P^{α_4}	{ODoor}	p_4

$$\tilde{P} = \left\{ \frac{p_1}{\alpha_2}, \frac{p_2}{\alpha_2}, \frac{p_3}{\alpha_3}, \frac{p_4}{\alpha_4} \right\} = \left\{ \frac{p_1}{0.6}, \frac{p_2}{0.6}, \frac{p_3}{0.8}, \frac{p_4}{1} \right\} \quad (9)$$

5 Performance Evaluation

In this section, our objective is to evaluate fuzzy model versus crisp model presented in [7][8]. In other hand, we present advantages of modifying the windows using quantified sentences. To do this proof, we used a database that represents daily activities of the user. To be precise, we focus study on the analysis of Leave Home behaviour, to simplify the process. We study two databases. These databases try to force the temporal model, thus they imply different temporal relations between their items. In the first one, the interval between two consecutive actions is always the same. In the second database, we break this temporal relation.

With these experiments, we want to control the fuzzy window influence over the database.

As well as, we need a goodness measure to analyze obtained outcomes. We design a measure to evaluate the method that works with a sequence of words. So, we have to design a way to transform these sequences in a numerical way. We have to design this measure because it tries to pay attention all important aspects: correction, rubbish, alpha, minsup, time, etc.

Therefore, we have to define a goodness measure that makes the result evaluation easier:

$$c = \alpha_1 correction + \alpha_2 alpha + \alpha_3 \frac{1}{time} \quad (10)$$

where,

- correction.- similarity measure of every pattern to the correct pattern. We define a function that associates a numeric value with sequences that a higher value indicates greater similarity.
- alpha.- mean of the alpha-values used for the α-cut for every pattern.
- time.- amount of millisecond that the method inverts in obtaining results.

Goodness measure is inversely proportion with time raises to the power of minus one and directly proportion with the correction and confidence measure. So, if the absolute value of the correction gets bigger, the absolute value of goodness measure gets bigger too.

For the experiments, we use a specific values for α. We have chosen this values because we believe they are rougher to the reality. We use: α_1 and α_2 equals 0.3 and α_3 equals 0.4.

5.1 Crisp model vs Fuzzy model

We were evaluating our system with three different fuzzy distribution: a trapezoidal distribution, a the Gauss bell shape

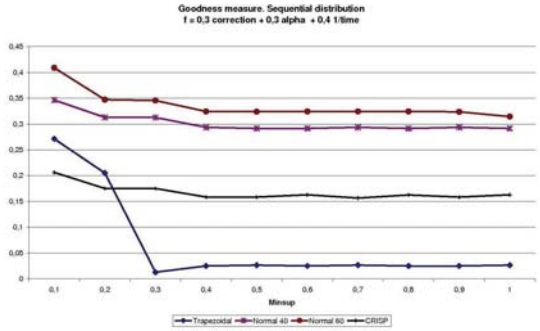


Figure 4: Crisp vs Fuzzy. Database a

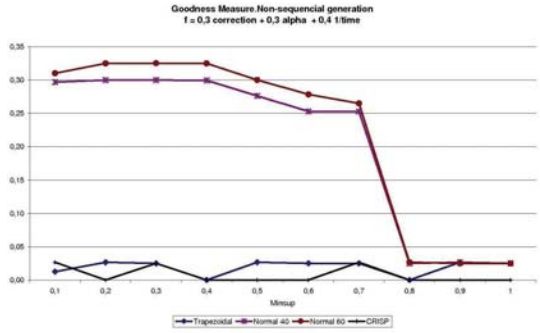


Figure 5: Crisp vs Fuzzy. Database b

with mean 40 minutes and a the Gauss bell shape with mean 60 minutes. We have used some distributions more, however we believe these distributions show the most interesting results. In figure 4 and 5, two graphics show goodness measure for every window. In addition, we show results of crisp model to compare the methods. If we observe results, we extract some conclusions which are convenient to study:

- In general, we conclude that the Fuzzy model obtains better results than the Crisp one. The Fuzzy method includes a way to provide importance to some actions that are key but are irregular. On the other hand, the Crisp method consider all actions at the same importance, and can lose this key actions. In addition, the Crisp method has to study all the data, while the Fuzzy method can reduce the size of transactions and can do the elements become frequent, minimizing the computational cost. This reduction could be very interesting when we have a forgetful user, because we would have a database with many mistakes that we would have to avoid. Moreover, this size reduction of the database makes it easier to find if the order of the sequences are valid.
- Among the Fuzzy Temporal Windows, the best results are obtained with the Gauss bell shape. It relaxes the degree of membership of their elements to the fuzzy window. When we perform the cut with a specific α, we include more items to study. Instead, in the trapezoidal shape the α-cuts are more abrupt, then more elements are lost.
- For the second database, the Crisp method provide good results: if we study all possible data, you always get

ISBN: 978-989-95079-6-8

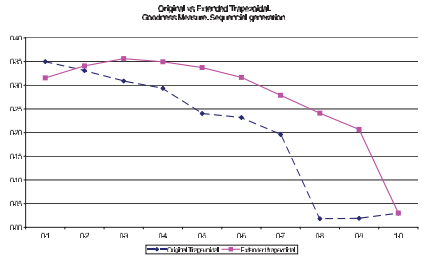


Figure 6: Original vs Extended Trapezoidal. Database a

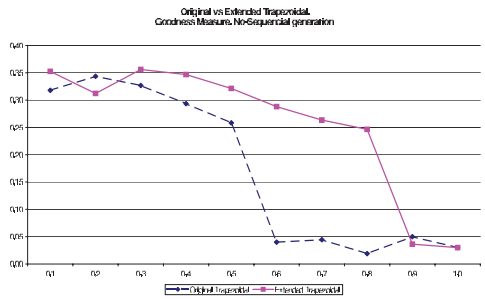


Figure 7: Original vs Extended Trapezoidal. Database b

key actions. However, with any Fuzzy Temporal Window with a Gauss bell shape we can extract almost the same results than the Crisp model minimizing the computational cost. Other alternative is to extend the variance of the fuzzy window, as we will see in the next subsection.

- Thus, the outcomes obtained with the fuzzy model are better than using crisp model.

5.2 *Analysis of process of extending the Fuzzy Temporal Window*

Once we have checked our fuzzy method, we analysed the process of extending the Fuzzy Temporal Window from a knowledge expressed as a quantified sentence. We are going to continue with example 3.4 that we proposed on section 3.3. We applied original Trapezoidal Temporal Window over two databases used in the previous section, and the new window obtained after expanding the window. Goodness measure is represented on figures 6 and 7: From these graphics, we observe that sometimes we extract results with the second window where the first window do not. This event happens because with the second window we have expanded the amount of data we studied to obtain the results. Although, always, we studied less data than the Crisp Method. Therefore, we adjusted the window to get better results than original window, without studying the whole interval.

6 Conclusions

In this paper, we have presented a method to detect user behaviour. As we have indicated, this problem is random imprecision by definition, because we do not know the interval of the time in which the action happen. Imprecision appears due of a behaviour is not static on a context or domain and is specific for every user. Thus, we have designed a method which uses fuzzy logic to limit events in a specific behaviour using a Fuzzy Temporal Window. With a Window, we can assign a membership degree for every event for the studied behaviour. In addition, we extend the Fuzzy Temporal Window concept, to adjust the interval using the knowledge from user we have. This adjustment is made with quantified sentences. With this adjustment we solve the possibility of a person changes his habits.

System's output is the fuzzy pattern sequences which define the behaviour. To manage non-random imprecision, we use the α-cuts and the Decomposition Theorem to obtain a fuzzy result applying the crisp method.

We have demonstrated that the fuzzy method is better than crisp method and that the fuzzy method obtains better outcomes than crisp one. As well as we conclude that results depend on the type of window we use. In addition, we have studied that outcomes are independent from the database.

In future work, we want to distinguish between three problems of our system:

- Non-random imprecision
- Uncertainty, because it could exists an error in the sensor reading. There are many reasons to provoke mistakes in sensors reading: the sensors or the reader could fail, the environment affects to the sensors, sensors damage, etc.
- Uncertainty in the Fuzzy Temporal Window, because the behaviour does no always happen exactly equal.

Our objective consists on extracting patterns that could control these kinds of problems, obtaining an adjusted pattern.

References

[1] M. Weiser. The computer for the twenty-first century. *Scientific American*, 1:94, 1994.

[2] Roy Want. Rfid: A key to automating everything. *Scientific American Magazine*, -:10, Enero 2004.

[3] Hiroyuki Yamahara et al. An individual behavioral pattern to provide ubiquitous service in intelligent space. volume 6, pages 562–569, 2007.

[4] T. Mori et al. Sensing room: Distributed sensor environment from measurement of human daily behaviour. *Intelligent Robots and Systems, 2005. (IROS 2005).IEEE/RSJ International Conference*, 2:1703 – 1709, 2-6 Aug. 2005.

[5] Kyohei Koyamaet al. Embedded action detector to enhance freedom from care. In *Proc. of 11th International Conference on Human-Computer Interaction*, page 8, Las Vegas, July 2005.

[6] Matthai Philipose et al. Inferring activities from interactions with objects. *IEEE Pervasive Computing*, 3(4):50–57, 2004.

[7] María Ros et al. Un sistema para identificacin de comportamientos correctos en un mundo marcado. Primeras Jornadas Cientficas sobre RFID, 21-23 Noviembre 2007.

[8] María Ros et al. Correct behavior identification system in a tagged world. *Expert systems with applications*, 36:9899–9906, 2009.

[9] L A Zadeh. A computational approach to fuzzy quantifiers in natural languages. *Computing and Mathematics with Applications*, (9):149–184, 1983.

[10] A. Gonzalez et al. Dealing with uncertainty and imprecision by means of fuzzy numbers. *International Journal of Approximate Reasoning*, 21:233–256(24), August 1999.

[11] Hung T. Nguyen and Elbert A. Walker. *A first Course in Fuzzy Logic*. Chapman & Hall/CRC Taoylr & Frances Group, 2007.

ISBN: 978-989-95079-6-8

Organizational Risk Assessment using Adaptive Neuro-Fuzzy Inference System

J.Jassbi [1] S.khanmohammadi [2]

1 Islamic Azad University, Science & Research Branch,
Poonak sq. Ashrafi Esfahani Blv., Hesarak, Tehran, Iran
2 Faculty of electrical Engineering, University of
Tabriz, Tabriz, Iran
Email: Jassbi@srbiau.ac.ir, khan@tabrizu.ac.ir

Abstract— *In this paper a fuzzy model based on Adaptive Neuro-Fuzzy Inference System (ANFIS) is introduced for calculating the level of risk in managerial problems. In this model, affecting factors on the level of risk are considered as inputs and the level of risk as the output. Using the fuzzy model, the risky condition changes smoothly in a fuzzy environment as it is the case in the real world; while in classic models, we may have some stepwise changes in the state of the system caused by an infinity small deviations in input parameters. The main advantage of the introduced model is that for continuous values of input factors, the counters of risk surface represent a more realistic behavior for different systems. The model is designed for a system of three inputs (probability, impact and ability to react) and one output (risky situation). The ability of using historical data as well as experts' knowledge and flexibility of adaptation to unusual risky situations are some benefits of the introduced model. This model which is originally used in strategic management system to analyze the external environment and the level of threats can also be used in contingency management for incidents (CMI) or as a tool for Comprehensive Emergency Management Program (CEMP). The strategic Risk of Roche is considered as the bench mark for comparing the difference between fuzzy and classic systems.*

Keywords— Risk assessment, Adaptive Neuro-Fuzzy Inference System, Risk Counters, Impact, Ability to react,

1 Introduction

Risk management has many applications in different fields such as financial management, industrial engineering, military and manufacturing systems as a Meta-disciplinary field. Companies and organizations have to monitor, scan and investigate the environmental factors and to update their strategic plans based on the level of recognized threats to select appropriate policies and strategies for reducing the risks. There are several steps for implementation of global risk management, starting with risk identification and calculation of its level for taking an appropriate decision. One of the most important sub-processes for this purpose is the methodology of calculation of risk and comparison of risky factors to determine the priorities. Considerable quantitative models have been introduced for this purpose in literature, where it is tried to calculate the level of the risk, which is simply defined as the rate of threat or future deficit of any system imposed by controllable or uncontrollable variables (Chavas, 2004; Doherty, 2000). Several factors such as probability of occurrence, impact, severity and ability to react are introduced as effecting factors on the risk. Then it is tried to find the mathematical relation between affecting factors and the value (level) of the risk (McNeil and Frey and Embrechts 2005; Li and Liao, 2007). The concept of risk is considerably wide. It can contain strategic, financial, operational or any other type of the risk. Based on literature review, the models which are mostly used in different fields of risk analysis can be classified to three types:

a) Data oriented models
b) Analytical models
c) Models based on judgment

Although it should be mentioned that this classification is not only related to Risk analysis, but in general it can be applied to system identification tasks.

a) Data oriented models:

In these models the structure is not needed to be known and the only important issue is the system's behavior. Some of the features of this type of models are: dependence to historical data, being behavioral, and storing experience and knowledge, based on the recognition of the input and output patterns of the system. Artificial intelligent models such as ANFIS and Neural Networks as well as statistical methods by concept of math average approach can be classified as this type of models. The more the number of the historical data is, the more reliable the models and their results would be. Figure 1 presents this type of model which is considered as a black box.

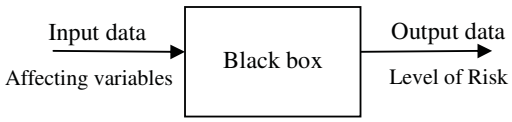


Figure. 1. Data-based System as a black box

This black box shows that there is no information on the analytical model of the system. It means that the relation among the variables as well as the way they affect each other is still unknown for user.

b) Analytical models.

In these systems it is possible to design a homomorphous model based on the system analysis process and structure of the problem. Being structure-based, analytical and independent from the data, are some of the features of this class of models. Many of the existing mathematical models,

ISBN: 978-989-95079-6-8

System Dynamics and a variety system analysis models are included in this class (Marrison2002). In this type of systems the structures must be recognized. It means that components and their relations must be identified to develop the analytical model. Figure 2 represents the concept of a structure based system. As it is mentioned the component and their relations must be recognized for developing an analytical model of the system.

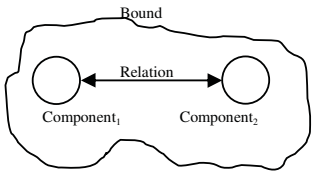


Figure. 2:. Structure-Based Systems (analysis)

For example if component$_1$ is Risk of investment when constructing a building, and component$_2$ represent the probability of earthquake in the region, the relation between them can be presented as :
Risk of investment=f (probability of earthquake)
Although it is a dream for researches to find an analytical model for such problems, however the major weakness of this kind of models is that they need precise data for modeling and the process of verification is considerably hard.

c) Models based on judgment

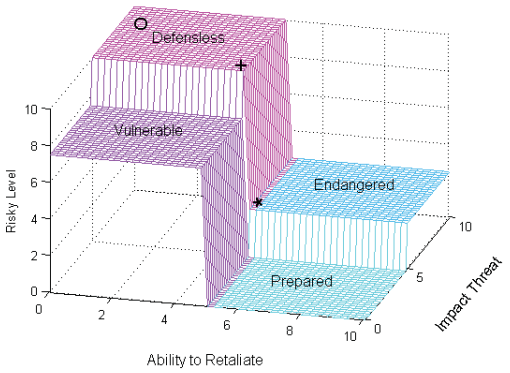
As it is mentioned, sometimes the first and second methods can not be used in real world problems. For example when the risk is based on the human behavior as a complex system, it is impossible for researches to find a pattern or mathematical model for it with available tools. So the judgment and knowledge of experts can help us to determine the level of risk and to find suitable solutions for risk management.

In this type of models, it is tried to overcome the major weakness of the first type consisting their uselessness in the case of no preexisting patterns. The growth of environmental dangers and the rapid increase of their variations as well as the increase of demand for such models by insurance companies have caused rapid and accelerated developments of such models (Roberets, 2005; Michael, 2004; Bernadell and CArdon and Coche and Diebold and Manganelli, 2004). This classification will help us to use the appropriate model in different situations.

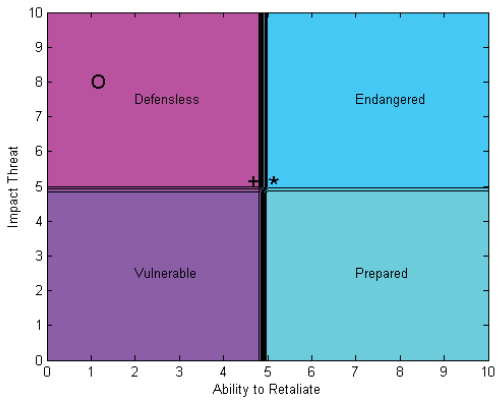
In this paper, the classic models of second type, commonly used in risk assessment of threats in strategic planning, are analyzed and criticized. Afterward a new model based on fuzzy inference is introduced for calculation of risk levels.

2. Problem Statement

Figure 3 shows a classic model of risk analysis. It consists of two factors: Impact threat and the ability to retaliate. In this model the risk value is classified in four groups. Each group represents a risky condition for the organization. During the implementation, first the opinions of the experts on the impact threat and the ability to retaliate are processed by means of any appropriate method such as group decision making, Delphi, …, and then the risky situation of organization is recognized. The distributions of the points with the same risk levels (contours of different levels) are also presented in Fig. 3(b), Points O and + represent the risky situation for two organizations with ability to retaliate and impact threats of (1, 8) and (4.9, 5.1) respectively.



(a)



(b)

Figure 3: (a) Risky situations classified in 4 levels, (b) counters

This model is very simple, but it has some structural drawbacks. For example the organization + which is in Endangered situation will change to completely opposite condition (Vulnerable) point(*) with infinity small deviations (A-da) in ability to retaliate. Also because of its geometrical structure, this model suffers from the lack of considering additional parameters such as ability to react.

Another method which has gained more attraction in the risk analysis literature is the model based on the linear combination of ability to retaliate and impact threat as:

$$Risk = (\text{ability to retaliate}) \times (\text{impact threat}) \quad (1)$$

Figure 4 represents a practical continuous increasing surface (levels), instead of stepwise levels for risk values. Two particular levels are shown by the cutting planes K1 and K2. Positions O, + and * are also presented in this figure. Figure 4(b) shows some contours of risky surface. As it is seen, in this model any small change in the values of probability and severity will cause a very small deviation in its risky level of organization. Also sometimes the risky level remains unchanged. This model is more realistic than the one

ISBN: 978-989-95079-6-8

presented by figure 3. However, it also has its limitations for real world applications because it simplifies the complicated relation between different factors that affect the risky conditions of organizations to a simple multiplication of two factors (ability to retaliate and impact threat).

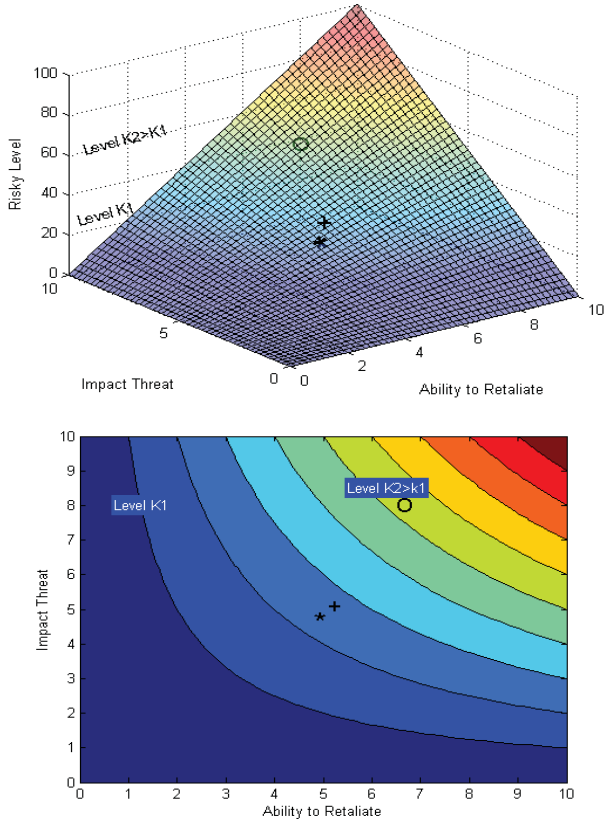


Figure. 4: Continuous surface for risk levels and counters

These kinds of models suffer from some weaknesses such as considering the relation between variables as a linear relation. Also when the number of inputs increases, the complexity of the system will increase dramatically while the accuracy of the model will decrease. Moreover due to inflexibility of the model, it can not be improved considering our experiences based on real world situations or the growth of experts' knowledge.

Hence we can formulate our problem as an input output system by:

$$R=F(X) \qquad (2)$$

Where X is the set of input variables which affect the level of the risk, R is the level of the risk and F(.) is a nonlinear function (Kreinovich and Nguyen, 2000).

The problem here is to find an appropriate solution by which the level of risk of the system (Organization) can be determined in complex situations when there is no access to all data, or the historical data is useless.

In this paper we try to develop a rule based model by using fuzzy inference system.

3. Fuzzy model

Fuzzy inference systems (FIS) are rule-based systems with concepts and operations associated with fuzzy set theory (Zadeh, 1965) and fuzzy logic (Ross, 2004; Mendel, 2001). These systems are mappings from an input space to an output state; therefore, they allow constructing structures that can be used to generate responses (outputs) to certain stimulations (inputs), based on stored knowledge on how the responses and stimulations are related. Sometimes this knowledge is obtained by eliciting information from specialists, in which case these systems are known as fuzzy expert systems (Takács, 2004). A fuzzy system also can be created to mach any set of input-output data. Adaptive Neuro-fuzzy Inference System (ANFIS) is one of the well known methods for creating Input-Output based models. (Krus , Gebhart and Palm, 1994). ANFIS only support Sugeno systems (Sugeno, 1985) subject to the structure of the system such as unit weight for each rule. Since its introduction, ANFIS has successfully been proved in many engineering applications (Jang, 1993). Another common denomination for FIS is fuzzy control systems (see for example (Mendel, 2001)).

FIS are usually divided in two categories (Mendel, 2001; Takagi and Sugeno, 1985): multiple input, multiple output (MIMO) systems, where the system returns several outputs based on the inputs it receives; and multiple input, single output (MISO) systems, where only one output is returned from multiple inputs. Since MIMO systems can be decomposed into a set of MISO systems working in parallel, all that follows will be exposed from a MISO point of view (Mamdani and Assilian, 1999).

FIS suffers of adjusting the linguistic knowledge of the expert with available data; so in this paper a fuzzy model based on ANFIS is introduced for calculating the risky situations of organizations by considering different factors such as probability, Impact threat and ability to retaliate (Nguene and Finger, 2007; Hyo and Hyun, and Yoon, 2002). This model is developed in three phases. To clarify the process of modeling in different phases a benchmark case study adopted from "strategic management by Rowe, et all" (Row et all, 1999) is used as a test example.

Phase 1: Data generation

The expert s' judgments may be extracted by means of different group decision making methods. Our experience in this work shows that the Delphi method as well as Nominal group is more practical. In case there are hard data as a result of historical behavior, a similar methodology can be used. By the way combination of hard data and information extracted from expertise can be used to develop first table which is necessary for next step.

Table 1 shows the set of generated data for test example.

Table.1

Probability	Impact	Ability to React	Risk (Vulnerability)
1	10	4	7.5
0	0	10	0
.5	5	5	1
1	5	5	3
.2	5	2	0
.8	2	8	5
.4	7	3	1

ISBN: 978-989-95079-6-8

1	0	5	0
.7	8	2	4
.8	8	7	2
.2	9	5	.5
.2	3	5	0
.7	10	3	4.5
.5	10	2	2.5
.5	10	10	0
1	8	0	10
1	2	8	.5
.6	10	4	3
.1	2	6	0
.3	6	8	0

Phase 2: Rule making

In this step the Adaptive Neuro-Fuzzy Inference System (ANFIS) is used for generating the rules.

Figure 5 Shows the surfaces of the rule base system adapted for the data of table 1 with:

numMFs = 25

mfType = gaussmf

epoch_n = 20

After constructing rule-base surface the model is improved by investigating the critical states (which are not approved by expert or when there is illogical behavior). Then by adding more rules or by imposing some minor deviations to data set the final rules may be obtained as shown in figure.5. By this way any specific behaviors cause by unusual relation between some particular inputs and output of the system can be supported which is significant advantage benefit of introduced model compare to other model such as Neural Network.

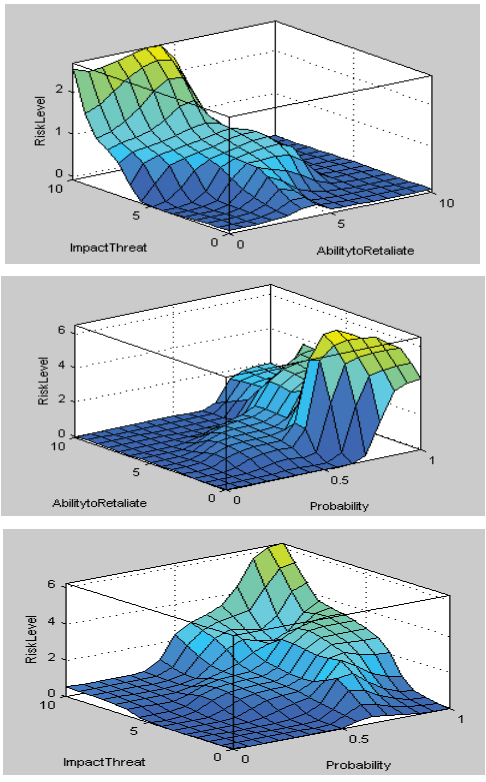
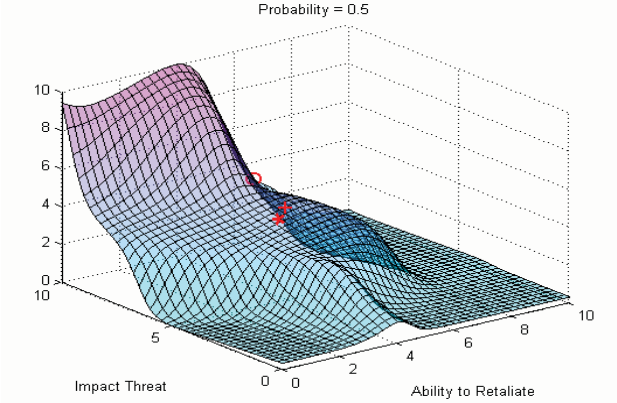
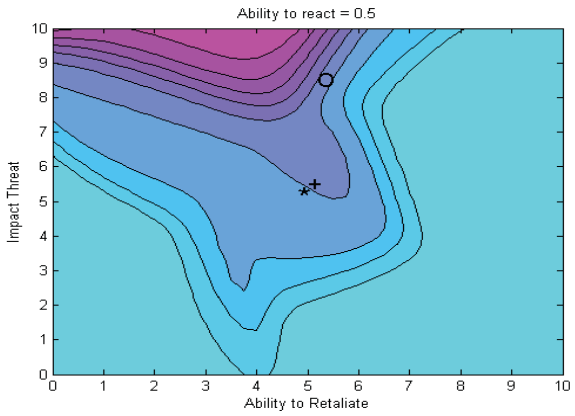


Figure.5: Surfaces of generated rules.

Phase 3. Implementation

This model is implemented to the simple example of section one to have an idea on the main difference between this method and the classic model. Figure 6 shows the surface and counters of risky levels of organizations +, and O. The results are shown for 50% of probability of impact to visualize the obtained results.





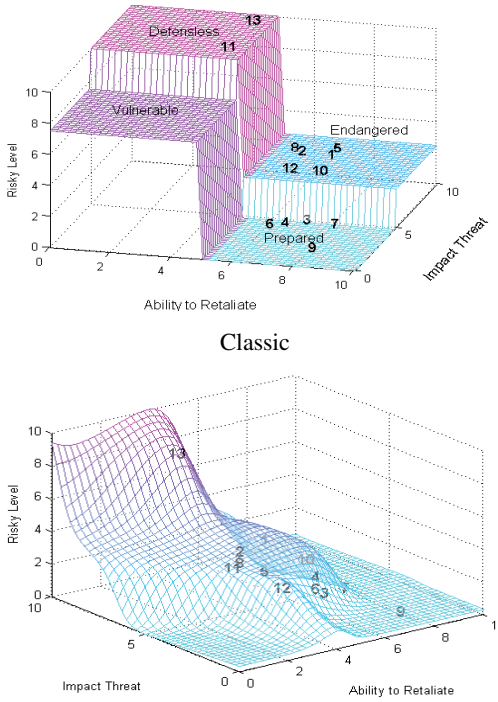
(b)

Figure. 6: (a) Risky surfaces, (b) counters for simple example

As it is seen, organization + which is appropriately Endangered situation will change to appropriately defenseless, point(*), with infinity small deviations (A-da) and (T-dt) in ability to retaliate and impact threat which is more realistic than the classic one.

The vulnerability analysis for Roche Company is also considered as a benchmark for implementing and comparing the obtained results with the classic methods. Figures 7 and 8 compare the results obtained by implementing the introduced model. Note that in the rest of paper only the surfaces and counters for probability of 50% are shown.

ISBN: 978-989-95079-6-8



Classic

Fuzzy

Figure.7: application of classic and fuzzy models for vulnerability analysis of Roche Company

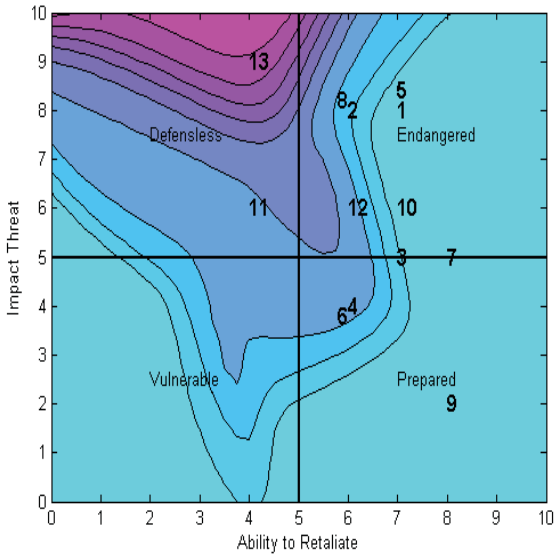


Figure 8: Risky levels of Roche by using classic and fuzzy models

Table 2 compares the results of this method by the classic one.

Table 2

Assumptions	Impact 0-10	Probability 0-1	Capability 0-10	Vulnerability Assessment Strategic management Approach	Vulnerability Assessment Classic	Vulnerability Assessment Fuzzy
1. Needs and wants served by products	8	0.7	7	Endangered	2.5	1.9960
2. Resources and Assets	8	0.6	6	Endangered	2.5	1.6119
3. Cost position relative to competition	5	0.4	7	Prepared	0	0.1490
4. Customer base	4	0.5	6	Prepared	0	1.9093
5. Technologies	8	0.4	7	Endangered	2.5	0.0446
6. Special skills	4	0.4	6	Prepared	0	1.1455
7. Corporate identity	5	0.5	8	Prepared	0	0.0298
8. Institutional barriers to competition	8	0.4	6	Endangered	2.5	0.8793
9. Social values	2	0.2	8	Prepared	0	0.0015
10. Sanctions, supports, and incentives	6	0.7	7	Endangered	2.5	1.7019
11. Customer goodwill	6	0.5	4	Defenseless	10	2.1834
12. Complementary products or services	6	0.3	6	Vulnerable	7.5	0.2943
13. Regulatory agencies	9	0.7	4	Defenseless	10	7.8129

It is seen that using the fuzzy method appropriate continues vulnerability levels are obtained to analyze the risky levels.

As it can be seen in Table.2 the Result of Vulnerability assessment by using Strategic management approach (as shown in fig.3) is stepwise and limited to 4 categories; hence the real value of risk is unclear. The second model (Classic Vulnerability Assessment) suffers from rigidity and can not be customized in different situations. So its efficiency can not be improved by considering available data and knowledge of experts. It is why in real world applications the decision Makers are not generally satisfied (trust the model) by the results obtained by this method.

ISBN: 978-989-95079-6-8

The introduced fuzzy based method overcomes the mentioned drawbacks. For example considering rows 1,2,5,8,10 in table.2, by using classical method as well as strategic management approach the same levels of risk are obtained for different assumptions where by using introduced method different level of risk is determined.

4. Conclusion

In this paper the fuzzy risk analysis procedure is compared with classic methods. It is demonstrated that the fuzzy analysis seems to be more reasonable and applicable because of its smoothness in calculating the risky situations. By fuzzy analysis the dynamic (time dependent) behavior of risk conditions can be defined and used for convenient and reliable decision makings. The model presented in this paper has some features as follows: a) Relations between variables in real life are nonlinear. Abstracting the situation and simplifying the problem to a linear model will cause the missing of some vital data where by utilizing the introduced model the relation between Risk and variables can be considered as a nonlinear function. b) Using FIS brings the advantage of developing and improving the model based on historical data. It is noted that the original model is constructed only based on the experts' knowledge due to lack of historical data. c) The model can be extended to be used for any number of inputs, where expanding the classic models to more inputs is not an easy task. d) This model considers probability as an input where other models usually solve the problem in a probabilistic environment. e) Any Specific behavior caused by unusual relation between some particular inputs and output of the system can be supported by imposing new rules to the model.

References

[1] Bernadell, C., CArdon, P., Coche, J., Diebold, F, X., Manganelli, S., *Risk management for central bank foreign reserves*. European Central Bank. 2004

[2] Carr, V., Tah, J, H, M., *A Fuzzy approach to construction project risk assessment and analysis*: construction project risk management system. Advances in Engineering Software, 32 (10 – 11), 847-857. 2001,

[3] Chavas, J, P., *Risk Analysis in Theory and Practice*. English, ISBN – 10: 0121706214. 2004,

[4] Doherty, N., *Integrated Risk Management*: Techniques and Strategies for Managing Corporate Risk, McGraw-Hill. 2000,

[5] Hall, L. O., Kandel, A., *The evolution from expert systems to fuzzy expert systems.* In Fuzzy Expert Systems, edited by A. Kandel, CRC press. 1991,

[6] Hyo, N, C., Hyun, H, C., Yoon, B, K., *A risk assessment methodology for incorporating uncertainties using fuzzy concepts.* Reliability Engineering & System Safety, 78, 173-183. 2002,

[7] Jang, J.S.R., *Adaptive-Network-based Fuzzy Inference System, IEEE Transactions on systems, Man, and Cybernetics* ,23(3) ,665-685,1993

[8] Kreinovich, V., Nguyen, H. T., *Fuzzy Systems are Universal Approximators for a Smooth Function and its Derivatives.* International Journal of Intelligent Systems, 15, 565-574. 2000,

[9] Kruse R. and Gebhardt J. and Palm R., Fuzzy Systems in Computer Science, Verlag Vieweg, P.O. Box 5829, D-65048 Wiesbaden, 1994"

[10] Li, Y., Liao, X,. *Decision support for risk analysis on dynamic alliance.* Decision Support Systems, 42, 2043-2059. 2007,

[11] Mamdani, E. H., Assilian, S., *An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller.* In. International Journal of Human-Computer Studies, 51 (2) pp. 135-147. 1999,

[12] Marrison C. , The Fundamentals of Risk Measurement, McGraw-Hill, 2002

[13] McNeil, A. J., Frey, R., Embrechts, P., *Quantitative Risk Management:* Concepts, Techniques, and Tools, Princeton University Press. 2005,

[14] Mendel, J., *Uncertain rule-based fuzzy inference systems:* Introduction and new directions. Prentice-Hall. 2001,

[15] Michael, P., *Systems modeling*; theory and practice, John Wiley & Sons, Ltd, England. 2004,

[16] Nguene, G, N., Finger, M., *A fuzzy – based approach for strategic choices in electric energy supply. The case of a swiss power provider on the eve of electricity market opening.* Engineering applications of Artificial Intelligence, 20, 37-48. 2007,

[17] Roberets, B. B., *The benefits of integrated quantitative risk management.* 12th annual international symposium council on system engineering, Melbourne. 2005,

[18] Ross, T. J., *Fuzzy logic with engineering applications.* John Wiley & Sons, Ltd. 2004,

[19] Rowe, Alan J., Mason, Richard O., Dickel, Karl E., Mann, Richard B., Mockler, Robert J. *Strategic Management: A Methodological Approach.* Addison N.Y. Page 204. 1999.

[20] Sugeno M., Industrial Applications of Fuzzy Control, Amsterdam, The Netherlands: North Holland, 1985

[21] Takács, M., *Critical Analysis of Various Known Methods for Approximate Reasoning in Fuzzy Logic Control.* 5th International Symposium of Hungarian Researchers on Computational Intelligence, Budapest, Hungary. 2004,

[22] Takagi, T., Sugeno, M., *Fuzzy identification of systems and its applications to modelling and control,* IEEE Trans. on Systems, Man and Cybernetics, 15:116-132,85. 1985,

[23] Zadeh L.A., Fuzzy Sets, Information and Control, vol.8, pp. 338--353, 1965"

ISBN: 978-989-95079-6-8

On the use of min-based revision under uncertain evidence for possibilistic classifiers

Salem Benferhat[1] Karim Tabia[1]

1.CRIL CNRS UMR8188
Artois University, France
Email: benferhat@cril.univ-artois.fr, tabia@cril.univ-artois.fr

Abstract— *Possibilitic networks, which are compact representations of possibility distributions, are powerful tools for representing and reasoning with uncertain and incomplete knowledge. According to the operator conditioning is based on, there are two possibilistic settings: quantitative and qualitative. This paper deals with qualitative possibilistic network classifiers under uncertain inputs. More precisely, we first present and analyze Jeffrey's rule for revising possibility distributions by uncertain observations in the qualitative setting. Then, we propose an efficient algorithm for revising possibility distributions encoded by naive possibilistic networks for classification purposes. This algorithm consists in a series of efficient and equivalent transformations of initial naive possibilistic classifiers.*

Keywords— Min-based possibilistic networks, classification under uncertain inputs

1 Introduction

Graphical models such as Bayesian networks [1][2], influence diagrams and possibilistic networks [3][4] are well-known formalisms widely used for representing and reasoning with uncertain and incomplete knowledge. Possibilistic networks are like Bayesian ones but lie on possibility theory [5][6] to handle imprecise and incomplete knowledge. They allow to factor a global joint possibility distribution into a set of local possibility distributions. There are two types of possibilistic networks: quantitative and qualitative. In the former, conditioning is based on the $product$ operator while it is based on the min-operator in the latter.

Classification is an important task in many real world applications consisting in predicting the class instance corresponding to an observation. This task is a special kind of inference: given an observed instance of each observable variable A_i, it is required to determine the class instance c_k for the instance to classify among a predefined set of class labels. It is important to note that possibilistic classifiers have not been sufficiently studied in spite of the fact that they are very appropriate for problems where knowledge is imprecise or missing as in real-time classification problems or problems where some inputs are missing or uncertain. These problems require classifiers under uncertain inputs. However, only few works used possibilistic classifiers [7][8][9][10] and to the best of our knowledge, there is only one preliminary work [11] addressing possibilistic network classifiers under uncertain inputs in the quantitative setting. More precisely, this work proposes an algorithm suitable of classification under uncertain inputs using product-based possibilistic networks.

This paper also addresses possibilistic-based classification with uncertain observations but in the qualitative setting. While quantitative possibilistic networks are quite similar to the Bayesian ones, the qualitative ones show significant differences. For instance, normalizing conditional possibility distributions in product-based possibilistic networks is the same as in the probabilistic setting while it is significantly different in min-based networks. We first recall the qualitative possibilistic counterpart of Jeffrey's rule [12] for revising possibility distributions by uncertain observations and show that this rule cannot be directly applied for revising possibilistic knowledge encoded by a possibilistic network. Indeed, classification by directly using the possibilistic counterpart of Jeffrey's rule is exponential in the number of attributes and attribute domains. Then, we proposed an efficient method for revising naive min-based possibilistic networks suitable for classification with uncertain inputs. This algorithm is based on a series of equivalent and polynomial transformations of initial possibilistic networks taking into account uncertain inputs.

The rest of this paper is organized as follows: Section 2 briefly presents basic background about possibility theory and possibilistic networks. In section 3, we address possibilistic belief revision based on the possibilistic counterpart of Jeffrey's rule. Section 4 proposes a new efficient algorithm for naive possibilistic network classification with uncertain inputs. Finally, section 5 concludes this paper.

2 Basic background on possibility theory and possibilistic networks

Let us first fix our notations: V=$\{A_1, A_2, .., A_n\}$ denotes the set of variables. D_A=$\{a_1, a_2, .., a_m\}$ denotes the finite domain of variable A. a_i denotes an instance (value) of variable A_i. A, X,.. denote subsets of variables from V. D_X=$\times_{A_i \in X} D_{A_i}$ represents the cartesian product relative to variables A_i involved in subset X. Ω=$\times_{A_i \in V} D_{A_i}$ denotes the universe of discourse and consists in the cartesian product of all variable domains involved in V. A tuple w=$(a_1, a_2, .., a_n)$ or w=$a_1 a_2 .. a_n$ which is an instance of Ω represents a possible state of the world. ϕ, φ denote subsets of Ω called events while $\overline{\phi}$ denotes the complementary of ϕ in Ω ($\overline{\phi}$=Ω-ϕ).

2.1 Possibility theory

Possibility theory was introduced by Zadeh [5] and developed by Dubois and Prade [6]. It is an uncertainty theory based on a pair of dual measures in order to evaluate knowledge/ignorance relative to event in hand. The concept of possibility distribution π is one of the important building blocks of possibility theory: It is a mapping from the universe of discourse Ω to the unit scale $[0, 1]$ which can be either quantitative or qualitative. In both these settings, a possibility degree $\pi(w_i)$ expresses to what extent it is consistent that w_i can be

ISBN: 978-989-95079-6-8

the actual state of the world. In particular, $\pi(w_i)$=1 means that w_i is totally possible and $\pi(w_i)$=0 denotes an impossible event. The relation $\pi(w_i)>\pi(w_j)$ means that w_i is more possible than w_j. A possibility distribution π is said to be normalized if $max_{w_i \in \Omega}(\pi(w_i))$=1. It is said to be sub-normalized otherwise.

The second important concept in possibility theory is the one of possibility measure denoted $\Pi(\phi)$ and computing the possibility degree relative to an event $\phi \subseteq \Omega$. It evaluates to what extent ϕ is consistent with the current knowledge encoded by possibility distribution π on Ω. It is defined as follows:

$$\Pi(\phi) = max_{w_i \in \phi}(\pi(w_i)). \quad (1)$$

The term $\Pi(\phi)$ denotes the possibility degree relative to having one of the events involved in ϕ as the actual state of the world.

The necessity measure is the dual of possibility measure and evaluates the certainty implied by the current knowledge of the world. Namely, $N(\phi)$=1-$\Pi(\overline{\phi})$ where $\overline{\phi}$ denotes the complementary of ϕ.

Given a possibility distribution π on Ω, marginal distributions π_X relative to subset of variables X ($X \subseteq V$) are computed using the max operator as follows:

$$\pi_X(x) = max_{w_i \in \Omega}(\pi(w_i) : w_i[X] = x), \quad (2)$$

where term $w_i[X] = x$ denotes the fact that x is the instantiation of X in w_i.

According to the interpretation underlying the possibilistic scale [0,1], there are two variants of possibility theory:

- **Qualitative possibility theory**: In this case, the possibility distribution is a mapping from the universe of discourse Ω to an "ordinal" scale where only the "ordering" of values is important.
- **Quantitative possibility theory**: In this case, the possibilistic scale [0,1] is numerical and possibility degrees are like numeric values that can be manipulated by arithmetic operators.

In this paper, we only focus on qualitative setting. Conditioning is a fundamental notion concerned with updating the current knowledge (encoded by a possibility distribution π) when an evidence (a sure event) is observed.

In the qualitative setting, conditional possibility degree of w_i given an event ϕ is computed as follows (we assume that $\Pi(\phi) \neq 0$) [13]:

$$\pi_m(w_i|\phi) \begin{cases} 1 & \text{if } \pi(w_i)=\Pi(\phi) \text{ and } w_i \in \phi; \\ \pi(w_i) & \text{if } \pi(w_i)< \Pi(\phi) \text{ and } w_i \in \phi; \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

2.2 Brief description of possibilistic networks

Like Bayesian networks, possibilistic ones [4] involves two components:

1. A graphical component consisting in a DAG[1] which encodes direct influence relationships existing between domain variables.

2. A numerical component which is a "numerical" component composed of a set of local a priori and conditional possibility distributions. The latter measure the influence endured by each domain variable A_i in the context of its parents U_{A_i}. Local possibility distributions should satisfy the normalization condition denoted as follows:

$$max_{a_{ij} \in D_{A_i}}(\pi(a_{ij}|U_{A_i})) = 1 \quad (4)$$

The joint possibility distribution encoded by the network is computed using the min-based chain rule. Namely,

$$\Pi(A_1, A_2, .., A_n) = min_{i=1..n}(\pi(A_i|U_{A_i})) \quad (5)$$

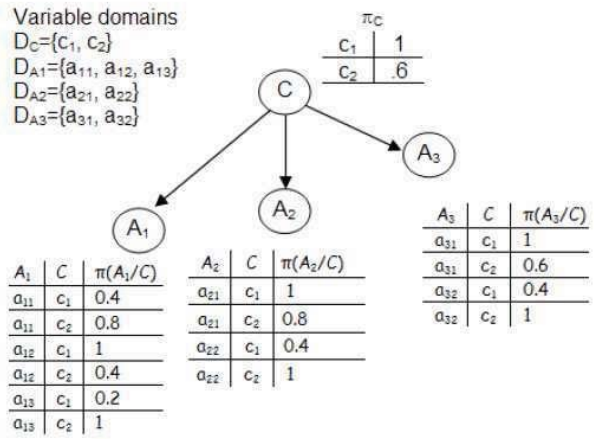
In classification problems, there exists one node associated with the class variable C which is not observable (it is the target variable) while the remaining nodes represent the attributes A_1, A_2,..,A_n that may be observable. Classification is ensured by computing the most plausible class instance given the instance to classify. Namely, given an observation denoted A=(a_1,a_2,..,a_n) of $\{A1, A_2, .., A_n\}$, the predicted class c is determined as follows:

$$c = argmax_{c_k \in D_C}(\Pi(c_k|A)) \quad (6)$$

Note that term $\Pi(c_k|A)$ denotes the possibility degree of having c_k the actual class instance given the observation A=(a_1,a_2,..,a_n). It is important to note that a class instance c_k is candidate for a given observation $a_1..a_n$ implies that $\Pi(c_k|a_1..a_n)$=1. Note also that several class instances may be totally possible for the observation to classify.

2.3 Min-based naive possibilistic network classifiers

A naive network classifier is the simplest form of possibilistic network classifiers. It lies on the strong independence assumption of attributes in the context of the parent node: attributes are assumed independent in the context of the class node. As it is shown in Figure 1, the only dependencies allowed in naive classifiers are from the class node C to each attribute A_i. As for the quantitative component of a naive possibilistic network, it involves the prior possibility distribution relative to the class node and the conditional possibility distributions relative to attributes given the class node. Figure 2 gives the joint possibility distribution encoded by network of Figure 1.



C	$\pi(C)$
c_1	1
c_2	.6

A_1	C	$\pi(A_1/C)$
a_{11}	c_1	0.4
a_{11}	c_2	0.8
a_{12}	c_1	1
a_{12}	c_2	0.4
a_{13}	c_1	0.2
a_{13}	c_2	1

A_2	C	$\pi(A_2/C)$
a_{21}	c_1	1
a_{21}	c_2	0.8
a_{22}	c_1	0.4
a_{22}	c_2	1

A_3	C	$\pi(A_3/C)$
a_{31}	c_1	1
a_{31}	c_2	0.6
a_{32}	c_1	0.4
a_{32}	c_2	1

Figure 1: Naive possibilistic network structure

Classification is ensured using min-based conditioning (see Equation 3) by computing the a posteriori possibility degree for each class instance c_k given the observation $A_1..A_n$ (namely $\Pi(c_k|A_1..A_n)$). Namely,

$$\Pi(c_k|A_1..A_n) = min(\pi(c_k, A_1..A_n), \Pi(A_1..A_n)) \quad (7)$$

[1] Direct Acyclic Graph

ISBN: 978-989-95079-6-8

A_1	A_2	A_3	C	$\pi(CA_1A_2A_3)$
a_{11}	a_{21}	a_{31}	c_1	0.4
a_{11}	a_{21}	a_{31}	c_2	0.6
a_{11}	a_{21}	a_{32}	c_1	0.4
a_{11}	a_{21}	a_{32}	c_2	0.6
a_{11}	a_{22}	a_{31}	c_1	0.4
a_{11}	a_{22}	a_{31}	c_2	0.6
a_{11}	a_{22}	a_{32}	c_1	0.4
a_{11}	a_{22}	a_{32}	c_2	0.6
a_{12}	a_{21}	a_{31}	c_1	1
a_{12}	a_{21}	a_{31}	c_2	0.4
a_{12}	a_{21}	a_{32}	c_1	0.4
a_{12}	a_{21}	a_{32}	c_2	0.4

A_1	A_2	A_3	C	$\pi(CA_1A_2A_3)$
a_{12}	a_{22}	a_{31}	c_1	0.4
a_{12}	a_{22}	a_{31}	c_2	0.4
a_{12}	a_{22}	a_{32}	c_1	0.4
a_{12}	a_{22}	a_{32}	c_2	0.4
a_{13}	a_{21}	a_{31}	c_1	0.2
a_{13}	a_{21}	a_{31}	c_2	0.6
a_{13}	a_{21}	a_{32}	c_1	0.2
a_{13}	a_{21}	a_{32}	c_2	0.6
a_{13}	a_{22}	a_{31}	c_1	0.2
a_{13}	a_{22}	a_{31}	c_2	0.6
a_{13}	a_{22}	a_{32}	c_1	0.2
a_{13}	a_{22}	a_{32}	c_2	0.6

Figure 2: Joint possibility distribution encoded by naive classifier of Figure 1

Note that according to min-based conditioning, Equation 7 is only valid when term $\pi(c_k, A_1..A_n) < \Pi(A_1..A_n)$ (See min-based conditioning of Equation 3). To the best of our knowledge, there is no work that addresses the problem of classification under uncertain inputs using min-based possibilistic networks.

3 Min-based revision for classification under uncertain inputs

In our context, uncertainty relative to uncertain/missing attribute A_i is represented by a possibility distribution π'_{A_i} given for example by the expert. Moreover, uncertainty can bear on any attribute subset or on the whole attribute set.
Jeffrey proposed in [12] a method based on probability kinematics for revising a probability distribution p into p' given uncertainty bearing on a set of mutually exclusive and exhaustive events λ_i. In this method, uncertainty is of the form (λ_i, α_i) with α_i=$p'(\lambda_i)$. Jeffrey's rule states that although there is uncertainty about events λ_i, conditional probability of any event $\phi \subseteq \Omega$ given any uncertain event λ_i. The possibilistic counterpart of this rule has been investigated in [14]. In the possibilistic framework, revised possibility distribution π' must comply with the principle stating that uncertainty about events λ_i must not alter the conditional possibility degree of any event $\phi \subseteq \Omega$ given any event λ_i. Namely,

$$\forall \lambda_i \in \Omega, \quad \forall \phi \subseteq \Omega, \Pi'(\phi|\lambda_i) = \Pi(\phi|\lambda_i) \tag{8}$$

Note that contrary to the probabilistic and quantitative possibilistic settings, Equation 8 may have several solutions. In this case, the least specific distribution is selected according to the principle stating that if an event is not explicitly discarded, then it must remain possible. Min-based conditioning allows to obtain from Equation 8 the following one:

$$\forall \phi, \Pi'(\phi) = max_{\lambda_i}(min(\pi'(\lambda_i, \phi), \pi(\lambda_i, \phi), \Pi(\phi)) \tag{9}$$

In classification with uncertain inputs problems, the possibilistic counterpart of Jeffrey's rule cannot directly be applied to revise the possibility distribution encoded by the classifier because of two major problems:
1- The first problem concerns the fact that Jeffrey's rule can be applied only if uncertainty bears on a set of exhaustive and mutually exclusive events while in classification with uncertain inputs problems, uncertainty is bearing on a set of attributes A_i which are not mutually exclusive. Indeed, if inputs A_1,..,A_n are uncertain, then this uncertainty is encoded by $\pi'(A_1)$,..,$\pi'(A_n)$ respectively. In order to apply Jeffrey's rule, we must compute a joint possibility distribution $\pi'(A_1..A_n)$ relative to $A_1..A_n$ where uncertain events $a_1..a_n$ are exhaustive and mutually exclusive. Obviously, the term $\pi'(A_1..A_n)$ is function of $\pi'(A_1)$,..,$\pi'(A_n)$ and will be henceforth denoted by $\pi'(A_1..A_n)$=$f(\pi'(A_1)$,..,$\pi'(A_n))$. Uncertain inputs $\pi'(A_1)$,..,$\pi'(A_n)$ can be combined using a combination operator according to the problem constraints and objectives. It is important to note that the algorithm we propose in next section works regardless the combination function f provided that this latter satisfies the following natural propriety (called henceforth unanimity):
i) If $\forall$i=1..n, $\pi'(a_i)$=1 then $\pi'(a_1..a_n)$=1.
ii) If $\exists a_i \in D_{A_i}$ such that $\pi'(a_i)$=α<1, then $\pi'(a_1..a_i..a_n)$<1 whatever are the possibility degrees of the other variable values.
One can easily check that the $product$ and min (and more generally t-$norm$ operators), which are the most used combination operators in the possibilistic framework satisfy this natural propriety. However, max and more generally, t-$conorm$ operators do not satisfy it. In this paper, we will use the min operator in order to induce the joint possibility distribution $\pi'(A_1..A_n)$ from uncertain inputs $\pi'(A_1)$,..,$\pi'(A_n)$. Namely,

$$\pi'(a_1..a_n) = min_{k=1..n}(\pi'(a_k)) \tag{10}$$

2- The second problem concerns the computational complexity of performing classification under uncertain inputs based on revising the possibility distribution encoded by the classifier by uncertain inputs using the possibilistic counterpart of Jeffrey's rule. Namely, given the initial possibility degrees of class instances c_k, we want to revise this possibility distribution given uncertain attributes A_1,..,A_n using Jeffrey's rule. More precisely, we need to compute $\pi'(c_k)$=$\pi(c_k|\pi'(A_1, .., A_n))$ defined as follows:

$$\pi'(c_k) = max_{A_1..A_n}(min(\Pi(c_k|A_1..A_n), \pi'(A_1..A_n))) \tag{11}$$

It is clear that Equation 11 cannot be used to compute $\pi'(c_k)$ because this computation is exponential in the number of attributes and attribute domains.

Example:
Let us illustrate possibilistic classification with uncertain inputs using Jeffrey's rule on the naive possibilistic classifier of Figure 1. The uncertain inputs to classify are provided by Figure 3. Classification is ensured by determining the most

π'_{A1}	
a_{11}	1
a_{12}	.4
a_{13}	1

π'_{A2}	
a_{21}	1
a_{22}	1

π'_{A3}	
a_{31}	1
a_{32}	.4

Figure 3: Uncertain inputs to classify

plausible class instance(s) given these uncertain inputs. Figure 4 gives the results of revising the initial joint possibility distribution (see Figure 2) using the possibilistic counterpart of Jeffrey's rule (see Equation 11). One can easily deduce from joint possibility distribution π' of Figure 4 that only class instance c_2 is totally plausible ($\pi'(c_2)$=1) given the uncertain observations of Figure 3. However, in order to obtain such a result, we computed the revised joint possibility distribution. We would like to ensure the same classification without computing the joint possibility distribution. Next section proposes

ISBN: 978-989-95079-6-8

A_1	A_2	A_3	C	$\pi'(CA_1A_2A_3)$
a_{11}	a_{21}	a_{31}	c_1	0.4
a_{11}	a_{21}	a_{31}	c_2	**1**
a_{11}	a_{21}	a_{32}	c_1	0.4
a_{11}	a_{21}	a_{32}	c_2	0.4
a_{11}	a_{22}	a_{31}	c_1	0.4
a_{11}	a_{22}	a_{31}	c_2	**1**
a_{11}	a_{22}	a_{32}	c_1	0.4
a_{11}	a_{22}	a_{32}	c_2	0.4
a_{12}	a_{21}	a_{31}	c_1	0.4
a_{12}	a_{21}	a_{31}	c_2	0.4
a_{12}	a_{21}	a_{32}	c_1	0.4
a_{12}	a_{21}	a_{32}	c_2	0.4

A_1	A_2	A_3	C	$\pi'(CA_1A_2A_3)$
a_{12}	a_{22}	a_{31}	c_1	0.4
a_{12}	a_{22}	a_{31}	c_2	0.4
a_{12}	a_{22}	a_{32}	c_1	0.4
a_{12}	a_{22}	a_{32}	c_2	0.4
a_{13}	a_{21}	a_{31}	c_1	0.2
a_{13}	a_{21}	a_{31}	c_2	**1**
a_{13}	a_{21}	a_{32}	c_1	0.2
a_{13}	a_{21}	a_{32}	c_2	0.4
a_{13}	a_{22}	a_{31}	c_1	0.2
a_{13}	a_{22}	a_{31}	c_2	**1**
a_{13}	a_{22}	a_{32}	c_1	0.2
a_{13}	a_{22}	a_{32}	c_2	0.4

Figure 4: Revised joint possibility distribution π'

an efficient algorithm suitable for naive min-based possibilistic network classification with uncertain inputs.

4 A polynomial algorithm for naive possibilistic classifier under uncertain inputs

Classification based on possibilistic networks is ensured by determining if each class instance $c_k \in D_C$ is totally possible given the inputs ($\Pi(c_k|A_1..A_n)$=1). This is the basic idea of our algorithm: only search for class instances having a posteriori possibility degrees equal to 1 on the basis of uncertain inputs. Our algorithm ensures this search through a series of equivalent transformations on the initial possibilistic network taking into account the inputs to classify. The five steps of our algorithm are detailed in the following.

4.1 Step 1: Eliminating not totally possible observations

In our context, classification is ensured by revising the possibility distribution encoded by the classifier according to the possibilistic counterpart of Jeffrey's rule (using Equation 11). Since we are only interested in determining if a given class label c_k has an a posteriori possibility degree equal to 1, it is possible to discard all instances of $A_1A_2..A_n$ where $\pi'(A_1A_2..A_n)$<1 because such instances force the value of $\pi'(c_k)$ to be less than 1. According to the *unanimity* propriety, each input $a_i \in D_{A_i}$ which is not totally possible ($\pi'(a_i)$<1) will prevent every input configuration $a_1..a_i..a_n$ from being totally possible. It is clear that such configurations are useless for the classification task and can be discarded. This leads to eliminating from each attribute domain $D_{A_i}^{\Pi G}$ (relative to attribute A_i in initial network ΠG) values whose possibility degrees in π'_{A_i} are less than 1. In this step, attribute domain $D_{A_i}^{\Pi G}$ is changed to $D_{A_i}^{\Pi G_{S_1}}$ which denotes A_i's domain in network ΠG^{S_1} obtained from ΠG after Step 1. Namely, $D_{A_i}^{\Pi G_{S_1}} = D_{A_i}^{\Pi G} - \{a_i, if\ \ \pi'_{A_i}(a_i)<1\}$. All the remaining attribute instances are totally possible after this step. Then we have the following proposition:

Proposition 1 Let ΠG be the naive possibilistic classifier encoding the initial knowledge and π'_{A_1}, π'_{A_2},..,π'_{A_n} be the possibility distributions encoding uncertainty relative to attributes A_1, A_2,.., A_n respectively. Let ΠG^{S_1} be the naive possibilistic network obtained by eliminating not totally possible instances. Then,
$\pi'^{\Pi G}(c_k) = max_{A_1..A_n}(\Pi^{\Pi G}(c_k|a_1..a_n) * \pi'(a_1..a_n))$=1
if and only if
$\pi^{\Pi G_{S_1}}(c_k) = max_{A_1..A_n}(\Pi^{\Pi G_{S_1}}(c_k|a_1..a_n)) = 1$

This proposition states that if there is a class instance c_k which is totally possible in the initial network ΠG given the uncertain inputs then it is also totally possible in network ΠG^{S_1} obtained after eliminating not totaly possible inputs.

Example (continued): Network of Figure 5 is obtained by respectively substituting $D_{A_1}^{\Pi G}$=$\{a_{11}, a_{12}, a_{13}\}$, $D_{A_2}^{\Pi G}$=$\{a_{21}, a_{22}\}$ and $D_{A_3}^{\Pi G}$=$\{a_{31}, a_{32}\}$ by $D_{A_1}^{\Pi G_{S_1}}$=$\{a_{11}, a_{13}\}$, $D_{A_2}^{\Pi G^{S_1}}$= $\{a_{21}, a_{22}\}$ and $D_{A_3}^{\Pi G^{S_1}}$=$\{a_{31}\}$.

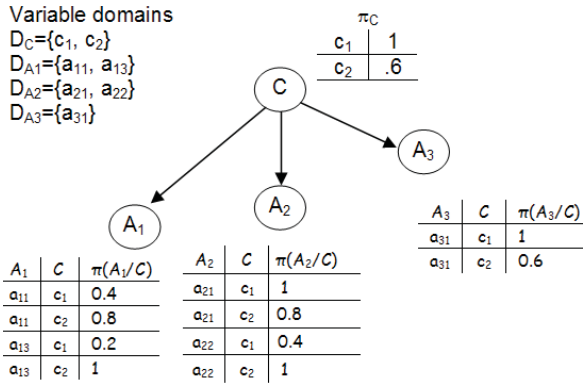


Figure 5: ΠG^{S_1}: Possibilistic network after Step 1

It is important to note that the possibility distribution relative to node C in ΠG^{S_1} is exactly the same as in ΠG.

4.2 Step 2: Eliminating unary variables

After the instance elimination step, some attributes may have their domains only containing one instance. In the example of Figure 5, attribute A_3's domain $D_{A_3}^{\Pi G_{S_1}}$ exactly contains one element (which is a_{31}). Such unary variables can be eliminated provided that class node distribution is adapted to keep the joint possibility distribution unchanged. The possibility degree of any instance $c_k a_1 a_2..a_n$ in network ΠG^{S_1} is computed using the min-based chain rule as follows:

$$\Pi^{\Pi G_{S_1}}(c_k a_1..a_n) = min(\pi^{\Pi G_{S_1}}(c_k), \pi^{\Pi G_{S_1}}(a_1|c_k), .., \pi^{\Pi G_{S_1}}(a_n|c_k)) \quad (12)$$

Unary attributes can be eliminated giving a new network ΠG^{S_2} where class node distribution has to be adjusted in order to guarantee that $\pi^{\Pi G_{S_1}}(c_k a_1..a_n)$ and $\pi^{\Pi G_{S_2}}(c_k a_1..a_n)$ remain equal. Hence, for an unary attribute A_i whose domain $D_{A_i}^{\Pi G_{S_1}}$ contains only one value a_i, it is possible to achieve this transformation by substituting each $\pi^{\Pi G_{S_1}}(c_k)$ with $\pi^{\Pi G_{S_2}}(c_k)$=min($\pi^{\Pi G_{S_1}}(c_k), \pi^{\Pi G_{S_1}}(a_i|c_k)$). This finding is formalized by the following proposition:

Proposition 2 Let ΠG^{S_1} be the naive possibilistic classifier whose nodes involve class node C and attribute nodes A_1, A_2,..,A_n. Assume that A_1 is a unary attribute whose domain only contains the instance a_1 and let ΠG^{S_2} be the naive possibilistic network involving C, A_2,..,A_n such that: $\pi^{\Pi G_{S_2}}(a_i)$= $\pi^{\Pi G_{S_1}}(a_i)$ for i=2,..,n and $\pi^{\Pi G_{S_2}}(c_k)$=min($\pi^{\Pi G_{S_1}}(a_1|c_k), \pi^{\Pi G_{S_1}}(c_k)$). Then,
$\forall c_k \in D_C, \forall a_i \in D_{A_i}^{\Pi G_{S_1}}$ for i=1..n,
$\pi^{\Pi G_{S_1}}(c_k a_1 a_2..a_n)$=$\pi^{\Pi G_{S_2}}(c_k a_2..a_n)$.

Example (continued): Let us continue our example. In network of Figure 5, attribute A_3 is unary and will be removed. After Step 2, we obtain network of Figure 6.

ISBN: 978-989-95079-6-8

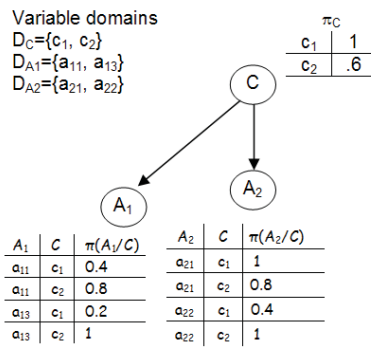


Figure 6: ΠG^{S_2}: Possibilistic network after Step 2

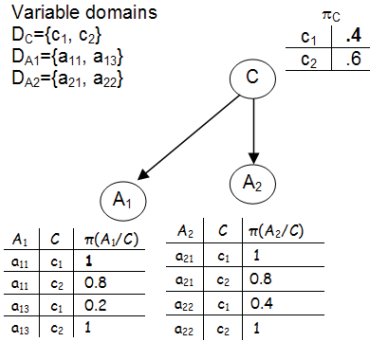


Figure 7: ΠG^{S_3}: Possibilistic network example after Step 3

4.3 Step 3: Re-normalizing local possibility distributions

As a consequence of the instance elimination step (Steps 1), local possibility distributions of some attributes might not be normalized. Indeed, it may exist a variable A_i and a class label c_k such that $max_{A_i}(\pi^{\Pi G_{S_2}}(a_i|c_k))=\alpha$ ($\alpha<1$). Step 3 deals with this problem considering two cases:

- If $\alpha=0$, then $\forall a_i \in D_{A_i}^{\Pi G_{S_2}}$, $\pi^{\Pi G_{S_2}}(c_k a_1..a_i..a_n)=0$ meaning that whatever is the value of A_i, this forces $\pi^{\Pi G_{S_2}}(c_k a_1 a_2..a_n)=0$. Hence, the class c_k cannot be among plausible ones, and can be removed from $D_C^{\Pi G_{S_2}}$.
- If $0<\alpha<1$, re-normalizing the conditional possibility distribution relative to attribute A_i can be done by building a new network ΠG^{S_3} by substituting $\pi^{\Pi G_{S_2}}(a_i|c_k)$ with

$$\begin{cases} 1 & \text{if } \pi(a_i|c_k)=max_{a_j \in D_{A_i}}(\pi^{\Pi G_{S_2}}(a_j|c_k)); \\ \pi(a_i|c_k) & \text{otherwise.} \end{cases}$$

and substituting $\pi^{\Pi G_{S_2}}(c_k)$ with $min(\pi^{\Pi G_{S_2}}(c_k), max_{a_j \in D_{A_i}}(\pi^{\Pi G_{S_2}}(a_j|c_k)))$. After this transformation, $\pi^{\Pi G_{S_3}}(c_k a_1..a_n)$ is conserved while sub-normalized local possibility distribution becomes normalized.

Proposition 3 Let ΠG^{S_2} be the naive possibilistic network involving nodes C,A_1,A_2,..,A_n obtained from Step 2. Assume that conditional possibility distribution relative to node A_1 is not normalized ($\exists c_k$ such that $max_{A_1}(\pi^{\Pi G_{S_2}}(a_1|c_k))=\alpha$ and $0<\alpha<1$). Let ΠG^{S_3} be the naive possibilistic network having same structure as ΠG^{S_2} where $\pi^{\Pi G_{S_3}}(a_1|c_k)=1$ if $\pi(wa_i|c_k)=max_{a_j \in D_{A_i}}(\pi^{\Pi G_{S_2}}(a_j|c_k))$, $\pi(a_i|c_k)$ otherwise and $\pi^{\Pi G_{S_3}}(a_i|c_k)=\pi^{\Pi G_{S_2}}(a_i|c_k)$ for i=2,..,n and $\pi^{\Pi G_{S_3}}(c_k)$=min$(\pi^{\Pi G_{S_2}}(c_k), max_{a_j \in D_{A_i}}(\pi^{\Pi G_{S_2}}(a_j|c_k)))$. Then, $\forall c_k \in D_C^{\Pi G_{S_3}}$, $\forall a_i \in D_{A_i}^{\Pi G_{S_3}}$, $\pi^{\Pi G_{S_3}}(c_k a_1 a_2..a_n)=\pi^{\Pi G_{S_2}}(c_k a_1 a_2..a_n)$

Note that contrary to product-based possibilistic networks, normalizing conditional possibility distribution of min-based networks is significantly different from normalizing local probability distributions of Bayesian networks.

Example (continued): In network ΠG^{S_2} of Figure 6, local possibility distribution of attribute A_1 is not normalized ($max_{a_i \in D_{A_1}}(\pi^{\Pi G_{S_2}}(a_i|c_1))$=.4). After re-normalizing ΠG^{S_2}, we obtain network of Figure 7. Possibility distribution relative to A_1 has been re-normalized and possibility distribution relative to class node C has been adjusted accordingly. One can easily check that joint possibility distributions encoded by networks of Figure 6 and Figure 7 are equal.

4.4 Step 4: Prior totally possible class lookup

The aim of Steps 1, 2 and 3 is simplifying and re-normalizing the initial network. Step 4 allows to search for class instances which are totally possible in network NP^{S_3}. Once the network is simplified and re-normalized, it is immediate that each class instance c_k having the utmost prior possibility degree in network NP^{S_3} (c_k=$argmax_{c_j \in D_C^{NP_{S_3}}}(\pi^{NP_{S_3}}(c_j))$, is totally possible given the uncertain inputs to classify. In particular, if $\pi^{NP_{S_3}}(c_k)$=1, then we can assert that there exists an attribute configuration $a_1..a_n$ allowing c_k to be totally plausible. This result is formalized in the following proposition:

Proposition 4 Let NP^{S_3} be the naive possibilistic network obtained after Steps 1, 2 and 3. Then, $\forall c_k \in D_C^{\Pi G_{S_3}}$ such that $\pi^{\Pi G_{S_3}}(c_k)$=1 or c_k=$argmax_{c_j \in D_C^{NP_{S_3}}}(\pi^{NP_{S_3}}(c_j))$, then $max_{A_1..A_n}(\pi^{\Pi G_{S_3}}(c_k|a_1..a_n))=1$

The attribute configuration guaranteeing that c_k is totally possible is the one where $\pi^{\Pi G_{S_3}}(a_i|c_k)$=1 for i=1..n.

Example (continued): Proposition 4 allows to assert that in network ΠG^{S_3} of Figure 7, class instance c_2 is totally possible since c_2=$argmax_{c_j \in D_C^{NP_{S_3}}}(\pi^{NP_{S_3}}(c_j))$. The attribute configuration allowing this result is $a_{13}a_{22}$. One can easily check that $\Pi^{\Pi G_{S_3}}(c_2|a_{13}a_{22})$=1.

4.5 Step 5: Conditionally totally possible class lookup

We assume in this step that the class variable C is a binary variable (namely, $D_C^{\Pi G_{S_3}}$={c_1, c_2}) and class label c_2 is totally possible ($\pi^{\Pi G_{S_3}}(c_2)>\pi^{\Pi G_{S_3}}(c_1)$).

In Step 4, only a subset of plausible class instances are found. Indeed, other class instances c_k having $\pi^{\Pi G_{S_3}}(c_k)=\alpha$ ($0<\alpha<1$) and $c_k \neq argmax_{c_j \in D_C^{NP_{S_3}}}(\pi^{NP_{S_3}}(c_j))$ can be totally possible given the uncertain inputs to classify. Namely, it may exist an attribute configuration $a_1 a_2..a_n$ such that $\Pi^{\Pi G_{S_3}}(c_k|a_1 a_2..a_n)$=1 even if $\pi^{\Pi G_{S_3}}(c_k)<1$ and $c_k \neq argmax_{c_j \in D_C^{NP_{S_3}}}(\pi^{NP_{S_3}}(c_j))$. Hence, there is a need to check if every class instance c_k which is not found totaly possible in Step 4 can be conditionally totally possible without exploring all configurations of $\Pi^{\Pi G_{S_3}}(c_k|a_1 a_2..a_n)$. If class instance c_1 is totally possible, then this implies

ISBN: 978-989-95079-6-8

that there exists an attribute configuration $a_1..a_n$ where $\Pi^{\Pi G_{S_3}}(c_1|a_1a_2..a_n)$=1. In order to find such a configuration, we can use the following decomposition:
$\Pi^{\Pi G_{S_3}}(c_1|a_1..a_n)$=1 if and only if
$\pi^{\Pi G_{S_3}}(c_1a_1..a_n)$=$\Pi^{\Pi G_{S_3}}(a_1..a_n)$. Recall that
$\Pi^{\Pi G_{S_3}}(a_1..a_n)$=$max(\pi^{\Pi G_{S_3}}(c_1a_1..a_n), \pi^{\Pi G_{S_3}}(c_2a_1..a_n))$
Hence, $\pi^{\Pi G_{S_3}}(c_1a_1..a_n)\geq\pi^{\Pi G_{S_3}}(c_2a_1..a_n)$. Then
$\frac{min(\pi(c_1),\pi(a_1|c_1),..,\pi(a_n|c_1)}{min(\pi(c_2),\pi(a_1|c_2),..,\pi(a_n|c_2)}\geq 1$.
Let us define $\pi_{c_1}(a_i/c_1)$=1 if
$\pi(a_i|c_1)\geq\pi(a_i|c_2)$ and $\pi_{c_1}(a_i/c_1)$=$\pi(a_i|c_2)$ otherwise. Then
$\frac{min(\pi(c_1),\pi(a_1|c_1),..,\pi(a_n|c_1)}{min(\pi(c_2),\pi(a_1|c_2),..,\pi(a_n|c_2)}\geq 1$ implies that

$$\pi(c_1) \leq min(\pi_{c_1}(a_1|c_1),..,\pi_{c_1}(a_n|c_1)). \qquad (13)$$

Decomposition of Equation 13 can be used to build a new network ΠG^{S_5} by transforming conditional possibility distributions relative to each attribute A_i in ΠG^{S_3} by substituting every term $\pi^{\Pi G_{S_3}}(a_i|c_1)$ by 1 if $\pi^{\pi G_{S_5}}(a_i|c_1)\geq\pi^{\Pi G_{S_3}}(a_i|c_2))$ and by $\pi^{\Pi G_{S_5}}(a_i|c_1)$ otherwise and discarding c_2 since it is known to be totally possible. After this transformation, some local conditional possibility distributions in network ΠG may not be normalized. Repeating Step 2 (Re-normalization) on sub-normalized local distributions in network ΠG^{S_5}) allows to re-normalize them. Once re-normalization accomplished, the new possibility distribution $\pi_C^{\Pi G_{S_5}}$ relative to class node shows whether class instance c_1 is totally possible. Then we have the following proposition:

Proposition 5 Let ΠG^{S_3} be the naive possibilistic network obtained after Steps 1, 2 and 3, and let $D_C^{\Pi G_{S_3}}$=$\{c_1, c_2\}$, $\pi^{\Pi G_{S_3}}(c_1)$<$\pi^{\Pi G_{S_3}}(c_2)$. Let ΠG be the possibilistic network having same structure as ΠG^{S_3} where $D_C^{\Pi G^{S_5}}$=$\{c_1\}$. Let for i=1..n $\pi^{\Pi G^{S_5}}(a_i|c_2)$=1 if $\pi^{\Pi G_{S_3}}(a_i|c_1)\geq\pi^{\Pi G_{S_3}}(a_i|c_2))$ and $\pi^{\Pi G^{S_5}}(a_i|c_2)$= $\pi^{\Pi G^{S_3}}(a_i|c_2)$ otherwise. Then, $\pi^{\Pi G^{S_5}}(c_1)$=1 if and only if there exists an attribute configuration $a_1..a_n$ such that $\Pi^{\Pi G_{S_3}}(c_1|a_1..a_n)$=1.

It is important to note that in comparison with the lookup for conditionally totally possible class instances in product-based networks (see [11]), this step shows significant differences in min-based networks.

Example (continued): Transformation of Step 5 on network ΠG^{S_3} and its re-normalization gives network ΠG^{S_5} of Figure 8 where network of the left side gives the network obtained after Step 5 before re-normalization while network of right side represents ΠG^{S_5} after re-normalization according to Step 3. After this transformation, we can assert that class instance c_1 is not totally possible ($\pi^{\Pi G^{S_5}}(c_1) < 1$). This result confirms the one obtained by directly applying the possibilistic counterpart of Jeffrey's rule (see Figure 4).

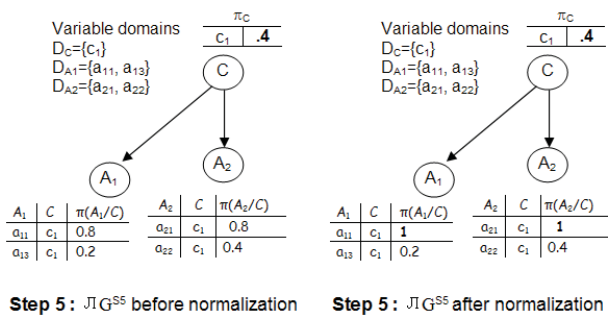


Figure 8: ΠG^{S_5}: Possibilistic network after Step 5

5 Conclusion

This paper dealt with min-based possibilistic network classifiers under uncertain inputs. It first addressed the min-based possibilistic counterpart of Jeffrey's rule for revising possibilistic knowledge encoded by a qualitative naive possibilistic network. After showing that classification based on the min-based revision of a possibility distribution encoded by a qualitative possibilistic classifier is exponential in the number of uncertain inputs, we proposed a polynomial algorithm for revising a naive min-based possibilistic network given uncertain inputs. This algorithm applies a series of equivalent and polynomial transformations on the initial network taking into account the uncertain evidence to classify. In future works, we will address general possibilistic classifiers under uncertain inputs.

References

[1] F. V. Jensen. *An Introduction to Bayesian Networks*. UCL press, London, 1996.

[2] A. Mittal. *Bayesian Network Technologies: Applications and Graphical Models*. IGI Publishing, USA, 2007.

[3] S. Benferhat, D. Dubois, L Garcia, and H. Prade. Possibilistic logic bases and possibilistic graphs. In *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence UAI 99*, pages 57–64, Stockholm, 1999.

[4] C. Borgelt and R. Kruse. *Graphical Models: Methods for Data Analysis and Mining*. John Wiley & Sons, Inc., USA, 2002.

[5] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1(3), 1978.

[6] D. Dubois and H. Prade. *Possibility theory*. Plenum Press, New-York, 1988.

[7] B.a Haouari, Z. Elouedi N. Ben-Amor, and K. Mellouli. Naive possibilistic network classifier. In *Proceedings of The Thirteenth Congress of International Association for Fuzzy-Set Management and Economy, SIGEF 2006*, 2006.

[8] T. Denoeux and L. M. Zouhal. Handling possibilistic labels in pattern classification using evidential reasoning. *Fuzzy Sets and Systems*, 122(3):409 – 424, 2001.

[9] H. Eyke. Possibilistic induction in decision-tree learning. In *ECML '02: Proceedings of the 13th European Conference on Machine Learning*, pages 173–184, London, UK, 2002.

[10] I. Jenhani, N. Benamor, and Z. Elouedi. Decision trees as possibilistic classifiers. *Int. J. Approx. Reasoning*, 48(3):784–807, 2008.

[11] S. Benferhat and K. Tabia. An efficient algorithm for naive possibilistic classifiers with uncertain inputs. In *The Second International Conference on Scalable Uncertainty Management(SUM'08)*, Naples, Italy, oct 2008. Springer LNAI.

[12] R. C. Jeffrey. *The Logic of Decision*. McGraw Hill, NY, 1965.

[13] E. Hisdal. Conditional possibilities independence and non interaction. *Fuzzy Sets and Systems*, pages 283–297, 1978.

[14] D. Dubois and H. Prade. A synthetic view of belief revision with uncertain inputs in the framework of possibility theory. *Int. J. of Approximate Reasoning*, 17(2–3):295–324, aot 1997.

ISBN: 978-989-95079-6-8

Parallel Coordinates-based Fuzzy Color Distance and its Applications

Constantin Vertan

Image Analysis and Processing Laboratory, Politehnica University of Bucharest
Bucharest, Romania
Email: cvertan@alpha.imag.pub.ro

Abstract— *Although simple and based on physical realizable color primaries, the RGB color space cannot allow the direct definition of a topology-preserving, perceptually-compatible inter-color distance. The classical normalized inter-color distances are rather complex to implement and tune and do not account for the natural incertitude and confusion regarding both the numerical color values and the color perception. We propose the use of a planar color representation plot, derived from multivariate data visualization, which allows the introduction of a luminance-invariant, geometry-based fuzzy inter-color distance, used with good results in the implementation of distance-based color filters and edge detectors.*

Keywords— fuzzy distance, fuzzy color, fuzzy image filtering.

1 Introduction

The classical inter-color distances are based on standard colorimetric representations, being metrics (or almost metrics) in the color gamut space (a subset of $\mathbf{R}^3$). We propose to investigate a new approach, inspired by the reduced ordering principle of Barnett [1] and also used in multivariate data visualization [2]: namely, the multivariate data (colors) are mapped to some familiar, two-dimensional objects, that can be grouped, compared, and plotted with more ease and are more suited for human perception. Examples of such mappings are the Chernoff faces [3], the Andrews curves [4] and their possible extensions [5], the basic and modified parallel coordinates [6] and the star glyphs [2]. Previous work showed that we could extend the geometric distance between simplified, plane color representations (such as the star glyphs) to color distances [7].

This work proposes the embedding of the uncertainty regarding the exact values of the color components describing the color into a fuzzy color distance, inspired by the two-dimensional geometrical representation of the color by means of the parallel coordinates.

The remainder of the paper is organized as follows: section 2 describes the basic parallel coordinates representation of multivariate data and its application to the representation of colors, section 3 introduces the distance between colors based on their parallel coordinates representation, section 4 introduces the fuzzy extension of the inter-color distance and, finally, section 5 presents some applications of the proposed fuzzy distance in non-linear color image filtering and color edge extraction.

2 Parallel coordinates: the basics

The parallel coordinates representation is a visualization technique that basically allows plotting n-dimensional points and patterns into the bi-dimensional plane. Thus, the parallel coordinates representation transforms multidimensional problems into two-dimensional patterns, without loss of information.

Let us consider a n-dimensional data point $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, originally represented in a Cartesian coordinate system. If each of the n axes of the coordinate system are lined up in parallel with all others into the plane, separated by a fixed distance Δ, we obtain the parallel coordinates system (as presented in Fig. 1) [6]. Obviously, the approach is scalable with respect to the data dimension, n. Visualization is facilitated by viewing the two-dimensional representation of the n-dimensional data points as lines, crossing the n parallel axes, each of them representing one dimension of the original feature space. A n-dimensional point $\mathbf{x}$ is represented in the plane by the series of $n-1$ connected line segments defined by the end points $[(kD, x_1), ((k+1)D, x_2)]$, $[((k+1)D, x_2), ((k+2)D, x_3)]$, , $[((k+n-2)D, x_{n-1}), ((k+n-1)D, x_n)]$ (as shown in Fig. 1).

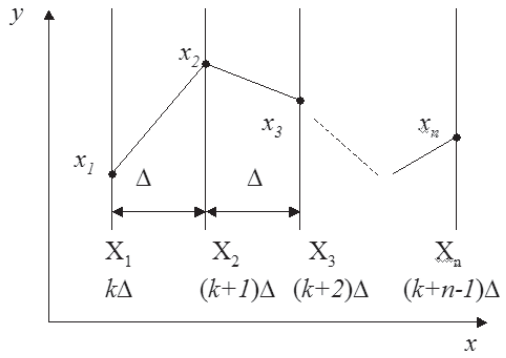


Figure 1: Basic parallel coordinates representation of a n-dimensional data point.

Several interesting mathematical results have been proven for the parallel coordinates representation [6], [8] and several applications were proposed, such as clustering [9], fuzzy rule representation [10], air traffic control [8], and color image filtering [11] based on the independent processing of the points along the individual coordinates axes.

We may notice the strong resemblance of the parallel coordinate representation with the Andrews curves [4], [2] representation method. The Andrews curve representation maps each n-dimensional point into a continuous, analytical curve, obtained as a n-term series expansion having the point coordinates as coefficients, according to a fixed functional basis. The original method uses the Fourier expansion (sine and cosines functions), but modifications were also proposed (e.g. the use of Haar wavelets for color image filtering [5]). For a discrete representation of the Andrews curves, the continuous basis functions are sampled at regular intervals, similar to the

ISBN: 978-989-95079-6-8

use of some parallel axes.

In the following of the paper, we will focus on the representation of colors - three-dimensional data points. Thus, each data point is a triple, $\mathbf{x} = (x_1, x_2, x_3)$. This representation is straightforward for the representation of RGB color data, due to the similar nature of the RGB components; thus, the data points will be $\mathbf{x} = (R, G, B)$, as proposed in [11]. Still, we propose to use a slight modification of the classical model, by using a four-dimensional data vector for the representation of the color, namely $\mathbf{x} = (R, G, B, R)$ as shown in Fig. 2. A color will be thus represented by the polygonal line or by the subgraph (lower limited by the horizontal axis) of that polygonal line, which is itself a polygon and will be subsequently named color polygon.

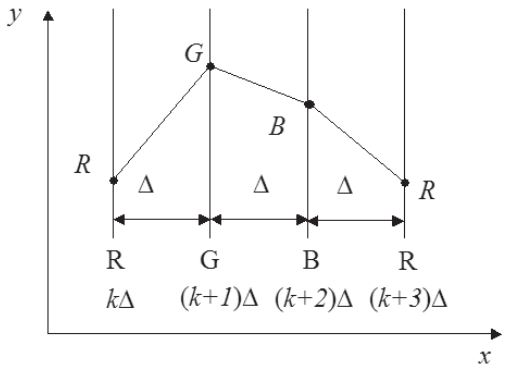


Figure 2: Proposed parallel coordinates representation of a RGB color, viewed as a 4-dimensional data vector.

3 Computing color distances

Color is represented by a three-component vector; the basic representation is the RGB model, for which the vector components are the relative amounts of normalized red, green and blue that additively mix in order to produce that color. Although simple and based on physical realizable color primaries, the RGB color space cannot allow the direct definition of a topology-preserving, perceptually-compatible inter-color distance. The same observation holds for the spectral color representations as well. As normalized by CIE (Commission Internationale de l'Eclairage), perceptual-compatible inter-color distances are obtainable from the Lab color representation. For any two colors $C_1 = (L_1, a_1, b_1)$ and $C_2 = (L_2, a_2, b_2)$, the simplest inter-color distance is the Euclidean distance in the Lab color gamut [12].

Advanced color difference formulas (such as the CMC, BFD or CIE94 [12]) have been introduced, since the basic Lab Euclidean metric do not accurately quantify small-to-medium-sized color differences. Such correctly-measured (as correlated to the subjective estimation) color difference are very complex and untractable for general-use in color image processing. Furthermore, they are not invariant to the change or luminance.

Obviously, the parallel coordinates polygon's shape is related to the nature of the color (properties like hue, saturation, colorfulness, etc.) and the overall area of the polygon relates to the luminance or brightness of the color. It can be accepted that a similarity measure between colors is given by the area of intersection of the two associated parallel coordinates polygons. In order to construct a symmetrical and normalized measure, we will follow an approach similar to the definition of the Canberra distance: the normalization is performed with respect to the average area of the two color polygons. Thus, the distance between colors C_1 and C_2, represented by their corresponding parallel coordinates polygons P_1 and P_2 can be introduced as:

$$d(C_1, C_2) = 2\frac{Area_{P_1 \nabla P_2}}{Area_{P_2} + Area_{P_2}} \tag{1}$$

From simple plane geometry considerations one can easily show that the distance in (1) can be approximately put in the form of:

$$d(C_1, C_2) = \left(D_{RG} + D_{GB} + D_{BR}\right)/3 \tag{2}$$

where D_{RG} is given in (3) and D_{GB} and D_{BR} have similar forms.

$$D_{RG} = 2\begin{cases} \frac{|R_1-R_2|+|G_1-G_2|}{R_1+R_2+G_1+G_2} & if \Delta R \Delta G \geq 0 \\ \frac{(R_1-R_2)^2+(G_1-G_2)^2}{(R_1+R_2)^2-(G_1+G_2)^2} & if \Delta R \Delta G < 0 \end{cases} \tag{3}$$

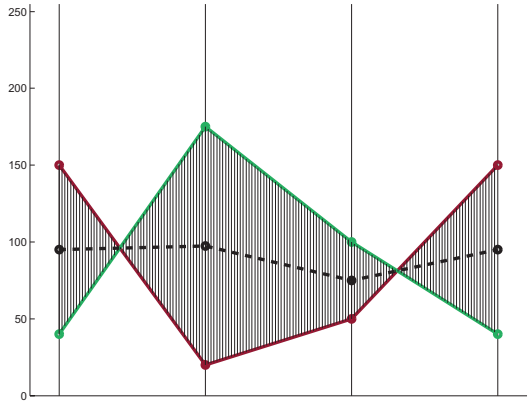


Figure 3: Proposed parallel coordinates distance for two colors (red with $RGB = (150, 25, 50)$ and green with $RGB = (40, 170, 110)$): the hashed area is the uncommon area of the parallel coordinates polygons associated to the colors; the dashed central line is the parallel coordinates polygon having the average area.

It can be easily shown that for unsaturated colors (grays), the proposed inter-color distance is the Canberra distance. The distance in (2) is normalized in the $[0; 1]$ range. It can be easily shown that the proposed distance is invariant to luminance changes, that is $d(C_1, C_2) = d(\alpha C_1, \alpha C_2), \forall \alpha \in \mathbf{R}$.

4 The fuzzy color distance

All image processing algorithms must deal with the imprecision and vagueness that naturally arise in the digital representation of visual information. Noise, quantization and sampling errors, the tolerance of the human visual system are the cause of this imprecision. This strongly suggests that fuzzy models may be used for taking them into account.

In the case of color images, color attributes and color differences play a particularly important role in the perception of

ISBN: 978-989-95079-6-8

object boundaries. The process of measuring color differences must be designed to maintain a balance between the computed and the perceived difference. Still, the simple use of any given color representation does not account for the similarity perception and the visual confusion of colors. We propose to deal with this factors in the framework of a fuzzy representation of the basic RGB color components and obtain a fuzzy distance between colors based on the parallel coordinate representation of the fuzzy colors.

The assumed model means that we will consider each of the R, G, and B color components as fuzzy numbers, centered at their ideal (correct, crisp) value. Thus, the polygonal line representing the given color in the parallel coordinates system will evolve into some fuzzy polygonal line, composed by fuzzy line segments.

We will define the fuzzy distance $\widetilde{D}$ between two fuzzy colors by means of its α-cuts. The corresponding α-cuts of any fuzzy color component are closed intervals along the corresponding parallel coordinate axes. Under these circumstances we can easily determine (from basic geometrical considerations) the minimal and maximal common areas of the corresponding color parallel coordinates polygon and thus to determine the lower and upper bounds of the corresponding α-cut of the fuzzy distance $\widetilde{D}$.

In the particular case of modelling the fuzzy RGB components by fuzzy triangular numbers, the limits of the α-cut intervals of the fuzzy distance are particulary simple to obtain. We will assume that the fuzzy color components $\widetilde{V}_i$ (where V_i is any of the RGB components) are defined by triangular fuzzy membership functions around their crisp value V_{i0}:

$$\mu_{\widetilde{V}_i}(V) = max(0, 1 - \frac{|V - V_{i0}|}{2 * \delta}) \tag{4}$$

In equation (4) above, δ is the support of the membership function, measuring its width around the crisp (correct) value and being thus the direct measure of the incertitude regarding the correct component value V_{i0}.

The α-cuts of any of the terms that sum to the color distance defined in (2) are bounded intervals, like $[D^-_{XY}(\alpha); D^-_{XY}(\alpha)]$. These upper and lower bounds of the α-cuts intervals of color distance terms are given by the expressions in (5) and (6) (for the D_{RG} term given in (3)), according to the cases when the line segments within two successive parallel coordinate axes intersect (corresponding to $\Delta R \Delta G < 0$) or not ($\Delta R \Delta G \geq 0$).

$$D^{\pm}_{RG}(\alpha) = 2\frac{|R_1 - R_2| + |G_1 - G_2| \pm 4(1-\alpha)\delta}{R_1 + R_2 + G_1 + G_2} \tag{5}$$

$$\begin{aligned} P_1 &= (R_1 + R_2 + G_1 + G_2 \mp 4(1-\alpha)\delta) \\ P_2 &= (|R_1 - R_2| + |G_1 - G_2| \mp 4(1-\alpha)\delta) \\ U_1 &= (R_1 - R_2 \pm 2(1-\alpha)\delta)^2 \\ U_2 &= (G_1 - G_2 \pm 2(1-\alpha)\delta)^2 \\ D^{\pm}_{RG}(\alpha) &= 2\frac{U_1 + U_2}{P_1 P_2} \end{aligned} \tag{6}$$

The fuzzy color distance is obtained by its α-cuts, by summing the α-cuts of the corresponding D_{XY} terms in (2). The choice of the width δ of the triangular membership function that models the RGB color components and the discrete nature of the component values imply that the number of distinct α-cuts of the color distance is limited to the rounded value of $\delta + 1$. For instance, the fuzzy distance between the two colors represented in parallel coordinates in figure 3 (red with $RGB = (150, 25, 50)$ and green with $RGB = (40, 170, 110)$) computed according to $\delta = 2.2$ is defined by three α-cuts: 2.53 for $\alpha = 1$, $[2.49; 2.58]$ for $\alpha \in [0.55; 1)$ and $[2.44; 2.62]$ for $\alpha \in [0.09; 0.55)$.

5 Applications

A significant part of nonlinear filters for color images are distance-based, i.e. they rely on the computation of inter-color distances between the colors selected by the filtering window. We shall prove the use of the proposed fuzzy inter-color distance in both smoothing filters (such as the vector median filter) and edge extraction filters, such as standard derivative-based filters.

5.1 Vector median filtering using the fuzzy inter-color distance

The classical VMF (Vector Median Filter) [13] defines the vector (or color) median as the vector characterized by a minimal aggregated distance with respect to all other vectors in the filtering neighborhood; this ordering of the color vectors is an instance of the reduced ordering principle [1]. Usually, the distance between color is computed as a L^k norm of the RGB color vectors. We will show here that the use of the proposed fuzzy inter-color distance provides good filtering results.

The fuzzy aggregate distance associate to each color vector within the filtering window is computed based on the α-cuts of the individual inter-color distances. Finally, the aggregated distances (which are also fuzzy numbers defined by their α-cuts) are ranked, and the minimal fuzzy aggregated distance corresponds to the median color.

The ranking of the fuzzy numbers (the fuzzy aggregated distances) is performed according to the classical total integral value. The total integral value of a fuzzy number $\widetilde{D}$ was introduced in [14] as:

$$I_T(\widetilde{D}) = \int_0^1 \mu^{-1}_{\widetilde{D}}(y)dy \tag{7}$$

According to the ranking proposed in [14], the smallest fuzzy number has the smallest total integral value (we may also notice that this ranking is an instance of Barnett's reduced ordering principle [1]).

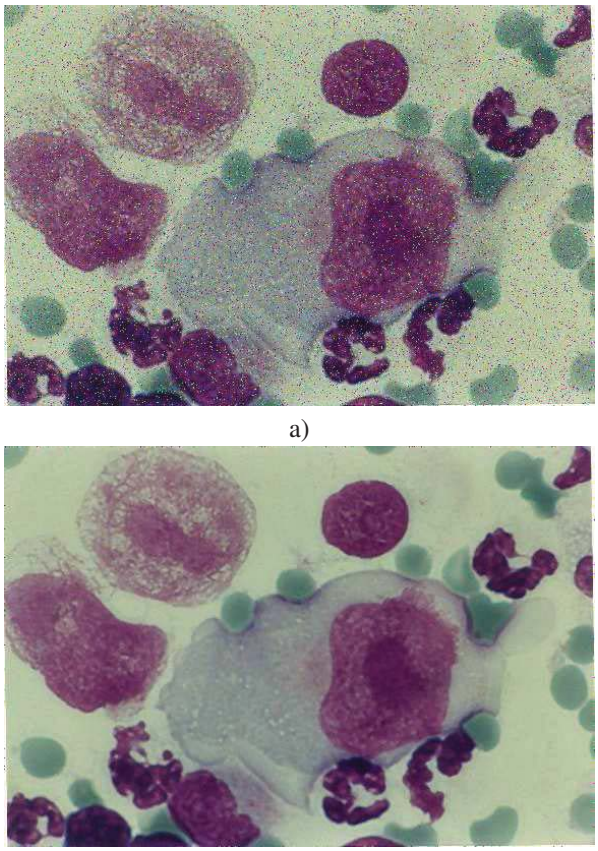
Figure 4 presents the result of a fuzzy-VMF filter applied for impulsive noise reduction in a color image.

5.2 Fuzzy color edge detection

Basically, all edge detection operators rely on the computation of an edge intensity map, which is further thresholded in order to obtain a binary edge map. The edge intensity map exhibits important values for the pixels that are on the boundaries of uniform regions, characterized by a discontinuity (or variation) of their colors.

One of the simplest contour extraction methods is the image Laplacian. We shall use the luminance-invariant fuzzy inter-color distance introduced in (2) for the implementation

ISBN: 978-989-95079-6-8



a) b)

Figure 4: a) 5% Impulsive noise degraded image and b) fuzzy-VMF filtered image within a 3×3 window, with a fuzzy color model characterized by $\delta = 2.2$.

of a fuzzy color Laplacian operator L. The proposed operator is a modification of the classical (derivative-type) V4-neighborhood Laplacian operator; the color Laplacian is the average inter-color distance within the color at the current processed location (i, j) and its immediate neighboring colors from the color image f. Mathematically we can express the proposed fuzzy Laplacian at location (i, j) as:

$$\widetilde{L}(i,j) = \frac{1}{4} \sum_{(k,l) \in V_4} \widetilde{D}(f(i+k, j+l), f(i,j)) \qquad (8)$$

The fuzzy Laplacian will be defined by the use of fuzzy color distances in (8). The summation of the fuzzy color distance is performed via the corresponding summation of the limits of their α-cut intervals, yielding an α-cut definition of the fuzzy color Laplacian at each image location.

The extraction of the binary edge map image implies the thresholding of the fuzzy Laplacian edge intensity map. According to the desired level of detail that we want to extract form the image, we may use as color edge intensity map either the lower or the upper limit of a specified α-cut interval, as shown in figures 5 and 6. The choice of a particular value of α acts as trimming parameter for the extraction of a more accurate and detailed or a more rough binary edge map, while keeping a same fixed general threshold. This is not possible when using a non-fuzzy color distance, either in RGB or Lab or other color representations.

6 Conclusions

In this contribution we presented a new fuzzy inter-color distance measure, derived from geometrical considerations related to a plane, two-dimensional, reversible color representation. This color representation by parallel coordinates (representing a color as a an open polygonal line) was primarily used in multivariate data representation. The proposed fuzzy inter-color distance can be used with good results in the median (non-linear) filtering of color images and fuzzy edge detection.

Acknowledgment

This work was supported in part by the Romanian National Agency for Scientific Research under the PNCDI2 51-021.2/2007 and 12-136/P2/2008 research grants.

References

[1] V. Barnett. The ordering of multivariate data. *J. of. Royal Stat. Soc. A*, 139(3):318–354, 1976.

[2] W. J. Krzanowski. *Principles of Multivariate Analysis: A User's Perspective*. Clarendon Press, Oxford, 1993.

[3] H. Chernoff. The use of faces to represent points in k- dimensional space graphically. *J. of the American Statistical Association*, 68:125–136, 1973.

[4] D. F. Andrews. Plots of high-dimensional data. *Biometrics*, 28:125–136, 1972.

[5] C. Vertan, M. Ciuc, and V. Buzuloiu. Multichannel median filtering by wavelet series expansion. In *Proc. of the 6th International Workshop on Systems, Signals and Image Processing IWSSIP 1999*, pages 77–80, Bratislava, Slovakia, 2-4 Jun. 1999.

[6] Inselberg A. and Dimsdale B. Multidimensional lines I: representation. *SIAM J. Appl. Math.*, 54(2):559–577, 1994.

[7] C. Vertan, M. Zamfir, E. Zaharescu, V. Buzuloiu, and C. Fernandez-Maloigne. Non-linear color image filtering by color to planare shape mapping. In *Proc. of ICIP 2003*, volume 1, pages 808–811, Barcelona, Spain, 10-13 Sept. 2003.

[8] Inselberg A. and Dimsdale B. Multidimensional lines II: proximity and applications. *SIAM J. Appl. Math.*, 54(2):578–596, 1994.

[9] S. Y. Chou, S. W. Lin, and C. S. Yeh. Cluster identification with parallel coordinates. *Pattern Recognition Letters*, 20:562–572, 1999.

[10] L. O. Hall and M. R. Berthold. Fuzzy parallel coordinates. In *Proc. of NAFIPS 2000*, pages 74–78, Atlanta GA, USA, 13-15 Jul. 2000.

[11] C. Vertan, M. Ciuc, and V. Buzuloiu. Parallel coordinates for color image processing. In *Proc. of International Conference on Signals, Circuits and Systems SCS 2001*, pages 469–472, Iaşi, România, 10-11 Jul. 2001.

[12] M. R. Luo. Colour science. In S. J. Sangwine and R. E. N. Horne, editors, *The Colour image processing handbook*, pages 26–66, London, UK, 1998. Chapman and Hall.

[13] J. Astola, P. Haavisto, and Y. Neuvo. Vector median filters. *Proc. IEEE*, 78(4):678–689, Apr. 1990.

[14] T. S. Lion and M. J. Wang. Ranking fuzzy numbers with integral value. *Fuzzy Sets and Systems*, 50:247–255, 1992.

ISBN: 978-989-95079-6-8

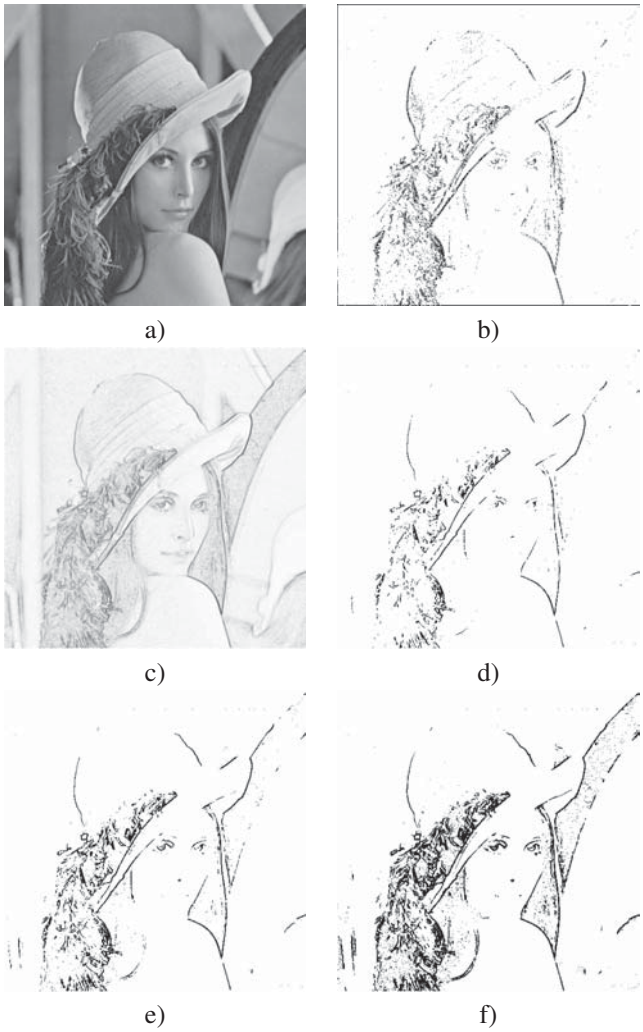


Figure 5: Fuzzy Laplacian edge extraction according to the proposed fuzzy color distance: a) Original color image; b) Classical Laplacian binary edge map; c) Fuzzy color Laplacian edge strength map computed according to (8) and (2) for $\delta = 3.2$ corresponding to an α-cut at α=1; d) Fuzzy color Laplacian binary edge map computed from c) according to an α-cut at $\alpha = 0.66$ and thresholding of the lower bound of the α-cuts interval; e) Fuzzy color Laplacian binary edge map computed from c) according to an α-cut at $\alpha = 1$; f) Fuzzy color Laplacian binary edge map computed from c) according to an α-cut at $\alpha = 0.66$ and thresholding of the upper bound of the α-cut interval. The fuzzy color model is defined by $\delta = 3.2$.

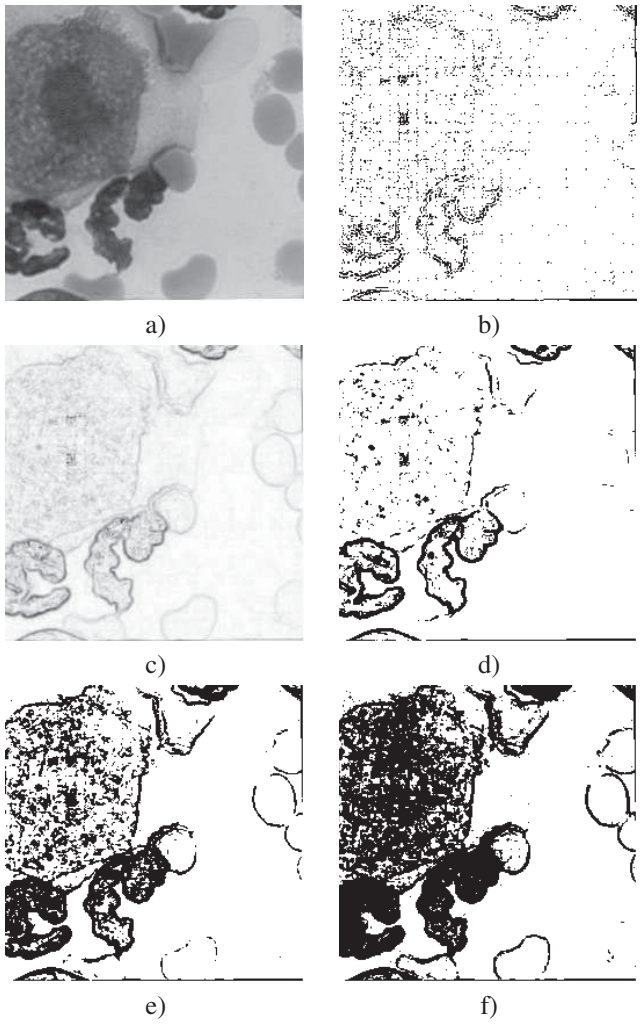


Figure 6: Fuzzy Laplacian edge extraction according to the proposed fuzzy color distance: a) Original color image; b) Classical Laplacian binary edge map; c) Fuzzy color Laplacian edge strength map computed according to (8) and (2) for $\delta = 3.2$ corresponding to an α-cut at α=1; d) Fuzzy color Laplacian binary edge map computed from c) according to an α-cut at $\alpha = 0.66$ and thresholding of the lower bound of the α-cuts interval; e) Fuzzy color Laplacian binary edge map computed from c) according to an α-cut at $\alpha = 1$; f) Fuzzy color Laplacian binary edge map computed from c) according to an α-cut at $\alpha = 0.66$ and thresholding of the upper bound of the α-cut interval. The fuzzy color model is defined by $\delta = 3.2$.

ISBN: 978-989-95079-6-8

Definition of fuzzy Pareto-optimality by using possibility theory

Ricardo C. Silva[1] Akebo Yamakami[1]

1. Department of Telematics, School of Electrical and Computer Engineering,
University of Campinas, P.O.Box 6101, 13083-970, Campinas-SP, Brazil.
Email: {rcoelhos,akebo}@dt.fee.unicamp.br

Abstract— *Pareto-optimality conditions are crucial when dealing with classic multi-objective optimization problems because we need to find out a set of optimal solutions rather than only one optimal solution to optimization problem with a single objective. Extensions of these conditions to the fuzzy domain have been discussed and addressed in recent literature. This work presents a novel approach based on the use of possibility theory as a comparison index to define a fuzzily ordered set with a view to generating the necessary conditions for the Pareto-optimality of candidate solutions in the fuzzy domain. Making use of the conditions generated, one can characterize fuzzy efficient solutions by means of carefully chosen single-objective problems. The uncertainties are inserted into the formulation of the studied fuzzy multi-objective optimization problem by means of fuzzy coefficients in the objective function. Some numerical examples are analytically solved to illustrate the efficiency of the proposed approach.*

Keywords— Possibility theory, multi-objective optimization, fuzzy Pareto-optimality conditions, fuzzy mathematical programming.

1 Introduction

One of the most significant characteristics of human beings is the decision making of day-by-day problems. This characteristics is used to solve several practical problems including economic, environmental, social and technical. These problems are multidimensional and have multiple objectives that are often non-commensurable and conflict with each other. Thus, they are inserted in the set of problems that are solved by using the theory of multi-objective optimization which is a generalization of traditional single objective optimization. Although multi-objective optimization problems differ from single objective problems only in the plurality of objective functions, it is important to realize that the notion of optimality condition change because now the decision maker must find solutions that satisfy or create a compromise among the multiple objectives. These solutions are called Pareto optimal or efficient or non-dominated solutions.

Optimization is a procedure of finding and comparing feasible solutions until no better solution can be found. These solutions are defined good or bad in terms of one or several objectives when is used any optimization model. The optimization models often use classical mathematical programming, which attempts to develop an exact model to the optimization problem of interest. Such modeling may overlook ambiguities that all too frequently exist in actual optimization operations. In recent years, Fuzzy Logic [16] has showed great potential for modeling systems which are non-linear, complex, ill-defined and not well understood. Fuzzy Logic has found numerous applications due to its ease of implementation, flexibility, tolerant nature to imprecise data, and ability to model non-linear behavior of arbitrary complexity because of its basis in terms of natural language. In [18] is discussed the use of fuzzy logic that is a precise logic of imprecision and approximate reasoning. We refer to [6, 10, 19], for some applications in the fields of pattern recognition, data analysis, optimal control, economics and operational research, among others.

The representation and arithmetic manipulation of uncertain numerical quantities can be defined by means of fuzzy sets. Unfortunately, the comparison among two or more fuzzy numbers, intervals and/or sets is not easy. Some approaches to compare them (see some examples in [2, 6, 9, 10, 14]) were developed, each one being based upon a different point of view. The possibility theory, which is analogous to the probability theory, was proposed by Zadeh [17] to aggregate the concept of a possibility distribution to the theory of fuzzy sets. Comparison indexes to rank fuzzy numbers and intervals employing possibility theory were proposed in [7]. Their importance stems from the fact that much of the information on which human decisions rely upon have a possibilistic, rather than a probabilistic nature. On the other hand, some works describe a fuzzy optimization problem in a classical problem and they use the classical theory to find the Pareto optimal set. [8] transform a fuzzy single objective problem into a classical multiple objective one where the number of objectives is defined by fuzzy coefficients from the fuzzy problem.

Possibility theory emerged from the notion of fuzzy sets and his concept tries to take account of the fact that an object may more or less correspond to a certain category in which one attempts to place it. In the calculus of degree of possibility emphasizes the double relationship between possibility theory and set theory and the concept of measure, respectively. One merit of possibility theory is to represent imprecision and to quantify uncertainty at the same time.

This work is organized as follows. Section 2 presents an overview about the formulation of classical multi-objective programming problems and classical concepts to obtain the set of Pareto optimal solutions. Also, it is shown an extension of these concepts to Pareto optimal solutions of fuzzy multi-objective programming problems. Section 3 introduces a novel approach to fuzzy Pareto-optimality. A fuzzily ordered set is defined by using a possibility distribution function as a comparison measure. This section also presents the characterization of the efficient solutions through the use of well defined scalar problems. This relation between efficient solutions and scalar problems can be determined by certain theorems. To clarify the above developments, two numerical

ISBN: 978-989-95079-6-8

examples are analyzed in section 4. Finally, conclusions are presented in Section 5.

2 Multiobjective programming problem formulation and concepts

Choosing the goal to be optimized is a critical step in the process of modeling real-world problems. The local or global optimal solution depends totally upon this choice. In the vast majority of real-world problems, various objective functions could be defined. Many times these functions are conflicting and/or non-measurable. Multi-objective optimization is the branch of mathematical optimization theory devoted to developing methods to solve problems with various objective functions. A classical multi-objective problem can be formulated as follows:

$$\begin{array}{ll} \min & F(\mathbf{x}) \\ \text{s.t.} & \mathbf{x} \in \Omega \end{array} \tag{1}$$

where $F = (f_1, f_2, \ldots, f_m),\ (m \geq 2)$ is a vector of objectives and $\Omega \subset \mathbb{R}^n$ is the set of feasible solutions.

Due to the issue of conflicting objectives, the classical concept of optimality that are used to obtain the optimal solution to optimization problems with an objective function does not fit into the multi-objective framework. Hence, in such a framework, one settles for the so-called efficient, non-dominated or Pareto-optimal solution.

The classical works [12] and [13], by Vilfredo Pareto, introduced the concept of Pareto-optimality and started the field of multi-objective optimization. A solution $x^* \in \Omega$ is said to be *Pareto-optimal* or *non-dominated* if there exists no alternative solution in Ω that improves some of the objective functions without degrading at least another objective function. Then, we can define mathematically a non-dominated solution as

Definition 1 (Pareto optimal solution) $\mathbf{x}^* \in \Omega$ *is said to be a* non-dominated solution *of Problem (1) if there exists no other feasible* $\mathbf{x} \in \Omega$ *such that* $f_i(\mathbf{x}) \leq f_i(\mathbf{x}^*),\ \forall\ i = 1, \ldots, m$ *with strict inequality for at least one* i.

However, when non-linear programming problems with single objective are solved by any non-linear programming method, only local optimal solutions are guaranteed in practical. Then, the concept of local non-dominated solution can be defined in the following way:

Definition 2 (local Pareto optimal solution) $\mathbf{x}^* \in \Omega$ *is said to be a* local non-dominated solution *of Problem (1) if and only if there exists a real number* $\delta > 0$ *such that* $\mathbf{x}^*$ *is non-dominated in* $\Omega \bigcap \mathcal{N}(\mathbf{x}^*, \delta)$*, i.e., there does not exist another feasible* $\mathbf{x} \in \Omega \bigcap \mathcal{N}(\mathbf{x}^*, \delta)$ *such that* $f_i(\mathbf{x}) \leq f_i(\mathbf{x}^*),\ \forall\ i = 1, \ldots, m$ *with strict inequality for at least one* i.

where $\mathcal{N}(\mathbf{x}^*, \delta)$ denotes the δ neighbourhood of $\mathbf{x}^*$ defined by $\{\mathbf{x} \in \mathbb{R} \| \|\mathbf{x} - \mathbf{x}^*\| < \delta\}$. Thus, it is possible to see by these definitions that the solution of a multi-objective programming problem consists of an infinite number of points.

Many methods to solve multi-objective programming problems were proposed and some specific methods can be found in [4]. These methods are classified according to the instant the decision maker applies their criteria. Three methods are proposed: (i) *a-Priori method*, where the decision maker assigns weights to the objective functions a-priori, thus obtaining a single mono-objective criterion; (ii) *a-Posteriori method*, where the decision maker strives to create some type of efficient solutions to chose from a-posteriori; and (iii) *interactive methods*, where the decision maker informs their preferences during the search process of an efficient solution.

Although the mathematical formulation of optimization problems with multiple objectives be well defined, some values of the real-world problems have vagueness, imprecision and uncertainty. These values which have been estimated by decision maker are parameters in the set of constraints and in one or several objective functions. These uncertainties can be formulated by logic fuzzy which is a way to describe this vagueness mathematically and it has found numerous and different applications due to its easy implementation, flexibility, tolerant nature to imprecise data, low cost implementations and ability to model non-linear behavior of arbitrary complexity because of its basis in terms of natural language. The concept of fuzzy decision to obtain a solution to fuzzy programming problems was introduced by Bellman and Zadeh[1] which proved that the fuzzy programming problems can be reduced to a conventional programming problem under some assumptions determined by decision maker.

Based on this work, many researches developed methods that solve optimization problems under fuzzy environment and in [15] is used α-cut sets to define a non-dominated solution to multi-objective programming problems with fuzzy parameters which is called α-Pareto optimal solution. In this case, the fuzzy parameters can be inserted in Problem (1) and it is transformed in the following way:

$$\begin{array}{ll} \min & F(\tilde{\mathbf{a}}; \mathbf{x}) = (f_1(\tilde{\mathbf{a}}_1; \mathbf{x}), \ldots, f_m(\tilde{\mathbf{a}}_m; \mathbf{x})) \\ \text{s.t.} & \mathbf{x} \in X(\tilde{\mathbf{b}}) \triangleq \{x \in \mathbb{R}^n | g_i(\tilde{\mathbf{b}}_i; \mathbf{x}) \leq 0, i = 1, \ldots, p\} \end{array} \tag{2}$$

where $\tilde{\mathbf{a}}_i$ and $\tilde{\mathbf{b}}_i$ represent a vector of fuzzy parameters involved in the objective functions and in the functions that form the set of constraints, respectively. These fuzzy parameters which reflect the expert's ambiguous understanding of the nature of the parameters in the problem formulation process, are assumed to be characterized as fuzzy numbers.

Definition 3 (α-level set) *The* α*-level set of the fuzzy numbers* $\tilde{c}$ *and* $\tilde{d}$ *is defined as the ordinary set* $(\tilde{c}, \tilde{d})_\alpha$ *for which the degree of their membership functions exceeds the level* α*:*

$$(\tilde{c}, \tilde{d})_\alpha = \{(c, d) | \mu_{\tilde{c}}(c) \geq \alpha \text{ and } \mu_{\tilde{d}}(d) \geq \alpha\}.$$

Now, suppose that the decision maker considers that the degree of all of the membership functions of the fuzzy numbers involved in Problem (2) should be greater than or equal to a certain value of α. Then, for such a degree α, Problem (2) can be interpreted as a conventional multi-objective programming problems in the following way:

$$\begin{array}{ll} \min & F(\mathbf{a}; \mathbf{x}) = (f_1(\mathbf{a}_1; \mathbf{x}), \ldots, f_m(\mathbf{a}_m; \mathbf{x})) \\ \text{s.t.} & \mathbf{x} \in X(\mathbf{b}) \triangleq \{x \in \mathbb{R}^n | g_i(x, b_i) \leq 0, i = 1, \ldots, p\} \\ & (\mathbf{a}, \mathbf{b}) \in (\tilde{\mathbf{a}}, \tilde{\mathbf{b}})_\alpha \end{array} \tag{3}$$

where the coefficient vector $(\mathbf{a}, \mathbf{b}) \in (\tilde{\mathbf{a}}, \tilde{\mathbf{b}})_\alpha$ and $(\mathbf{a}, \mathbf{b})$ are arbitrary for any value in $(\tilde{\mathbf{a}}, \tilde{\mathbf{b}})_\alpha$. This fuzzy subset is formatted for all the vectors whose degree of each membership function exceeds the level α. Here, it is possible to observe that the parameters $(\mathbf{a}, \mathbf{b})$ are treated as decision variables of Problem (3) rather than constants of Problem (2). Through the

ISBN: 978-989-95079-6-8

description of Problem (3), the concepts of (local) α-Pareto optimality is defined as follows.

Definition 4 ((Local) α-Pareto optimal solution) *$x^* \in X(b)$ is said to be a (local) α-Pareto-optimal solution to the α-MONLP(Multi-Objective Non-Linear programming) if and only if there exists no other $x \in X(b)(\bigcap \mathcal{N}(x^*, r))$ and $(a, b) \in (A, B)_\alpha(\bigcap \mathcal{N}(a^*, b^*, r'))$ such that $f_i(x, a_i) \leq f_i(x^*, a_i^*)$, $i = 1, 2, \ldots, k$, with strict inequality holding for at least one i, where the corresponding values of parameters a^* and b^* are called α-level optimal parameters (and $\mathcal{N}(x^*, r) \triangleq \{x \in \mathbb{R}^n | \|x - x^*\| < r\}$ denote the r neighborhood of x^*).*

The (local) α-Pareto-optimal solutions can be obtained through a direct application of the scalarization methods, which transform a multi-objective programming problem into a single-objective programming problem, and the optimal solution obtained is formed by point $(\mathbf{x}, \mathbf{a}, \mathbf{b})$. These set of solutions, however, generally comprises an infinite number of points and the decision should select a single (local) solution based on a subjective criterion.

3 The use of possibility theory in multi-objective optimization under fuzzy environment

Normally, when a multi-objective optimization problems is formulated, many parameters need to be assigned by the decision maker and they may be described by possible values. In most practical situations, it is natural to consider that the possible values of these parameters are often only vaguely known and it is appropriated to interpret them by the decision maker's understanding. This parameters can be represented by fuzzy numbers which intent to describe the possible values that are inserted in the real-world problems. Then, the resulting multi-objective programming problem with fuzzy parameters would be viewed as the more realistic version of the conventional one. In addition, it is necessary to define a model for the quantification of the imprecise data that interpret the decision maker's judgment. There are many comparison approaches among fuzzy numbers and one of them is the possibility theory which was chosen in this work.

3.1 Fuzzy basic concepts

Mathematical programming problems need a precise definition of both the constraints and the objective function to be optimized. Fuzzy sets help handle uncertainties when multi-objective programming problems are formalized in the following form:

$$\begin{array}{ll} \widetilde{\min} & F(\tilde{\mathbf{a}}; \mathbf{x}) \\ \text{s.t.} & \mathbf{x} \in \tilde{\Omega} \end{array} \tag{4}$$

where $F = (f_1, f_2, \ldots, f_m)(m \geq 2)$ is a vector of objectives, $\tilde{\mathbf{a}} \in \mathbb{F}(\mathbb{R}^{n \times m})$ represent the fuzzy parameters in the objective functions and $\tilde{\Omega} \subset \mathbb{F}(\mathbb{R}^n)$ is a subset of feasible solutions. $\mathbb{F}(\mathbb{R})$ defines the set of fuzzy numbers, $\mathbb{F}(\mathbb{R}^n)$ defines the set of n-dimensional vector with fuzzy parameters and $\mathbb{F}(\mathbb{R}^{n \times m})$ defines the set of $n \times m$-dimensional matrix with fuzzy parameters. However, we shall only address the uncertainties of the parameters in the objective functions, in this work.

In order to be able to sort fuzzy numbers in an increasing (decreasing) order, one has to opt for a comparison measure, therefore a possibility measure is used which is defined as:

Definition 5 (Possibility measure) *Let A be a fuzzy subset of U and let $\prod_X$ be a possibility distribution associated with a variable X which takes values in U. The* possibility measure*, $\pi(A)$, of A is defined by*

$$Poss\{X \text{ is } A\} \triangleq \pi(A) \triangleq \sup_{u \in U} \min(\mu_A(u), \pi_X(u)) \tag{5}$$

where μ_A is the membership function of A and π_X is the possibility distribution function of X. It can be interpreted as the possibility that the value X belongs to the set A *and it is defined to be numerically equal to the membership function of X.*

Then, it is possible define a way to compare two fuzzy numbers and this index can be formulated as follows

$$Poss\{\tilde{a}_1 \leq \tilde{a}_2\} = \sup_{u, v \in U; u \leq v} \min(\mu_{\tilde{a}_1}(u), \mu_{\tilde{a}_2}(v))$$

where $\mu_{\tilde{a}_1}$ and $\mu_{\tilde{a}_2}$ are membership functions of $\tilde{a}_1$ and $\tilde{a}_2$. Possibility degree $Poss\{\tilde{a}_1 \leq \tilde{a}_2\}$ shows to what extent $\tilde{a}_1$ is possibly less than or equal to $\tilde{a}_2$, as described in [7, 11].

The definition above enables one to define a fuzzily ordered set $\mathbb{F}(\mathbb{R})$ that is an extension of the classical ordered set. A set is said to be *completely ordered* if it satisfies the following conditions:

Definition 6 (Ordered fuzzily set) *A fuzzy subset $A \subset \mathbb{F}(\mathbb{R})$ is fuzzily ordered with respect to the possibility measure if each element in A satisfies the following basic properties:*

1. $Poss[\tilde{\mathbf{a}}_1 \leq \tilde{\mathbf{a}}_1] = 1$;
2. $Poss[\tilde{\mathbf{a}}_1 \leq \tilde{\mathbf{a}}_2] \geq \alpha_1$ *and* $Poss[\tilde{\mathbf{a}}_2 \leq \tilde{\mathbf{a}}_3] \geq \alpha_2 \Rightarrow$ $Poss[\tilde{\mathbf{a}}_1 \leq \tilde{\mathbf{a}}_3] \geq \min\{\alpha_1, \alpha_2\}$;
3. $Poss[\tilde{\mathbf{a}}_1 \leq \tilde{\mathbf{a}}_2] \geq \alpha_1$ *and* $Poss[\tilde{\mathbf{a}}_2 \leq \tilde{\mathbf{a}}_1] \geq \alpha_2 \Rightarrow$ $Poss[\tilde{\mathbf{a}}_1 = \tilde{\mathbf{a}}_2] \geq \min\{\alpha_1, \alpha_2\}$;

$\forall\, \tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_2, \tilde{\mathbf{a}}_3 \in A$ *and* $\forall\, \alpha_1, \alpha_2 \in [0, 1]$.

According to the expressions above, a fuzzy subset $A \subset \mathbb{F}(\mathbb{R})$ is completely ordered. However, a fuzzy subset of $\mathbb{F}(\mathbb{R}^m)$ is only partially ordered. Therefore, the concept of optimal solution for single objective problems, which was defined in [3, 15], does no fit into the multi-objective formulation, unless the problem admits the so-called *ideal solution*, i.e. a single solution that simultaneously minimizes all objective functions as below:

Definition 7 (Ideal solution) *The ideal solution $\tilde{\mathbf{y}}$ of the multi-objective problem is defined as*

$$\tilde{\mathbf{y}}_i = f_i(\tilde{\mathbf{a}}_i; \mathbf{x}^i), \;\; i = 1, \ldots, m$$

where $\mathbf{x}^i = \arg\min_{\mathbf{x} \in \mathbb{R}^n} f_i(\tilde{\mathbf{a}}_i; \mathbf{x})$.

The problem is said to admit an ideal solution whenever the set of arguments $\{\mathbf{x}^i,\ i = 1, \ldots, m\}$, possesses a single element. Because the multi-objective framework is most often

ISBN: 978-989-95079-6-8

employed in problems with conflicting objectives, a typical multi-objective problem is unlikely to admit such a solution. However, due the existence of an ideal solution is very rare, such a possibility will not be considered in the present analysis.

Conceptually, an efficient solution is one which is not dominated by any other feasible solution. Hence, the domination concept of a fuzzy multi-objective problem should reflect the decision maker's preferences. In this work, a fuzzy dominance concept is proposed which can be adjusted to the decision maker's preferences. This renders the proposed approach flexible and customizable, and possibly applicable to a wide range of problems. For any point $\mathbf{x}^0 \in \mathbb{R}^n$, consider the following subsets

$$\Omega_<(\mathbf{x}^0;\alpha) \triangleq \{\mathbf{x} \in \mathbb{R}^n : Poss[F(\tilde{\mathbf{a}};\mathbf{x}) \leq F(\tilde{\mathbf{a}};\mathbf{x}^0)] \geq \alpha \text{ and } Poss[F(\tilde{\mathbf{a}};\mathbf{x}) = F(\tilde{\mathbf{a}};\mathbf{x}^0)] < 1\}$$

$$\Omega_\geq(\mathbf{x}^0;\alpha) \triangleq \{\mathbf{x} \in \mathbb{R}^n : Poss[F(\tilde{\mathbf{a}};\mathbf{x}) \geq F(\tilde{\mathbf{a}};\mathbf{x}^0)] \geq \alpha\}$$

$$\Omega_\sim(\mathbf{x}^0;\alpha) \triangleq \{\mathbf{x} \in \mathbb{R}^n : \max\{Poss[F(\tilde{\mathbf{a}};\mathbf{x}) \leq F(\tilde{\mathbf{a}};\mathbf{x}^0)], Poss[F(\tilde{\mathbf{a}};\mathbf{x}) \geq F(\tilde{\mathbf{a}};\mathbf{x}^0)]\} \leq \alpha\}$$

The subset $\Omega_<(\mathbf{x}^0;\alpha)$ comprises the points in $\mathbb{R}^n$ that dominate $\mathbf{x}^0$, whereas $\Omega_\geq(\mathbf{x}^0;\alpha)$ encompasses the points in $\mathbb{R}^n$ that are dominated by $\mathbf{x}^0$. The set of points that neither dominate nor are dominated by $\mathbf{x}^0$ is denoted by $\Omega_\sim(\mathbf{x}^0;\alpha)$. The parameter α is a vector where each one of the terms, α_i with $i = 1, 2, \ldots, m$, belong to the interval $[0, 1]$. Those sets being defined, one can denote the set of fuzzy Pareto-optimal solutions as below:

Definition 8 (Fuzzy Pareto-optimal solution) *$\mathbf{x}^* \in \Omega$ is said be a fuzzy Pareto-optimal solution if there exists no other $x \in \Omega$ such that $Poss[f_i(\tilde{\mathbf{a}}_i;\mathbf{x}) \leq f_i(\tilde{\mathbf{a}}_i;\mathbf{x}^*)] \geq \alpha_i$, $\forall i$ and $Poss[f_j(\tilde{\mathbf{a}}_j;\mathbf{x}) = f_j(\tilde{\mathbf{a}}_j;\mathbf{x}^*)] < 1$ for at least one j, where $\alpha_i \in [0, 1]$, $\forall i$.*

For difficult optimization problems it is often the case that a local optimal solution is acceptable. A local efficient solution for the proposed problem is defined below:

Definition 9 (Fuzzy local Pareto-optimal solution) *$\mathbf{x}^* \in \Omega$ is said to be a fuzzy local Pareto-optimal solution if there is a real number $\delta \geq 0$ such that there exists no other $x \in \Omega \cap \mathcal{N}(x^*, \delta)$ such that $Poss[f_i(\tilde{\mathbf{a}}_i;\mathbf{x}) \leq f_i(\tilde{\mathbf{a}}_i;\mathbf{x}^*)] \geq \alpha_i$, $\forall i$ and $Poss[f_j(\tilde{\mathbf{a}}_j;\mathbf{x}) = f_j(\tilde{\mathbf{a}}_j;\mathbf{x}^*)] < 1$ in at least one j, where $\alpha_i \in [0, 1]$, $\forall i$.*

Note that the definition above implies that a candidate solution to the proposed fuzzy problem is (locally) non-dominated or efficient, if one cannot find (in a certain vicinity) another solution that simultaneously improves all the objective functions. This interpretation matches the classical counterpart of multi-objective optimization.

The convexity hyphotesis determine that the neighbourhood of each one local solution involves whole the feasible region.

Theorem 1 *Let $f_i : \Omega \subset \mathcal{X} \to \mathbb{F}(\mathcal{Y})$, $i = 1, \ldots, m$ a convex fuzzy functions about a convex subset Ω of a linear space $\mathcal{X}$. Then whole locally efficient solution is globaly efficient solution.*

Proof: Let $\mathbf{x}^* \in \Omega$ is a locally efficient solution. By definition of convex subset, we obtain $\lambda\mathbf{x}^* + (1-\lambda)\mathbf{x} \in \Omega$, $\forall \mathbf{x} \in \Omega - \mathcal{N}(\mathbf{x}^*, \epsilon)$, with $\epsilon > 0$ e $\lambda \in [0, 1]$. Suppose $\lambda\mathbf{x}^* + (1-\lambda)\mathbf{x} \in \Omega \cap \mathcal{N}(\mathbf{x}^*, \epsilon)$ then by the fuzzy Pareto-optimal solution definition, we obtain $Poss[f_i(\tilde{a}_i;\mathbf{x}^*) \leq f_i(\tilde{a}_i;\lambda\mathbf{x}^* + (1-\lambda)\mathbf{x})] \geq \alpha_i^1$, $i \in \mathcal{I} = \{1, 2, \ldots, m\}$, and $Poss[f_j(\tilde{a}_j;\mathbf{x}^*) = f_j(\tilde{a}_j;\lambda\mathbf{x}^* + (1-\lambda)\mathbf{x})] < 1$ for at least one $j \in \mathcal{I}$, where $\alpha_i^1 \in (0, 1]$, $\forall i \in \mathcal{I}$. By the convex fuzzy function definition, we obtain $f_i(\tilde{a}_i;\lambda\mathbf{x}^* + (1-\lambda)\mathbf{x}) \lesssim \lambda f_i(\tilde{\mathbf{a}}_i;\mathbf{x}^*) + (1-\lambda) f_i(\tilde{\mathbf{a}}_i;\mathbf{x})$, $\forall i \in \mathcal{I}$, which it can be rewritten by using Possibility Theory as $Poss[f_i(\tilde{a}_i;\lambda\mathbf{x}^* + (1-\lambda)\mathbf{x}) \leq \lambda f_i(\tilde{\mathbf{a}}_i;\mathbf{x}^*) + (1-\lambda) f_i(\tilde{\mathbf{a}}_i;\mathbf{x})\}] \geq \alpha_i^2$, where $\alpha_i^2 \in (0, 1]$, $\forall i \in \mathcal{I}$. Thus, by using the ordered fuzzily subset definition, we obtain $Poss[f_i(\tilde{a}_i;\mathbf{x}^*) \leq \lambda f_i(\tilde{\mathbf{a}}_i;\mathbf{x}^*) + (1-\lambda) f_i(\tilde{\mathbf{a}}_i;\mathbf{x})\}] \geq \min\{\alpha_i^1, \alpha_i^2\} \Rightarrow Poss[f_i(\tilde{a}_i;\mathbf{x}^*) \leq f_i(\tilde{a}_i;\mathbf{x})] \geq \min\{\alpha_i^1, \alpha_i^2\}$, $\forall i \in \mathcal{I}$. By selecting a determined objective function $k \in \mathcal{I}$ and $k \neq i$, by the Theorem 2, we guarantee that $Poss[f_k(\tilde{a}_k;\mathbf{x}^*) = f_k(\tilde{a}_k;\mathbf{x})] < 1$ for at least one $k \in \mathcal{I}$. ■

3.2 Characterization of fuzzy efficient solutions

The characterization of efficient solutions, $efi(\Omega)$, by means of well defined scalar problems is a recurrent approach in fuzzy multi-objective problems. The following theorem relates efficient solutions and scalar problems.

Theorem 2 *$\mathbf{x}^* \in efi(\Omega)$ if and only if $\mathbf{x}^*$ solves the m scalar problems*

$$P_k: \quad \min_{\mathbf{x}\in\Omega} \quad f_k(\tilde{\mathbf{a}}_k;\mathbf{x}) \qquad s.t. \quad f_l(\tilde{\mathbf{a}}_l;\mathbf{x}) \lesssim f_l(\tilde{\mathbf{a}}_l;\mathbf{x}^*), \quad l = 1, 2, \ldots, m, \quad \forall\, l \neq k. \tag{6}$$

Proof: ($\Rightarrow$) If $\mathbf{x}^* \in efi(\Omega)$, then there exist no other $x \in \Omega$ such that $Poss[f_i(\tilde{\mathbf{a}}_i;\mathbf{x}) \leq f_i(\tilde{\mathbf{a}}_i;\mathbf{x}^*)] \geq \alpha_i$, $i = 1, 2, \ldots, m$, and $Poss[f_j(\tilde{\mathbf{a}}_j;\mathbf{x}) = f_j(\tilde{\mathbf{a}}_j;\mathbf{x}^*)] < 1$, for any j. In this case $\mathbf{x}^*$ solves (6) for all k.
($\Leftarrow$) Suppose $\mathbf{x}^*$ solves (6), but $\mathbf{x}^* \notin efi(\Omega)$, then there exists another $\mathbf{x} \in \Omega$ such that $Poss[f_i(\tilde{\mathbf{a}}_i;\mathbf{x}) \leq f_i(\tilde{\mathbf{a}}_i;\mathbf{x}^*)] \geq \alpha_i$, $\forall i$, and for some j, $Poss[f_j(\tilde{\mathbf{a}}_j;\mathbf{x}) \leq f_j(\tilde{\mathbf{a}}_j;\mathbf{x}^*)] < 1$. Therefore, $\mathbf{x}^*$ does not solve Problem (6). This contradiction concludes the proof. ■

The development of analytical conditions to efficient solutions, based on the characterization of non-dominated solutions to problems P_k, $k = 1, 2, \ldots, m$, is an important tool in the theoretical analysis. However, such an analysis yields only m non-dominated solutions, one to each scalar problem and is therefore, unable to generate the whole Pareto-optimal set.

Employing a similar analysis to the one presented above, we now establish the relationship between non-dominated solutions of a fuzzy multi-objective problem and solutions to the weighting problem. An alternative characterization based on the linear combination of the objectives can be expressed as

Theorem 3 *Let $\mathbf{x}^* \in \Omega$ solve the problem*

$$P_w: \quad \min_{\mathbf{x}\in\Omega} \quad \langle \mathbf{w}, F(\tilde{\mathbf{a}};\mathbf{x})\rangle = \textstyle\sum_{i=1}^m \omega_i f_i(\tilde{\mathbf{a}}_i;\mathbf{x}) \tag{7}$$

for some $\mathbf{w} \in \mathbb{R}^m$, $\mathbf{w} \geq \mathbf{0}$ and $\sum_{i=1}^m w_i = 1$. Then $\mathbf{x}^* \in efi(\Omega)$ if

ISBN: 978-989-95079-6-8

(i) $\mathbf{x}^*$ *is the unique solution(7), or*

(ii) $w_i > 0,\ i = 1, \ldots, m.$

Proof: *(i)* If $\mathbf{x}^* \in \Omega$ is a unique solution of (7), then $\forall \mathbf{x} \in \Omega$ and by definition, we obtain $Poss\left[\sum_{i=1}^{m} w_i \left(f_i(\tilde{\mathbf{a}}_i; \mathbf{x}^*) - f_i(\tilde{\mathbf{a}}_i; \mathbf{x})\right) < 0\right] \geq \min_i\{\alpha_i\}$. Suppose $\mathbf{x}^* \notin efi(\Omega)$, i.e., there exists at least one $\mathbf{x}^0 \in \Omega$ such that $Poss[f_i(\tilde{\mathbf{a}}_i; \mathbf{x}^0) \leq f_i(\tilde{\mathbf{a}}_i; \mathbf{x}^*)] \geq \alpha_i,\ i = 1, 2, \ldots, m$ and $Poss[f_j(\tilde{\mathbf{a}}_j; \mathbf{x}^0) = f_j(\tilde{\mathbf{a}}_j; \mathbf{x}^*)] < 1$, for some j. This contradicts the uniqueness hypothesis, because $\mathbf{w} \geq \mathbf{0}$. Thus, $\mathbf{x}^* \in efi(\Omega)$.
(ii) Suppose $\mathbf{x}^* \notin efi(\Omega)$, but $\mathbf{x}^*$ is a solution of (7). Then there exists a $\mathbf{x}^0 \in \Omega$ such that $Poss[f_i(\tilde{\mathbf{a}}_i; \mathbf{x}^0) \leq f_i(\tilde{\mathbf{a}}_i; \mathbf{x}^*)] \geq \alpha_i,\ i = 1, 2, \ldots, m$ and $Poss[f_j(\tilde{\mathbf{a}}_j; \mathbf{x}^0) = f_j(\tilde{\mathbf{a}}_j; \mathbf{x}^*)] < 1$, for any j. Hence,

$$Poss\left[\sum_{i=1}^{m} w_i \left(f_i(\tilde{\mathbf{a}}_i; \mathbf{x}^*) \leq f_i(\tilde{\mathbf{a}}_i; \mathbf{x}^0)\right) > 0\right] \geq \min_i\{\alpha_i\},\ \forall i.$$

A contradiction and therefore $\mathbf{x}^* \in efi(\Omega)$. ■

4 Results and analysis

The problems we use to evaluate this method are two hypothetic mathematical formulations, which are described in [5], with the fuzzy approach described in Section 3. Nevertheless, they are efficiency in validating the realized study. They were resolved using an modified implementation of **NSGA-II** that solves multi-objective programming problems with constraints or not. This modification was made in the comparison of the objective functions with fuzzy parameters between two feasible solutions which uses the concepts of fuzzy Pareto optimal solutions described in this work.

The vagueness was inserted into the costs of the objective function and fuzzy numbers are interpreted in the form $(a, \underline{a}, \overline{a})_{LR}$ where a is the modal value, $\underline{a}$ is the scattering left and $\overline{a}$ is the scattering right of each fuzzy number.

Example 1 (Schaffer's problem)

$$\begin{aligned} \min \quad & f_1(\tilde{\mathbf{a}}_1; \mathbf{x}) = (x + \tilde{a}_1)^2 \\ \min \quad & f_2(\tilde{\mathbf{a}}_2; \mathbf{x}) = (x - \tilde{a}_2)^2 \\ s.t \quad & -5 \leq \mathbf{x} \leq 5 \end{aligned} \tag{8}$$

where $\tilde{a}_1 = (0, 0, 2)_{LR}$ *and* $\tilde{a}_2 = (2, 1, 1)_{LR}$.

The figures below present the fuzzy solution and fuzzy front.

In Figure 2, each star represents one solution of this multi-objective problem by using the fuzzy Pareto optimal concept defined by Sakawa to $\alpha = 0.8$*, while each square represents one solution by using the fuzzy Pareto optimal concept described in this work to* $\alpha = 0.8$*, too. It can be observed that a range of possible Pareto optimal solutions is formed when the squares are merged. We can also see that many stars are inside some squares, i.e., this solutions have a degree of possibility great or equal to 0.8 and belong to the range of Pareto optimal solutions obtained by the definition that uses possibility theory.*

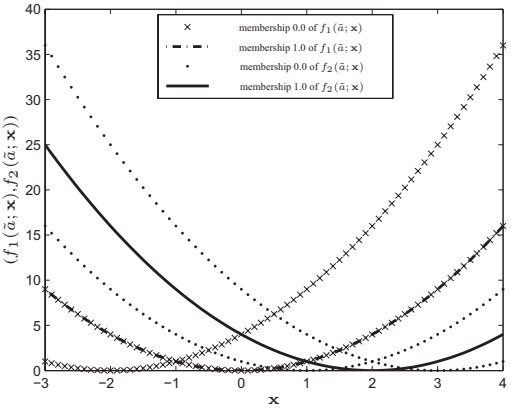


Figure 1: Objective functions

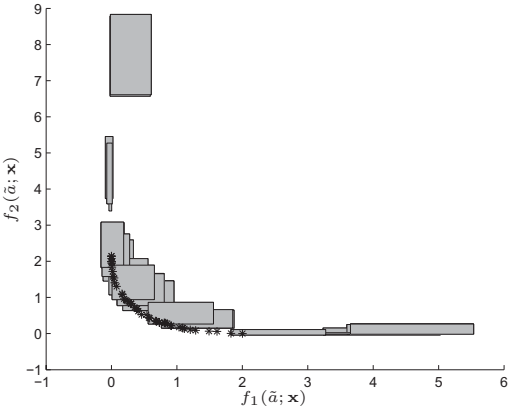


Figure 2: Pareto front

Example 2 (Binh and Kern's problem)

$$\begin{aligned} \min \quad & f_1(\tilde{\mathbf{a}}_1; \mathbf{x}) = \tilde{a}_{11} x_1^2 + \tilde{a}_{12} x_2^2 \\ \min \quad & f_2(\tilde{\mathbf{a}}_2; \mathbf{x}) = (x_1 - \tilde{a}_{21})^2 + (x_2 - \tilde{a}_{22})^2 \\ s.t. \quad & (x_1 - 5)^2 + x_2^2 \leq 25 \\ & (x_1 - 8)^2 + (x_2 + 3)^2 \geq 7.7 \\ & 0 \leq x_1 \leq 5,\ \ 0 \leq x_2 \leq 3 \end{aligned} \tag{9}$$

where $\tilde{a}_{11} = \tilde{a}_{12} = (4, 1, 1)_{LR}$, $\tilde{a}_{21} = \tilde{a}_{22} = (5, 1, 1)_{LR}$.

The figures below present the fuzzy solution and fuzzy front.

In Figure 3 are shown the function objectives of Problem (2) where the superior drawing represents the function $f_1(\tilde{\mathbf{a}}_1; \mathbf{x})$ *and the inferior one represents the function* $f_2(\tilde{\mathbf{a}}_2; \mathbf{x})$*. The drawings with solid line represent the value of the objective functions with* $\alpha = 1$ *while the ones with dotted line represent the value of the objective functions with* $\alpha = 0$*.*

Again, each star represents one solution by using the concept defined by Sakawa to $\alpha = 0.8$*, while each square represents one solution by using the concept described here to* $\alpha = 0.8$*, too. In this case, the stars are in the imaginary bound of the Pareto front range formed by possible Pareto solutions, i.e., the solutions obtained by Sakawa's definition have a satisfaction level closed in* $\alpha = 0.8$*.*

ISBN: 978-989-95079-6-8

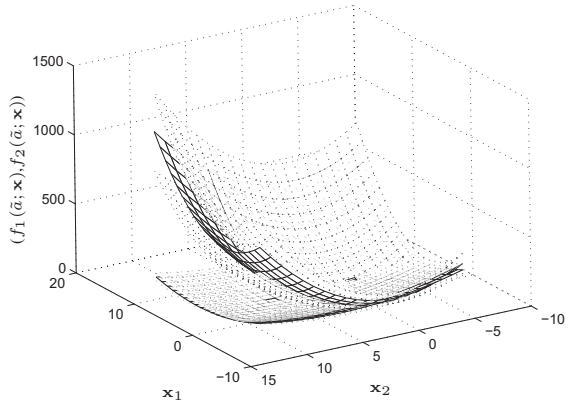


Figure 3: Objective functions

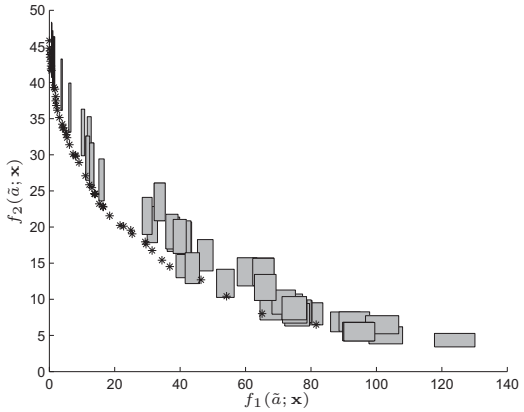


Figure 4: Pareto front

5 Conclusion

The use of defuzzification methods to solve fuzzy mathematical programming problems is very common. These methods transform a fuzzy number into a classical one, but some data are always lost in each case. The Pareto-optimality presented in this work makes use of different defuzzification methods in various stages of the definitions.

Multi-objective Programming problems are very important in a variety of both theoretical and practical areas. As ambiguity and vagueness are natural and ever-present in real-life situations that require precise solutions, it makes perfect sense to attempt to address these problems using Fuzzy Multi-objective Programming problems. In this context, this paper presented a novel theory to determine Pareto-optimality conditions which provide the user with a fuzzy solution. This theory is an expansion of the classical Pareto-optimality theory and demonstrates the necessary conditions for fuzzy Pareto-optimality. Some numerical examples are presented to validate the theory outlined.

The authors aim firstly to extend the line of investigation regarding Fuzzy Quadratic Programming problems in order to try to solve practical real-life problems by facilitating the building of Decision Support Systems. This requires the involvement of fuzzy costs as well as fuzzy coefficients, as a must.

Acknowledgments

The authors want to thank the support provided by the Brazilian agency CAPES.

References

[1] R. E. BELLMAN and L. A. ZADEH, Decision-marking in a fuzzy environment. *Management Science* 17(4):B141–B164, 1970

[2] L. CAMPOS and J. L. VERDEGAY. Linear programming problems and ranking of fuzzy numbers. *Fuzzy Sets and Systems* 32:1–11, 1989.

[3] L. A. P. CANTÃO and A. YAMAKAMI, Nonlinear programming with fuzzy parameters: Theory and applications In: International Conference on Computational Intelligence for Modelling, Control and Automation, CIMCA 2003, Vienna, Austria, 2003.

[4] V. CHANKONG and Y. Y. HAIMES. *Multiobjective decision making: Theory and Methodology*, volume 8 of *North Hollando series in system science and engineering*. North Holland, New York, USA, 1983.

[5] K. DEB. *Multi-objective optimization using evolutionary algorithms*. John Wiley & Sons, Chichester, UK, 2001.

[6] D. DUBOIS and H. PRADE. *Fuzzy sets and systems: Theory and Application*. Academic Press, San Diego, USA, 1980.

[7] D. DUBOIS and H. PRADE. Ranking fuzzy numbers in the setting of possibility theory. *Information Sciences*, 30(3):183–224, 1983.

[8] F. JIMÉNEZ and J. M. CADENAS and G. SÁNCHEZ and A. F. GÓMEZ-SKARMETA and J. L. VERDEGAY. Multi-objective evolutionary computation and fuzzy optimization. *International Journal of Approximate Reasoning* 43:59–75, 2006.

[9] A. KAUFMANN and M. M. GUPTA. *Introduction to fuzzy arithmetic: theory and applications*. Van Nostrand Reinhold, New York, USA, 1984.

[10] G. J. KLIR and B. YUAN. *Fuzzy sets and fuzzy logic: theory and applications*. Prentice Hall, New Jersey, USA, 1995.

[11] M. INUIGUSHI: On possibilistic/fuzzy optimization. In: 12th IFSA World Congress, Cancun, Mexico, June 2007, pag. 351-360

[12] V. PARETO. *Cours d'economique politique*, volume I. Macmillan, Paris, FR, 1987.

[13] V. PARETO. *Le cours d'economique politique*, volume II. Macmillan, London, UK, 1987.

[14] W. PEDRYCS and F. GOMIDE. *An introduction of fuzzy sets: analisys and design*. MIT press, Cambridge, UK, 1998.

[15] M. SAKAWA. *Fuzzy sets and interactive multiobjective optimization*. Plenum Press, New York, USA, 1993.

[16] L. A. ZADEH. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[17] L. A. ZADEH. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1(1):3–28, 1978.

[18] L. A. ZADEH. Is there a need for fuzzy logic?. *Information Sciences*. 178:2751–2779, 2008.

[19] H.-J. ZIMMERMANN. *Fuzzy set theory and its applications*. Kluwer Academic Publishers, Massachusetts, USA, third edition, 1996.

ISBN: 978-989-95079-6-8

Terrain Morphology Classification over Fuzzy Digital Elevation Models

Jorge M. F. Santos[1]
Departamento de Matemática da Universidade de Coimbra and INESC Coimbra, Portugal 1
Email: jmfs@mat.uc.pt

Abstract— *One way to incorporate uncertainty associated to the data and to the interpolation method in a digital elevation model (DEM) is to use Fuzzy Sets Theory. A fuzzy DEM can be generated from fuzzy data, where elevations are represented by fuzzy numbers. To use that type of terrain models in further applications, a generalization of known methods is needed. An example about the generalization of terrain morphology classification methods is proposed in this work. Given a fuzzy DEM, a procedure to find ridges and other morphological features will be analyzed. Two examples of this approach will be presented to illustrate the proposed methods.*

Keywords— Fuzzy DEM, fuzzy morphological feature, signal of a fuzzy number.

1 Introduction

Automated identification of objects representing terrain features (such as peaks, channels and ridges) from a digital elevation model is a well known procedure in Geomorphology. The development of those methods has important implications in terms of spatial analysis and for the generalization of spatial databases, for example. Another important application is the study of water flow over the terrain surface that represents a fundamental geomorphological process and it is closely related to its shape: it constraints the water route over the surface and it is mainly shaped by this flow.

The topographic surface is usually defined through a gridded DEM constructed by interpolation from a finite set of samples and it is a well-known problem in geographic modeling. The uncertainty associated to this process generates uncertain elevation values which can be modeled using fuzzy numbers [1, 2, 3]. Therefore, methods to identify morphological terrain features have to be generalized to fuzzy DEMs in order to express the impact of elevation uncertainty in terrain classification. In this work we will study one approach to this problems, generalizing known methods over fuzzy DEMs.

2 Morphology classification

To identify morphological objects over the terrain surface it is adopted here the strategy developed by Wood [4]. This problem can be seen as an image processing problem to classify each *pixel* (grid cell) in one class of *geomorphological feature* (usually, *peak*, *ridge*, *pass*, *plane*, *channel* or *pit*). The feature classification is based on table 1, where the signals [1] of the four morphological parameters: *slope* (S), *cross sectional curvature* ($crosc$), *maximum profile convexity* ($maxic$) and *minimum profile convexity* ($minic$), are given by

$$sign(S) = sign(\arctan\sqrt{d^2+e^2}), \quad (1)$$

$$sign(crosc) = sign(cde - bd^2 - ae^2), \quad (2)$$

$$sign(maxic) = -sign(a+b-\sqrt{(a-b)^2-c^2}), \quad (3)$$

$$sign(minic) = -sign(a+b+\sqrt{(a-b)^2-c^2}). \quad (4)$$

The values of a, b, c, d and e are evaluated from the first and second order derivatives of the quadric surface

$$z = ax^2 + by^2 + cxy + dx + ey + f, \quad (5)$$

usually fitted to a 3x3 data window [4].

Table 1: Feature classification criteria [4].

Feature	$slope$	$crosc$	$maxic$	$minic$
Peak	0	all	+	+
Ridge	0	all	+	0
	+	+	all	all
Pass	0	all	+	-
Plane	0	all	0	0
	+	0	all	all
Channel	0	all	0	-
	+	-	all	all
Pit	0	all	-	-

3 Fuzzy morphology classification

The above procedures are based on a regular elevation grid. Considering that the elevations have some degree of uncertainty which is expressed by fuzzy numbers, we will study the generalization of those procedures to a fuzzy DEM. The first step is to evaluate the first and second order fuzzy derivatives. This is done using interval arithmetic over the fuzzy numbers $\alpha - levels$. Those fuzzy derivatives will be used in equations (1 to 4) to obtain fuzzy morphological parameters using again interval arithmetic [5].

3.1 Signal of a fuzzy number

To use the fuzzy version of the above *slope*, *cross sectional curvature*, *maximum profile curvature* and *minimum profile curvature* it is need now a generalization of the $signal$ function. It is proposed here to adapt the concept of *overtaking between fuzzy numbers* in [3] to decide how much a fuzzy value is negative, positive or zero. This concept is based on the notion of *overtaking between intervals* defined by

[1] We assume that $sign$ is a three-valued function in $\{-1, 0, 1\}$.

ISBN: 978-989-95079-6-8

$$\sigma(X,Y) = \begin{cases} 0, & X_u \leq Y_l \\ \frac{X_u - Y_l}{X_u - X_l}, & X_u > Y_l \wedge X_l \leq Y_l \\ 1, & X_l < Y_l \end{cases}, \quad (6)$$

where X and Y are intervals defined by $X = [X_l, X_u]$ and $Y = [Y_l, Y_u]$.

Following, the *overtaking between fuzzy numbers* is

$$\sigma(\tilde{X}, \tilde{Y}) = \int_0^1 \sigma([\tilde{X}]_\alpha, [\tilde{Y}]_\alpha) w(\alpha) d\alpha, \quad (7)$$

where $[\tilde{X}]_\alpha$ and $[\tilde{Y}]_\alpha$ are the $\alpha - levels$ of fuzzy numbers $\tilde{X}$ and $\tilde{Y}$ and $w(\alpha)$ is a weight function $w : [0,1] \rightarrow \Re$ such that

$$\begin{aligned} & w(\alpha) \geq 0 \\ & \alpha < \alpha' \Rightarrow w(\alpha) < w(\alpha') \\ & \int_0^1 w(\alpha) d\alpha = 1. \end{aligned} \quad (8)$$

.

To apply this strategy, it is needed to define a fuzzy zero value. When fuzzy numbers like the fuzzy morphological parameters have very different support widths, ranging from 10 to 0.01, approximately, scaling problems may arise if comparison is needed. To prevent those problems, the zero value will depend on the fuzzy value to be evaluated. The fuzzy zero will have the same membership of the respective fuzzy value but there will be a shift over the real line such that the $\alpha - level$ of maximum membership will be centered in the zero of real line. For triangular fuzzy numbers it will be $\tilde{0} = (X_l - X_m/0/X_u - X_m)$ for fuzzy number $\tilde{X} = (X_l/X_m/X_u)$, see figure 1. This type of fuzzy zero value will be called here *twin zero* of $\tilde{X}$.

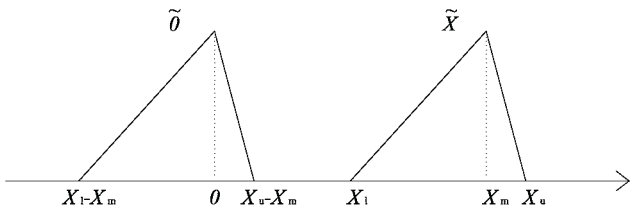


Figure 1: Twin zero of a triangular fuzzy number $\tilde{X}$.

First, to evaluate how much an interval X is positive, it is used a named $positive$ function defined by

$$pos(X) = \begin{cases} 0, & X_l \leq 0_l \\ \frac{X_u - 0_u}{X_u - X_l}, & 0_l < X_l < 0_u \\ 1, & X_l \geq 0_u \end{cases}, \quad (9)$$

where $[0_l, 0_u]$ is the twin zero of $X = [X_l, X_u]$, that is, it has the same width and it is centered in the real zero value. In the same way, $negative$ and $zero$ functions for intervals can be defined by

$$neg(X) = \begin{cases} 0, & X_u \geq 0_u \\ \frac{0_l - X_l}{X_u - X_l}, & 0_l < X_u < 0_u \\ 1, & X_u \leq 0_l \end{cases}, \quad (10)$$

and

$$zer(X) = \begin{cases} 0, & X_l \geq 0_u \\ \frac{0_u - X_l}{X_u - X_l}, & 0_l < X_l < 0_u \\ 1, & X_l = 0_l \\ \frac{X_u - 0_l}{X_u - X_l}, & 0_l < X_u < 0_u \\ 0, & X_u \leq 0_l \end{cases}. \quad (11)$$

Since the width of the zero interval and of X are the same, it is easy to see that[2]: $pos(X) + neg(X) + zer(X) = 1$ and $zer(X) = 1 - pos(X) \dot{\vee} zer(X) = 1 - neg(X)$. To extend those functions to fuzzy numbers it is used the same approach of (7):

$$pos(\tilde{X}) = \int_0^1 pos([\tilde{X}]_\alpha) w(\alpha) d\alpha, \quad (12)$$

where $w(\alpha)$ is the weight function in (8). The same can be done for $negative$ and $zero$ functions.

3.2 Fuzzy morphological features

From table 1, we can evaluate how much a grid cell belongs to a feature class using the above functions to evaluate the signals of the fuzzy morphological parameters and the standard fuzzy conjunction and disjunction defined by min and max operators [6]. This procedure will generate the fuzzy morphological classification, providing a method to quantify the membership of every terrain grid cell to the *fuzzy peak*, *fuzzy ridge*, *fuzzy pass*, *fuzzy plane*, *fuzzy channel* or *fuzzy pit* classes.

4 Examples

Two examples will be presented in this section to illustrate the methods discussed above. In the first case it will be used a digital elevation model of Coimbra region, crossed by Mondego river with narrow valleys and steep slopes (see figure 2). The second example is Kilimanjaro region, where a big volcanic cone lies down on the middle of a plain (see figure 5). In the following figures, all distance units are in meters and angles in degrees.

4.1 Coimbra region

The study region is a 2.0 by 2.1 kilometers rectangle, with elevations ranging from 20 do almost 300 meters. The elevation uncertainty model here is just the root mean square error (RMSE) provided by the map producer. It was introduced a small random gaussian variation added to that RMSE to define the support of symmetrical triangular fuzzy number to express elevation uncertainty. The modal values are taken from the elevations of the crisp DEM. Some results can be seen in figures 3 and 4.

4.2 Kilimanjaro mountain

It was used a rectangle with about 85 by 80 kilometers to enclose Kilimanjaro mountain, with elevations going from 800 to almost 6000 meters. The DEM for this region was taken from SRTM mission, version 4, with original gaps filled by interpolation and with other elevation data available available for that region. To quantify the distribution of DEM elevation uncertainty, about 100 GPS control points were measured around the mountain and in the top. The difference between the elevations of those points and the DEM were calculated

[2]The symbol $\dot{\vee}$ represents xor operator

ISBN: 978-989-95079-6-8

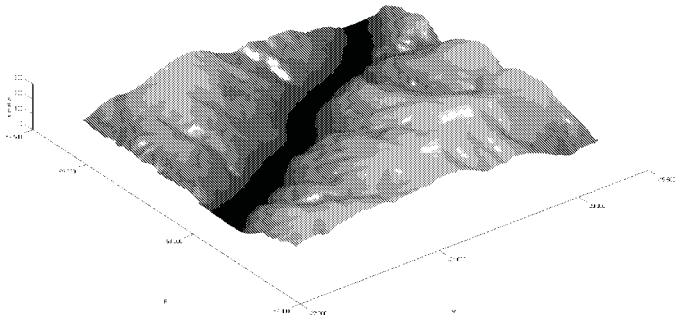


Figure 2: Coimbra digital elevation model.

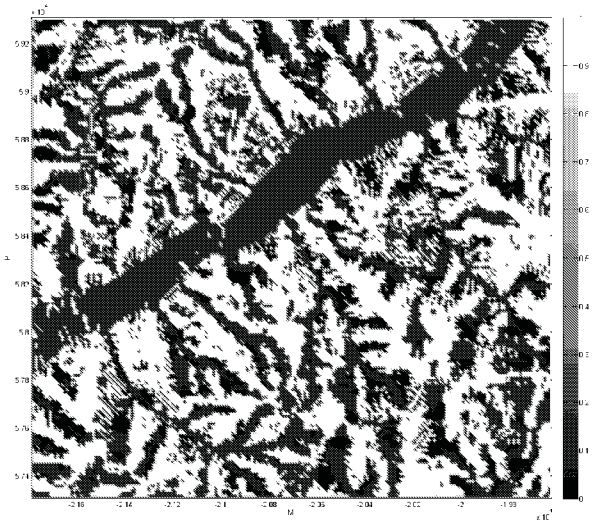


Figure 3: Fuzzy ridge membership for Coimbra study area.

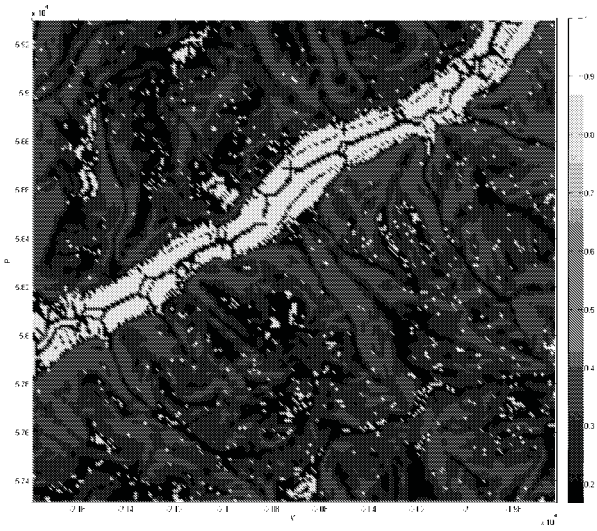


Figure 4: Fuzzy plane membership for Coimbra.

and an error surface were interpolated from those error samples. The error surface in figure 6 was used to estimate the support of symmetric triangular fuzzy numbers that will express the fuzzy elevations. The modal values of those fuzzy numbers are given by the crisp DEM. Some results can be seen in figures 7 and 8. Instead of crisp 0 or 1 membership values, now there are partial membership values to the fuzzy morphological objects, reflecting the uncertainty of the terrain model. Comparing figures 7 and 8, it is possible to detect visually some correlation between the error surface and the fuzzy membership values.

5 Conclusions

This work gave an example that methods applicable to crisp digital elevation models can be generalized to be also applicable to fuzzy DEMs or to similar gridded data expressed by fuzzy numbers. This approach has the advantage to allow the inclusion of uncertainty modeled by Fuzzy Set Theory in geographic analysis. The examples applications presented here are just a first step to develop more robust and advanced terrain classification methods to apply over fuzzy DEMs. Fuzzy values of morphological parameters, like slope and curvatures, have sometimes support intervals too wide. Therefore, those numerical operations are very sensitive to the input values and maybe more robust algorithms are needed here. These methods can also be combined with multi-scale classification methods used in [7, 8, 9, 10] or others approaches like in [11], for example, in order to improve the automated identification of morphological objects.

ISBN: 978-989-95079-6-8

Appendix

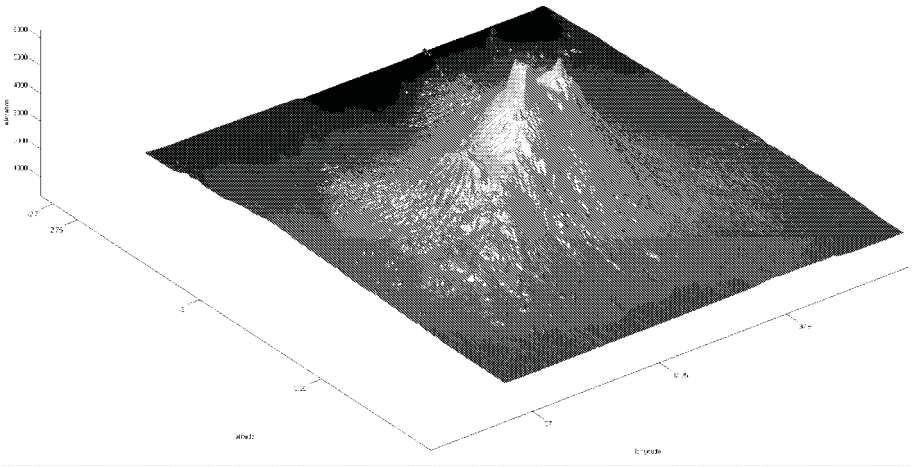


Figure 5: Kilimanjaro digital elevation model.

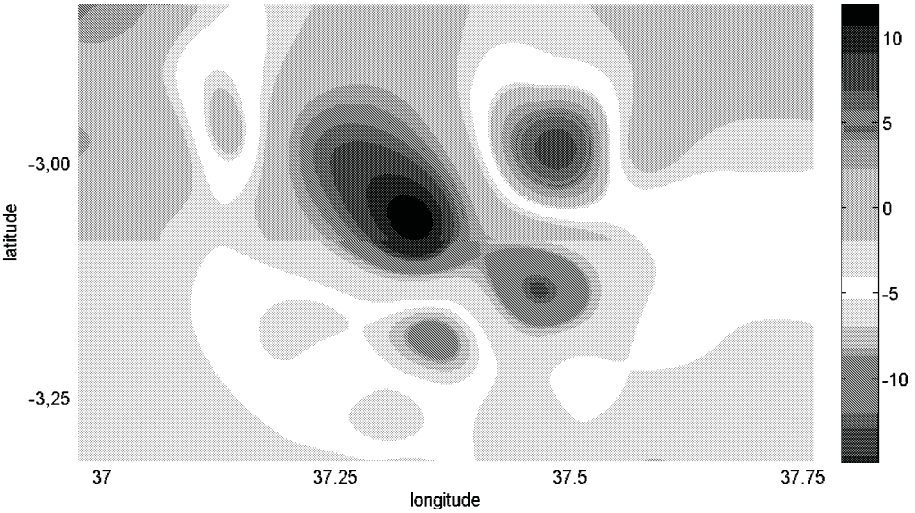


Figure 6: Kilimanjaro uncertainty spatial distribution.

ISBN: 978-989-95079-6-8

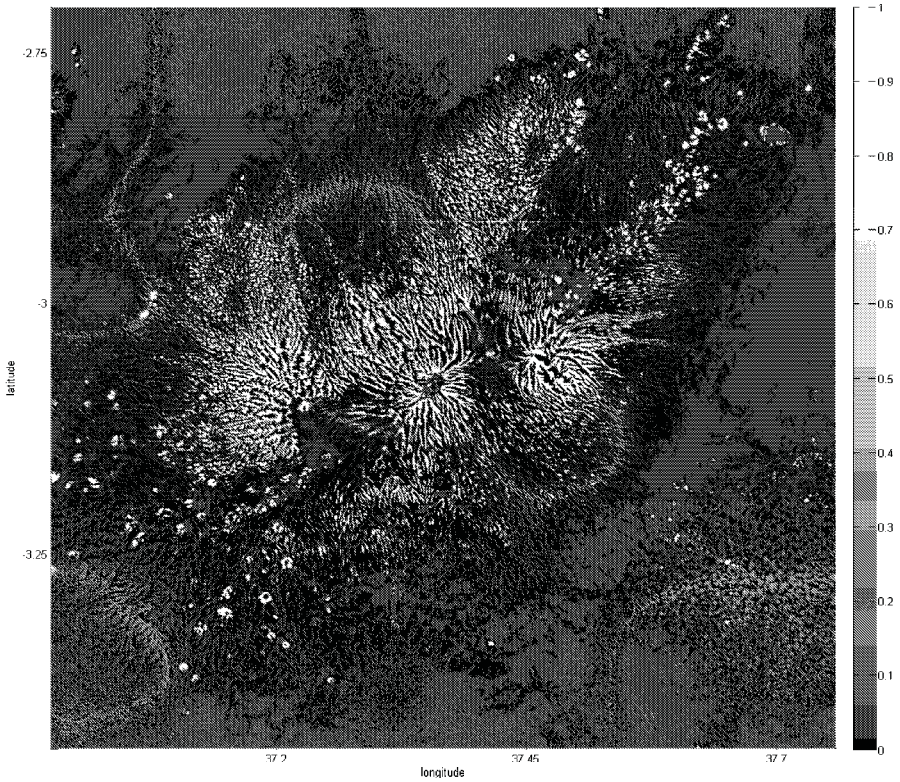


Figure 7: Fuzzy ridge membership for Kilimanjaro region.

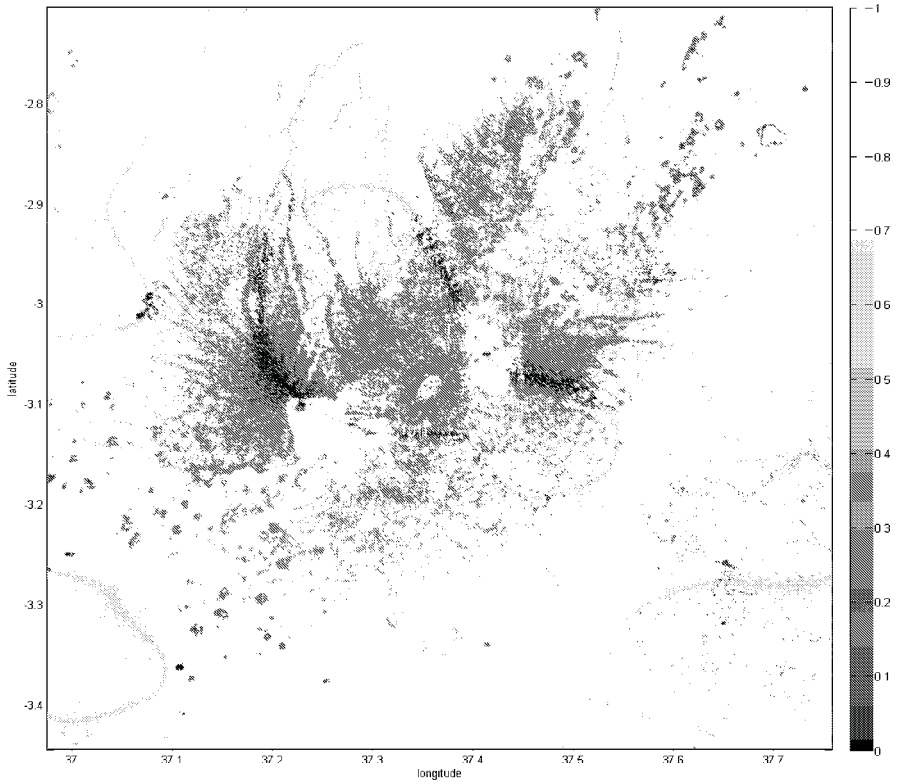


Figure 8: Fuzzy plane membership for Kilimanjaro region.

ISBN: 978-989-95079-6-8

Acknowledgments

Author thanks to the rest of *Kili 2008 Expedition* team (http://kili2008-expedition.blogspot.com/2008/09/ complete-list-of-participants.html) the hard work to provide GPS control points for DEM quality assessment of Kilimanjaro mountain and to Dora Santos, who helped to compile cartographic information of Coimbra.

References

[1] A. M. Anile, B. Falcidieno, G. Gallo, M. Spagnuolo, and S. Spinello. Modeling uncertain data with fuzzy b-splines. *Fuzzy Sets and Systems*, 113(3):397–410, 2000.

[2] W. Lodwick and J. Santos. Constructing consistent fuzzy surfaces from fuzzy data. *Fuzzy Sets and Systems*, 135(2):259–277, 2003.

[3] W. Lodwick, editor. *Fuzzy Surfaces in GIS and Geographical Analysis: Theory, Analytical Methods, Algorithms and Applications*. CRC Press, New York, 2008.

[4] J. Wood. *The geomorphological characterisation of digital elevation models*. PhD thesis, University of Leicester, UK, 1996.

[5] J. Santos. *Interpolação Difusa na Modelação de Fenómenos Geográficos*. PhD thesis, University of Coimbra, Portugal, 2004.

[6] G. J. Klir and B. Yuan. *Fuzzy Sets and Fuzzy Logic, Theory and Applications*. Prentice Hall Inc., New Jersey, 1995.

[7] P. Fisher, J. Wood, and T. Cheng. Where is helvellyn? fuzziness of multi-scale landscape morphometry. *Transactions of the Institute of British Geographers*, 29:106–128, 2004.

[8] P. Fisher, J. Wood, and T. Cheng. *Fuzzy Modelling with Spatial Information for Geographic Problems*, chapter Fuzziness and ambiguity in multi-scale analysis of landscape morphometry, pages 209–232. Springer, New York, 2005.

[9] P. Fisher, J. Wood, and T. Cheng. Higher order vagueness in geographical information: Empirical geographical population of type n fuzzy sets. *Geoinformatica*, 11:311–330, 2007.

[10] Y. Deng and J. P. Wilson. Multi-scale and multi-criteria mapping of mountain peaks as fuzzy entities. *International Journal of Geographical Information Science*, 22:205–218, 2007.

[11] P. A. Burrough, P. F. M. Van Gaans, and R. Hootsmans. High-resolution landform classification using fuzzy k-means. *Fuzzy Sets and Systems*, 113:37–52, 2000.

ISBN: 978-989-95079-6-8

Modeling of Polyester Dyeing Using an Evolutionary Fuzzy System

Maryam Nasiri[1] Stefan Berlik[2]

1 Institute of Knowledge Based Systems and Knowledge Management, Siegen University
Siegen,Germany
2 Software Engineering Institute, Siegen University
Siegen, Germany

Email: maryam.nasiri@uni-siegen.de, berlik@informatik.uni-siegen.de

Abstract*—The aim of this study is to apply and compare statistical regression and an evolutionary fuzzy system to model color yield in the polyester high temperature (HT) dyeing as a function of disperse dyes concentration, temperature and time. The predictive power of the obtained models was evaluated by means of MSE value. It seems that for modeling cases such as the one considered in this study, the evolutionary fuzzy system with a minimum MSE showed a better predictive capability than the model based on statistical regression.*

Keywords— Evolutionary Algorithm, Fuzzy Inference System, Polyester Dyeing, Modelling.

1 Introduction

After introducing and developing soft computing methods many attempts have been made to use and apply them in textile research. In this work, we describe two modeling methodologies, namely; statistical regression and evolutionary-fuzzy systems and compare their capability for modeling of color yield in polyester dyeing. So, a summary of the evolutionary fuzzy approach and a brief note on polyester dyeing are described followed by literature review in use of soft computing in textile engineering.

1.1 Evolutionary Fuzzy System

Many real world problems are so complex that classical computing methods fail to deal with them efficiently. An alternative approach to deal with these problems is soft-computing. The Term soft computing refers to a family of computing techniques comprising four different partners: fuzzy logic, evolutionary computation, neural network and probabilistic reasoning. The term soft computing distinguishes these techniques from hard computing that is considered less flexible and computationally demanding. From this set of techniques, Fuzzy Inference System (FIS) are a powerful tool for modeling real control systems [1,2].
After the initial fuzzy inference system has been set up, the parameters of its fuzzy sets have been optimized in a subsequent step using the covariance matrix adaptation evolution strategy (CMA-ES) [3,4]. CMA-ES is an advanced evolution strategy that relies on an enhanced update mechanism for the mutation distribution's covariance matrix. A survey about genetic-fuzzy systems and its application can be found in [2].

1.2 Polyester Dyeing

Polyester is the most important man made fiber, which is produced by melt spinning process [5]. The dyeing of polyester is limited to only disperse dyes [6] and requires special conditions such as high temperature (≈ 130°C), dry heat (190-220°C), or using carrier in the dye bath [7]. The chemical structure of disperse dyes contains polar groups but there are no ionic groups present which leads to their very low solubility in water. The three main chemical structures of disperse dyes are azo, anthraquinone, and nitro diphenylamine [8]. Temperature, time and disperse dye concentration are the primary factors affecting the color yield in dyeing polyester. The relative importance of these factors can be seen in models representing the color yield as a function of them. These models may also have application in processing and cost minimization. The color yield is shown by K/S. K/S shows the ratio of the absorbed light by an opaque substrate relative to the scattered light from it. This ratio is calculated by Kubelka-Munk theory as [1,9]:

$$(K/S)_\lambda = \frac{(1-R_\lambda)^2}{2R_\lambda} \quad (1)$$

1.3 Literature Review

The research related to the subject of this work can be summarized as follows:
Kim et al. studied Fuzzy modeling, control and optimization of textile processes [10]. Soft computing methods in textile sciences has been reviewed by Sztandera and Pastore [11]. Zarandi et al. presented a fuzzy expert system for textile manufacturing system using fuzzy cluster analysis [12]. FIS has been applied by Hung and to control continuous dyeing [13]. Rautenberg et al. used fuzzy sets for color recipe specification in the textile print shop [14]. Jahmeerbacusa et al. studied fuzzy dye bath pH control in exhaust dyeing [15]. Fuzzy- based simulation model for declorization of industrial waste water has been presented by Abdou et al. [16] Marjoniemi and Mantysalo applied Adaptive Neuro Fuzzy Inference Systems (ANFIS) in modeling dye solution and concentration [17,18]. Active Tension Control of High Speed Splitting Machines using Fuzzy PID has been studied by Chung et al. [19]. Tavanai et al. used fuzzy regression method to model the color yield in dyeing [20]. Smith et al.

ISBN: 978-989-95079-6-8

studied improving computer control of batch dyeing operations [21]. Kim et al. proposed an emotion-based textile indexing system using colors, texture and patterns. Their system utilizes both fuzzy rules and neural networks [22]. A genetic-fuzzy approach has been applied by Nasiri et al. to model polyester dyeing [1]. Callhof and Wulfhort applied ANFIS [23], Veit et al. employed Neural Networks [24], and Nasiri used fuzzy regression in texturing [25]. Guifen et al. predicted the warp breakage rate in weaving using neural network techniques [26]. Predicting the properties of melt spun fibers using neural network has been studied by Chung-Feng [27]. Jeng-Jong applied an evolutionary algorithm to obtain the best combination of weaving parameters for woven fabric designs [28]. Sette et al. used soft computing techniques to Fiber-to-Yarn production process [29]. Peeva et al. examined fuzzy relational calculus theory with applications in various engineering subjects such as textile [30].Blaga studied the application of evolutionary algorithms in knitting technology [31]. An automatic textile sales forecast using fuzzy treatment of explanatory variables was used by Thomassey et al. [32]. Wong et al. studied genetic optimization of JIT operation schedules for fabric-cutting process in apparel manufacture [33]. Siddaiah applied automation in cotton ginning [34]. Aggregation as similarity in a morphological framework for the processing of textile images has been studied by Soria-Frisch [35].

2 Modeling of Polyester Dyeing

The aim of this study is to model variations of color yield of C.I. Disperse Blue 266 versus time, temperature, and disperse dye concentration in the high temperature (HT) polyester dyeing process using an evolutionary fuzzy system and the statistical regression method. A total number of 120 polyester samples were dyed according to the conditions in Table 1. The models based on statistical regression and the evolutionary fuzzy system was developed and compared [1].

Table 1: Dyeing conditions [1]

Dye Concentration (%owf)	0.75 1.5 3 4.5 6
Temperature (°C)	100 110 115 120 125 130
Time (min)	12 24 36 48

2.1 Modeling by Statistical Regression

Modeling of a dependent variable as a function of one or more independent variable(s) can be carried out by means of regression. A general multiple regression for modeling the color yield can be considered to take the following form:

$$Y = A_0 + A_1 x_1 + A_2 x_2 + A_3 x_3 \qquad (2)$$

After using the least squares method to obtain the coefficients in the above equation, the statistical regression model is obtained as fallows:

K/S=-79.4+1.54Conc.+0.11Time+0.71Temp.

To verify the necessary statistical regression conditions, the following four conditions can be used [1,36]:

1) Linear form for the normal plot of the residuals.
2) I chart of residuals should lie between the upper and lower control limits without any specific pattern.
3) The residuals histogram should be of an approximately normal form.
4) Residuals versus fitted values should show no specific pattern.

Fig. 1 shows the information related to the statistical regression model obtained for C.I. Disperse Blue 266 [1]. Regarding the above mentioned four conditions for the validity of the models, it can be said that the linear statistical model cannot be accepted [1,36].

Obviously, the model discussed above with its just four degrees of freedom is of limited capability and there are more powerful e.g. nonlinear regression models. However, these tend to be quite complicated and that is why in the following an evolutionary fuzzy approach will be considered.

2.2 Modeling by Evolutionary Fuzzy System

Variation of K/S function versus two variables of temperature, time and concentration for dyed samples in C.I. Disperse Blue 266 as curves and surfs in some figures like Fig. 2 are shown to study the effect and behaviour of each variable. In view of this figure, for example, it is clearly observed that temperature has a greater effect on K/S value than time has.

In the same way, the FIS model has been investigated along the following lines to model the colour yield of C.I. Disperse Blue 266 in polyester high temperature dyeing [1,37,38].

First, membership functions for input and output variables according to Table 2 and Table 3 have been determined. The Gaussian membership function, and Mamdani max-min Inference was used for all input and output variables [1].

Table 2: Parameters of fuzzy set for input variables [1].

Fuzzy set	Concentration		Time		Temperature	
	Mean	Std	Mean	Std	Mean	Std
Low	1.28	0.848	13.1	14.2	100	9.61
Medium	3.38	0.891	-	-	119	1.64
High	5.87	1.33	44.4	11.7	129	3.97

Table 3: Parameters of fuzzy set for output variable [1].

Fuzzy set	Color yield (K/S)	
	Mean	Std
Very low	3.29	0.62
Low	4.11	1.42
Medium	12.8	1.69
High	19.3	1.86
Very high	29.8	2.23

Secondly, the following nine rules were defined according to the physical and chemical structure of polyester fiber, HT dyeing of polyester, and the behavior of 120 samples dyed in C.I. Disperse Blue 266 [1].

1) If (temperature is low) and (time is low) and (concentration is low), then (K/S is very low).
2) If (temperature is medium) and (concentration is high), then (K/S is high).

ISBN: 978-989-95079-6-8

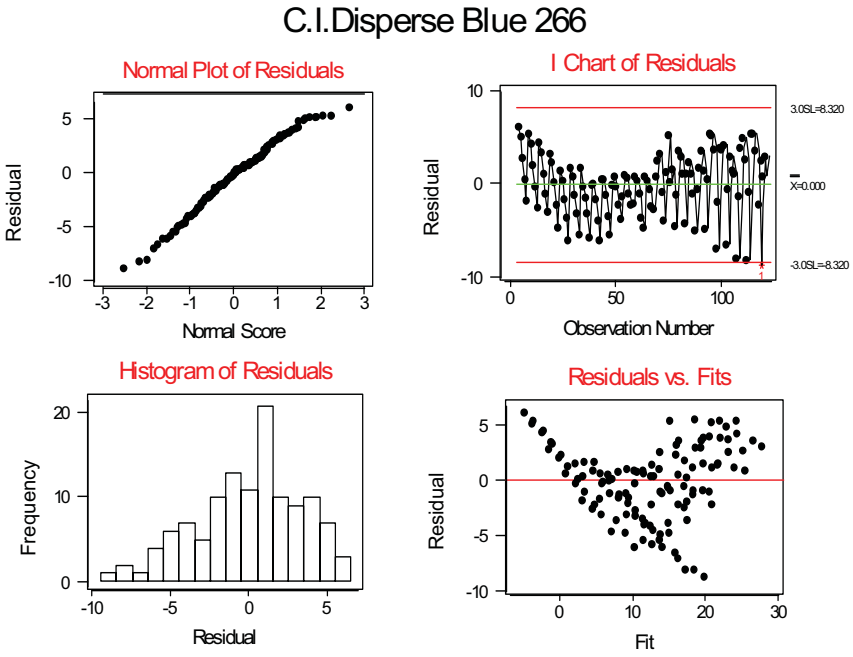


Figure 1: Plot of Residuals for K/S of C.I. Disperse Blue 266 [1].

3) If (temperature is high) and (concentration is low), then (K/S is medium).
4) If (temperature is high) and (concentration is medium), then (K/S is high).
5) If (temperature is low) and (time is high) and (concentration is low), then (K/S is very low).
6) If (temperature is high) and (concentration is high), then (K/S is very high)
7) If (temperature is medium) and (time is low) and (concentration is high), then (K/S is medium).
8) If (temperature is medium) and (time is high) and (concentration is high), then (K/S is high).
9) If (temperature is low) and (time is low) and (concentration is high), then K/S is low.

Then, after a centroid defuzzification of predicted K/S by FIS a Mean Square Errors (MSE) of 3.307 for the FIS model has been obtained [1].

2.3 Optimization of the Initial Fuzzy System

In a final step the fuzzy sets were tuned using the evolutionary algorithm. Hence, input of the CMA-ES algorithm is a vector of parameters of the predefined fuzzy inference system, in this setting the 16 values as given in Table 2. The CMA-ES was used without any modification, i.e. according to the 16 dimensions of the problem, 6 parents breeding 12 offspring were used as recommended. Also the build-in recombination and mutation operators have been applied. The initial step size of the optimization algorithm, also known as mutation strength, was set to 0.1. During the evolutionary loop the fuzzy sets' parameters are successively modified by the CMA-ES and in turn the squared error of the resulting fuzzy inference system is computed for the sample set. The optimization ends after 1000 generations.

3 Results and Interpretation

The FIS model has been improved along the following lines to model the colour yield of C.I. Disperse Blue 266 in polyester high temperature dyeing using an evolutionary algorithm. Table 4 shows parameters of improved fuzzy sets for input variables. As it can be seen in Table 4, concentration has two fuzzy sets in this new model. In the same way, the effect of rule 4 has been tested. Because of its low influence on the results, this rule can be neglected.

Table 4: Parameters of fuzzy sets for input variables.

Fuzzy set	Concentration		Time		Temperature	
	Mean	Std	Mean	Std	Mean	Std
Low	0.53	0.60	12.3	14.21	104.5	9.76
Medium	-	-	-	-	119.2	1.69
High	5.05	1.64	44.1	12.1	129.5	6.10

Regarding the remaining eight rules and changed input parameters according to Table 4, a Mean Square Error (MSE) of 2.333 for evolutionary fuzzy system has been obtained.

Fig. 3 shows the results of the FIS system applied to dyed samples in C.I. Disperse Blue 266.

ISBN: 978-989-95079-6-8

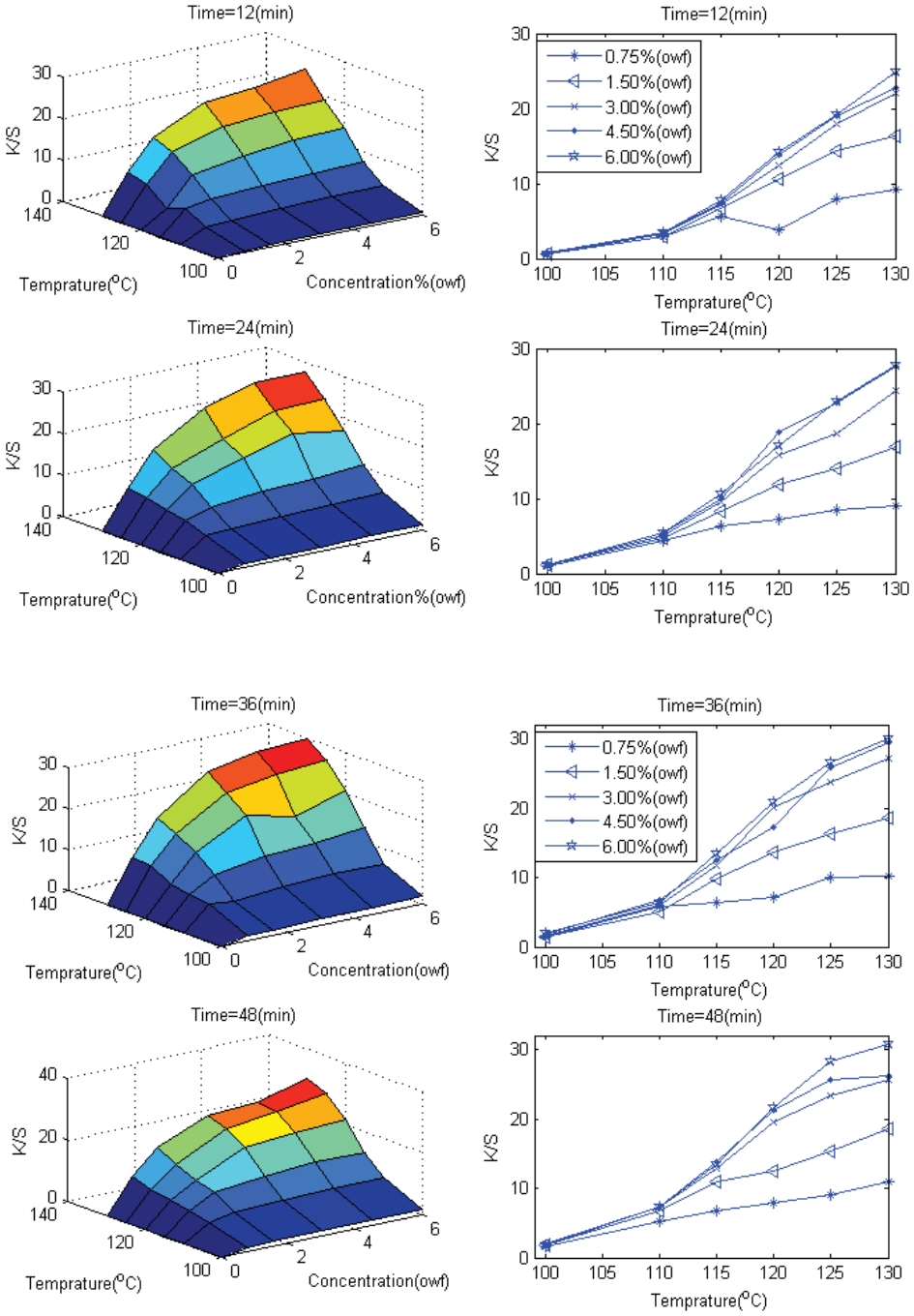


Figure 2: K/S in terms of temperature and concentration [1].

ISBN: 978-989-95079-6-8

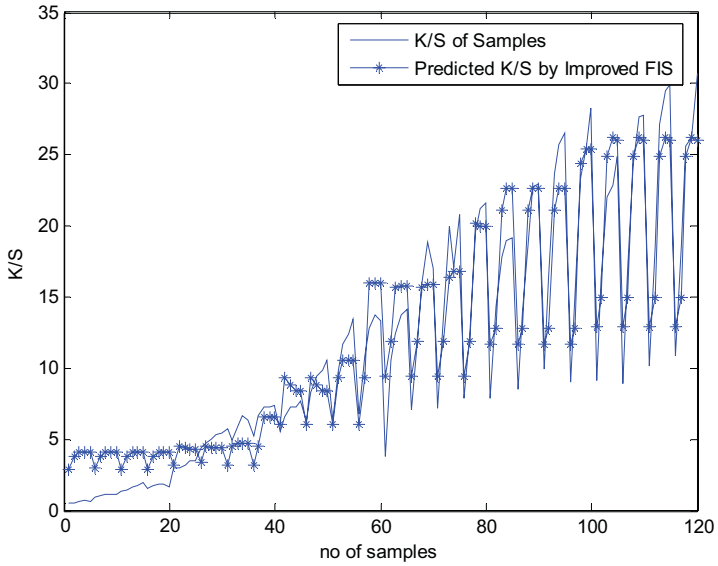


Figure 3: K/S of dyed polyester and predicted K/S by improved FIS.

3.1 Comparison of Statistical Regression and Evolutionary Fuzzy System

Table 5 shows a comparison of the predictive powers by statistical regression, the fuzzy inference system and the evolutionary fuzzy system using MSE. According to this table, the evolutionary fuzzy with minimal MSE and simpler structure showed the best predictive capability in comparison to statistical regression and the fuzzy inference system.

Table 5: Parameters of fuzzy set for output variables.

Method	MSE
Statistical Regression	3.382
Fuzzy Inference System	3.307
Evolutionary Fuzzy System	2.333

On the basis of the above considerations, it can be said that the evolutionary fuzzy system provides an appropriate method to predict the color yield.

4 Conclusions

This research employed statistical regression and an evolutionary fuzzy system to model the (HT) polyester dyeing process. Color yield has been predicted in terms of time, temperature, and disperse dye concentration. We improved the fuzzy sets and rules of the FIS model using an evolutionary algorithm. The results show that the prediction performance is best for the evolutionary fuzzy system.

Acknowledgment

The authors thanks Prof. Hossein Tavanai for his supervision during the first author's work on her MS dissertation when he lent his full support and knowledge to obtain the data used here and to study the behavior of the dyes and also Dr. S. Mahmoud Taheri for his helpful suggestions and assistance.

References

[1] M. Nasiri, S.M. Taheri and H. Tarkesh, Applying Genetic-Fuzzy Approach to Model Polyester Dyeing, Analysis *and Design of Intelligent Systems Using Soft Computing Techniques*, 41: 608-617, 2007.

[2] O. Cordon, F. Herrera, F. Hoffmann and L. Magdalena, Genetic-Fuzzy Systems, Evolutionary *Tuning and Learning of Fuzzy Knowledge Bases*, Advances in Fuzzy Systems - Application and Theory, 19, 2001.

[3] N. Hansen and A. Ostermeier, Adapting Arbitrary Normal Mutation Distributions in Evolution Strategies: The Covariance Matrix Adaptation. IEEE *International Conference on Evolutionary Computation*, 312–317, 1996.

[4] N. Hansen, The CMA Evolution Strategy: A tutorial, 2005. *URL www.bionik.tu-berlin.de/user/niko/cmatutorial.pdf*.

[5] J. Goorden Cook, HandBook *of Textile Fibers, Man Made Fiber*, Merro Technical Library, 1984.

[6] E. R. Trotman, Dyeing *and Chemical Technology of Textile Fibers*; Charles Griffin, 1970.

[7] R.W. Moncrieff , Man-*Made Fibers*, Newnes-Butterworths, 1971.

[8] A. Johnon, The *Theory of Coloration of Textiles*, Second Ed., Bardford: S.D.C. Pub., 1989.

[9] E. Allen, Optical Radiation *Measurements, 2 Color Measurement*, Academic Press, New York, 1980.

[10] S. Kim, A. Kumar, J.L. Dorrity, and G. Vachtsevanos, Fuzzy modeling, control and optimization of textile processes, *Proc 1 Int Jt Conf NAFIPS IFIS NASA*, 32-38,1994.

[11] L.M. Sztandera and C. Pastore, Soft *Computing in Textile Sciences*, Springer, Berlin, 2003.

[12] Mohammad H. Faze1 Zarandi and M. Esmaeilian, A Systematic Fuzzy Modeling for Scheduling of Textile Manufacturing System, *Fuzzy Information Processing Society, NAFIPS 2003. 22nd International Conference of the North American*, 359-364, 2003.

[13] C.C. Hung and W.H. Yu, Control of Dye Concentration, PH, and Temperature in Dyeing Processes, *Textile Research Journal* 69: 914-918, 1999.

ISBN: 978-989-95079-6-8

[14] S. Rautenberg, D. Fanderuff, R. Pacheco, A. Martins, Jose L. Todesco, Richardo M. Barcia, The Use of Fuzzy Sets for Color Recipe Specification in the Textile Print Shop, *Fuzzy Systems Proceedings, IEEE World Congress on Computational Intelligence.,* 885-890, 1998.

[15] M.I. Jahmeerbacusa, N. Kistamahb and R.B. Ramgulamb, Fuzzy Control of Dyebath PH in Exhaust Dyeing, *Coloration Technology*, 120:51-55, 2004.

[16] Abuda etal., Fuzzy- Based Simulation Model for Declorization of Industrial Waste Water, *Journal of Applied Sciences Research*, 4(2): 178-187, 2008

[17] M. Marjoniemi, E. Mantysalo, Neuro-Fuzzy Modeling of Spectroscopic Data. Part A- Modelling of Dye Solutions, *J.S.D.C 113*, 13-17, 1997.

[18] M. Marjoniemi, E. Mantysalo, Neuro-Fuzzy Modeling of Spectroscopic Data. Part B– Dye Concentration Prediction, *J.S.D.C 113*, 64-67, 1997.

[19] Byeong M. Chung, Sung G. Lee, and Che- Seung Cho, Active Tension Control of High Speed Splitting Machines using Fuzzy PID, *IEEE International Conference on Mechatronics*, Taipei, Taiwan, 2005.

[20] H. Tavanai, S.M. Taheri, and M. Nasiri, Modeling of Colour Yield in Polyethylene Terphethalate Dyeing with Statistical and Fuzzy Regression, *Iranian Polymer Journal,* 14: 954-968, 2005.

[21] B. Smith, J. Lu, Improving Computer Control of Batch Dyeing Operations, *American Dyestuff Reporter*, 82(9): 11, 1993.

[22] Soo J. Kim, Eun Y. Kim, K. Jeong and Jee I. Kim, Emotion-Based Textile Indexing Using Colors, Texture and Patterns, *Advances in Visual Computing,* 4292:9-18, 2006.

[23] C. Callhof, B. Wulfhort: Neuro-fuzzy Networks- A Tool For the Development of New Yarn Contacting Exemplary on Friction Discs, *Man Made Fiber Yarn Book* , 95, 1999.

[24] D. Veit, P. Batista-de –Souza, B. Wulfhorst, Application of a Neural Network in the False Twist Texturing Process, *Chemical Fibers International,* 48: 155-156, 1998.

[25] M. Nasiri, Fuzzy Regression Modeling of False-Twist Texturing Yarn, *IFSA 2005 Conference, Beijing, China*, 142-146, 2005.

[26] Y. Guifen, G. Jiansheng, Z. Yongyuan, Predicting the Warp Breakage Rate in Weaving by Neural Network Technicqes, *Textile Research Journal*, 75: 274-278, 2005.

[27] K.J. Chung-Feng, H. Kun-Iuan, W. Yi-Shiuan, Using Neural Network Theory to Predict the Properties of Melt Spun Fibers, *Textile Research Journal*, 74: 840-843, 2004.

[28] L. Jeng-Jong, A Genetic Algorithm for Searching Weaving Parameters for Woven Fabrics, *Textile research journal,* 73: 105-112, 2003.

[29] S. Sette, L. Boullart, Van Langenhof, L.: Building a Rule Set for the Fiber-to-Yarn Production Process by Means of Soft Computing Techniques, Textile Research Journal 70 (2000 Treatment of Explanatory Variables), *European Journal of Operational Research,* 16 : 275-284, 2005.

[30] K. Peeva K., Y. Kyosev, Fuzzy Relational Calculus, *Advanced in Fuzzy Systems- Application and Theory,* 22: 375-386, 2005

[31] M. Blaga, M. Draghici, Application of Genetic Algorithms in Knitting Technology, *The Journal of the Textile Institute*, 97: 175-178, 2004.

[32] S. Thomassey, M. Happiette, J.M. Castelain, An Automatic Textile Sales Forecast Using Fuzzy Treatment of Explanatory Variables, *European Journal of Operational Research,* 16: 275-284, 2005.

[33] W. K. Wong, C.K. Kwong, P.Y. Mok and W.H. Ip, Genetic Optimization of JIT Operation Schedules for Fabric-Cutting Process in Apparel Manufacture., *Journal of Intelligent Manufacturing*, 17(3): 341, 2006.

[34] M. Siddaiah, Michael A. Lieberman, Nadipuram R. Prasad, S.E. Hughs, Automationin Cotton Ginning, *International Journal of Intelligent Systems*, 19 (1-2): 111-129, 2004.

[35] A. Soria-Frisch, M. Koppen, B. Nickolay, Aggregation as Similarity in a Morphological Framework for the Processing of Textile Images *IEEE International Conference on Fuzzy Systems*, 3: 1739-1744, 2004.

[36] D.C. Montgomery, C.A. Peek, Introduction *to Linear Regression Analysis*, Second Ed., J. Wiley, 1991.

[37] H.J. Zimmerman, Fuzzy *Set Theory and Its Applications*, Third Ed. Kluwer Academic Publishers, 1998.

[38] W. Pedrycz, F. Gomidee, An *Introduction to Fuzzy Sets, Analysis and Design*, Prentice-Hall, 1998.

ISBN: 978-989-95079-6-8

Fuzzy concept lattice is made by proto-fuzzy concepts.

Ondrej Krídlo, Stanislav Krajči

Insitute of Computer Science, University of P.J.Šafárik in Košice
Email: ondrej.kridlo@upjs.sk, stanislav.krajci@upjs.sk

Abstract *— An L-fuzzy context is a triple consisting of a set of objects, a set of attributes and an L-fuzzy binary relation between them. An l-cut is a classical context over the same sets with relation as a set of all object attribute pairs, which fuzzy relation assigns truth degree geater or equal than l. Proto-fuzzy concept is a triple made of a set of objects and a set of attributes, which form a concept in some cut of L-fuzzy context and a supremum of all degrees in which cuts this concept exists. Aim of the paper is to show the connection of the structure of proto-fuzzy concepts and fuzzy concept lattice constructed in the way of [1][5]. This connection can help to generate of all fuzzy concepts.*

Keywords — formal concept analysis, fuzzy concept lattice, fuzzy Galois connection

1 Preliminaries

Basic notions of Formal Concept Analysis(FCA) are *formal context* and *formal concept*.

Definition 1 *A formal context $\langle B, A, R\rangle$ consists of a set of objects B, a set of attributes A and a relation R between B and A.*

Definition 2 *Define the mappings $\uparrow\colon {}^B2 \to {}^A2$ and $\downarrow\colon {}^A2 \to {}^B2$. The first assigns to the set $X \subseteq B$ the set of all attributes common to all objects of the set X*

$$\uparrow(X) = \big\{a \in A : \big(\forall o \in X\big)(o,a) \in R\big\}$$

and the second assigns to the set $Y \subseteq A$ the set of all objects common to all attributes of the set Y

$$\downarrow(Y) = \big\{o \in B : \big(\forall a \in Y\big)(o,a) \in R\big\}.$$

Definition 3 *A formal concept of the context $\langle B, A, R\rangle$ is a pair $\langle X, Y\rangle$ such that $X \subseteq B$, $Y \subseteq A$, $\uparrow(X) = Y$ and $\downarrow(Y) = X$.*

Ganter and Wille in [3] showed that the pair of mappings $\big(\uparrow, \downarrow\big)$ is a Galois connection and the composite mappings $\uparrow\downarrow\colon {}^B2 \to {}^B2$ and $\downarrow\uparrow\colon {}^A2 \to {}^A2$ are closure operators. Authors prooved an important theorem in FCA well known as The Basic Theorem On Concept Lattices.

Theorem 1 *(The Basic Theorem on Concept Lattices)The Concept Lattice (lattice of concepts with ordering $\langle X_1, Y_1\rangle \leq \langle X_2, Y_2\rangle$ iff $X_1 \subseteq X_2$ iff $Y_1 \supseteq Y_2$) is a complete lattice in which infimum and supremum are given by*

$$\bigwedge_{i\in I}\langle X_i, Y_i\rangle = \Big\langle \bigcap_{i\in I} X_i, \uparrow\downarrow\Big(\bigcup_{i\in I} Y_i\Big)\Big\rangle$$

$$\bigvee_{i\in I}\langle X_i, Y_i\rangle = \Big\langle \downarrow\uparrow\Big(\bigcup_{i\in I} X_i\Big), \bigcap_{i\in I} Y_i\Big\rangle.$$

A complete lattice V is isomorphic to the concept lattice of some context $\langle B, A, R\rangle$ if and only if there are mappings $\beta : B \to V$ and $\alpha : A \to V$, such that $\beta(B)$ is supremum-dense in V and $\alpha(A)$ is infimum-dense in V and $(o, a) \in R$ is equivalent to $\beta(o) \leq \alpha(a)$ for all $o \in B$ and $a \in A$. In particular V is isomorphic to the concept lattice of context $\langle V, V, \leq\rangle$.

Bělohlávek and Krajči in [1, 2, 5, 6] showed that above mentioned basic notions may be generalized by applying the fuzzy logic.

Everybody knows that reality provides situations where many of attributes are rather fuzzy than crisp. Answer of question "Does the object has the attribute?" is rather somewhere in the middle of false (0) and true (1).

Definition 4 An L-fuzzy formal context *is a triple $\langle B, A, r\rangle$ consists of a set of objects B, a set of attributes A and an L-fuzzy binary relation r, i.e. the L-fuzzy subset of $B \times A$ or mapping from $B \times A$ to L, where L is a complete residuated lattice.*

The class of all L-fuzzy sets in X will be denoted by ${}^XL$. If L is complete then the relation $\subseteq$ (defined by $f \subseteq g$ iff $f(x) \leq g(x)$ for all $x \in X$) makes ${}^XL$ into a complete lattice.

Definition 5 *A* complete residuated lattice *is an algebra $\boldsymbol{L} = \langle L, \wedge, \vee, \otimes, \to, 0, 1\rangle$ where*

(a) *$\langle L, \wedge, \vee, 0, 1\rangle$ is a complete lattice with the least element 0 and the greatest element 1,*

(b) *$\langle L, \otimes, 1\rangle$ is a commutative monoid,*

(c) *$\otimes$ and $\to$ satisfy adjointness, i.e.*

$$a \otimes b \leq c \iff a \leq b \to c$$

for each $a, b, c \in L$ ($\leq$ is the lattice ordering).

Definition 6 (Bělohlávek) *A triple $\langle B, A, r\rangle$ is an L-fuzzy context where $r : B \times A \to L$ and L is a complete residuated lattice. Define mappings $\uparrow\colon {}^BL \to {}^AL$ and $\downarrow\colon {}^AL \to {}^BL$ such that for every $f \in {}^BL$ and $g \in {}^AL$*

$$\uparrow(f)(a) = \bigwedge_{o\in B}\big(f(o) \to r(o,a)\big)$$

$$\downarrow(g)(o) = \bigwedge_{a\in A}\big(g(a) \to r(o,a)\big).$$

ISBN: 978-989-95079-6-8

Table 1: Example of L-fuzzy formal context.

	a_1	a_2	a_3	a_4
o_1	0,2	0,8	0,8	1
o_2	1	1	0,8	1
o_3	0,2	0,6	0,4	1
o_4	0,8	0,4	0,2	0
o_5	0,2	0,4	0,4	0,2

Table 2: 1-cut, 0,6-cut, 0,2-cut

	a_1	a_2	a_3	a_4
o_1				•
o_2	•	•		•
o_3				•
o_4				
o_5				

	a_1	a_2	a_3	a_4
o_1		•	•	•
o_2	•	•	•	•
o_3		•		•
o_4	•			
o_5				

	a_1	a_2	a_3	a_4
o_1	•	•	•	•
o_2	•	•	•	•
o_3	•	•	•	•
o_4	•	•	•	
o_5	•	•	•	•

The aim of this paper is to give a new equivalent definition of Belohlavek's mappings, but using so-called protto-fuzzy concepts as basic building units for constructing the fuzzy concepts. In the next section an l-cuts of L-fuzzy contexts will be defined for any truth degree $l \in L$, main properties of their concepts and a relationship of concepts of different cuts will be showed. Finally Proto-fuzzy concepts will be defined. Then an equality of new mappings with Bělohlávek ones will be showed.

2 Cuts of L-fuzzy context

Lets have an example of L-fuzzy context (Table 1). In our example is the lattice of truth degrees is $\langle\{1;0,8;0,6;0,4;0,2;0,\};\leq\rangle$.

Definition 7 *Let $l \in L$ be an arbitrary truth degree. An l-*cut *of some L-fuzzy set $f \in {}^X L$ is a classical set denoted by*

$$f_l = \{x \in X : f(x) \geq l\}.$$

Definition 8 *An l-*cut *of the L-fuzzy formal context $\langle B, A, r\rangle$ for $l \in L$ is the classical context $\langle B, A, r_l\rangle$ where $r_l = \{(o,a) \in B \times A : r(o,a) \geq l\}$.*

Some of cuts of our example are in the tables.

Definition 9 *For every truth value $l \in L$ lets define mappings $\uparrow_l$: ${}^B2 \to {}^A2$ and $\downarrow_l$: ${}^A2 \to {}^B2$. For every object or attribute subset $X \subseteq B$ and $Y \subseteq A$ put*

$$\uparrow_l (X) = \{a \in A : (\forall o \in X) r(o,a) \geq l\}$$

$$\downarrow_l (Y) = \{o \in B : (\forall a \in Y) r(o,a) \geq l\}.$$

Lemma 1 *Let $K \subseteq L$ be an arbitrary subset of truth degrees. Then for every set of objects $X \subseteq B$ and attributes $Y \subseteq A$ holds that*

$$\uparrow_{(\bigvee K)} (X) = \bigcap_{l \in K} \uparrow_l (X)$$

$$\downarrow_{(\bigvee K)} (Y) = \bigcap_{l \in K} \downarrow_l (Y).$$

Proof:

$\boxed{\subseteq}$ Cuts of the L-fuzzy context were defined such that for every $l_1, l_2 \in L$ if $l_1 \leq l_2$ then $r_{l_1} \supseteq r_{l_2}$. Hence for every subset of objects X or subset of attributes Y, $\uparrow_{l_1} (X) \supseteq \uparrow_{l_2} (X)$ and $\downarrow_{l_1} (Y) \supseteq \downarrow_{l_2} (Y)$, which for every $l \in K$ implies $\uparrow_{\bigvee K} (X) \subseteq \uparrow_l (X)$ and $\downarrow_{\bigvee K} (Y) \subseteq \downarrow_l (Y)$. And from above we have $\uparrow_{(\bigvee K)} (X) \subseteq \bigcap_{l \in K} \uparrow_l (X)$ and $\downarrow_{(\bigvee K)} (Y) \subseteq \bigcap_{l \in K} \downarrow_l (Y)$.

$\boxed{\supseteq}$ Let a be an arbitrary attribute from $\bigcap_{l \in K} \uparrow_l (X)$. For all $l \in K$ and for every object $o \in X$, $r(o,a) \geq l$. From the properties of supremum is $r(o,a) \geq \bigvee K$ for all objects $o \in X$. It means that $a \in \uparrow_{(\bigvee K)} (X)$. Hence $\bigcap_{l \in K} \uparrow_l (X) \subseteq \uparrow_{(\bigvee K)} (X)$.

The second part can be prooved dually. □

Lemma 2 *For all $l \in L$ the pair $(\uparrow_l, \downarrow_l)$ forms a Galois connection between the power-set lattices ${}^B2$ and ${}^A2$.*

Now lets define the concept on the l-cut for some truth degree $l \in L$.

Definition 10 *Let $\langle B, A, r\rangle$ be the L-fuzzy context. A pair $\langle X, Y\rangle$ is called an l-*concept *iff*

$$\uparrow_l (X) = Y, \text{ and } \downarrow_l (Y) = X,$$

hence the pair is a concept in a classical context $\langle B, A, r_l\rangle$. The set of all l-concepts will be assigned $\mathrm{C}_l(B, A, r)$, shortly C_l.

2.1 *Relationship of concepts in different cuts*

Lemma 3 *Let $l_1, l_2 \in L$ be an arbitrary truth values, such that $l_2 \leq l_1$. Let $\langle X, Y\rangle \in \mathrm{C}_{l_1}$. Then there exists an interval $\mathcal{I}$ in concept lattice C_{l_2}, such that for every l_2-concept $\langle Z, W\rangle \in \mathcal{I}$ holds that $X \subseteq Z$ and $Y \subseteq W$.*

Proof: From $\langle X, Y\rangle \in \mathrm{C}_{l_1}$ we know that $\uparrow_{l_1} (X) = Y$ and $\downarrow_{l_1} (Y) = X$. So as the greatest element of wanted interval we can use the $\langle \downarrow_{l_2}\uparrow_{l_2} (X), \uparrow_{l_2} (X)\rangle$ and the least one $\langle \downarrow_{l_2} (Y), \uparrow_{l_2}\downarrow_{l_2} (Y)\rangle$. From the fact $r_{l_1} \subseteq r_{l_2}$ we have inclusions

$$Y = \uparrow_{l_1} (X) \subseteq \uparrow_{l_2} (X)$$

$$X = \downarrow_{l_1} (Y) \subseteq \downarrow_{l_2} (Y).$$

From closure property of conclusion of mappings we have

$$\downarrow_{l_2}\uparrow_{l_2} (X) \supseteq X \text{ and } \uparrow_{l_2}\downarrow_{l_2} (Y) \supseteq Y.$$

From $Y = \uparrow_{l_1} (X) \subseteq \uparrow_{l_2} (X)$ we have

$$\downarrow_{l_2} (Y) \supseteq \downarrow_{l_2}\uparrow_{l_2} (X)$$

and from properties of concepts from [3] we know that it is equivalent to

$$\uparrow_{l_2}\downarrow_{l_2} (Y) \subseteq \uparrow_{l_2} (X).$$

□

ISBN: 978-989-95079-6-8

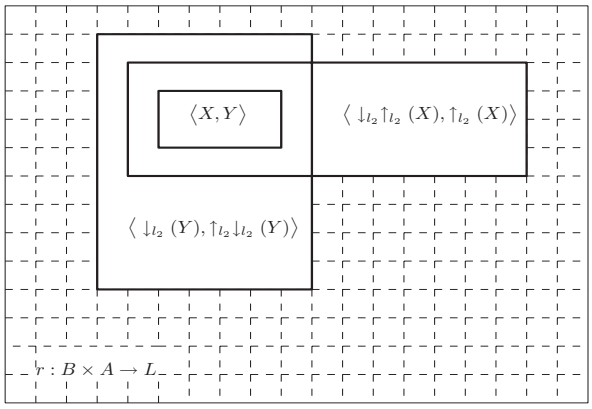


Figure 1: l_1-concept and its l_2-superconcepts

Definition 11 *Define an ordering "$\preceq$" on the set of all l-concepts of all cuts of L-fuzzy context. Lets have two concepts of two different cuts $\langle X, Y\rangle \in \mathrm{C}_{l_1}$ and $\langle Z, W\rangle \in \mathrm{C}_{l_2}$ (because of these pairs are concepts in two different cuts so we can assign them $\langle X, Y\rangle_{l_1}$ and $\langle Z, W\rangle_{l_2}$) then $\langle X, Y\rangle_{l_1} \preceq \langle Z, W\rangle_{l_2}$ iff $X \subseteq Z$ and $Y \subseteq W$ and $l_2 \leq l_1$.*

The lattice in the picture 3 isn't the concept lattice. It is the lattice of all l-concepts for all $l \in L$ and lines is assigning new ordering from definition.

Lemma 4 *Let $l_1, l_2 \in L$ are arbitrary truth values such that $l_2 \leq l_1$. Let $\langle X_1, Y_1\rangle, \langle X_2, Y_2\rangle \in \mathrm{C}_{l_1}$, such that $\langle X_1, Y_1\rangle \preceq \langle X_2, Y_2\rangle$. Than the geatest and least elements of corresponding intervals of l_2-concepts are ordered same as corresponding l_1-concepts.*

Proof: If $X_1 \subseteq X_2$ from properties of closure operator we have $\downarrow_{l_2}\uparrow_{l_2}(X_1) \subseteq \downarrow_{l_2}\uparrow_{l_2}(X_2)$ And from $Y_2 \subseteq Y_1$ we have $\downarrow_{l_2}(Y_1) \subseteq \downarrow_{l_2}(Y_2)$. □

3 Proto-fuzzy concepts

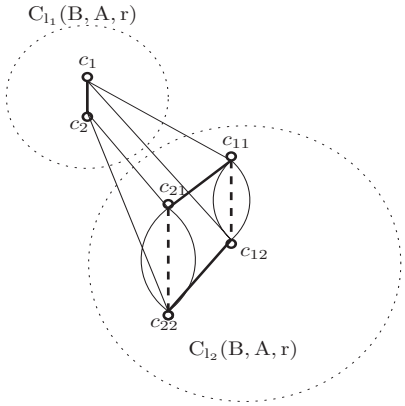


Figure 2: The relationship of concepts of different cuts

As we can see in the figure 3, some of l-concepts are equal, but in different cuts. If we fix some l-concept and look on the set of truth degrees in which cuts the conceot exists, we can see two important properties described in next theorems.

Theorem 2 is sayng that for every concept of some cut, a set of all truth degrees in which cut the concept exists is closed under its supremum.

Theorem 2 *Let $K \subseteq L$ be an arbitrary set of truth degrees, $\langle X, Y\rangle \in \mathrm{C}_l$ for all $l \in K$. Then $\langle X, Y\rangle \in \mathrm{C}_{\bigvee K}$.*

Proof: The lemma 1 implies

$$\uparrow_{\bigvee K}(X) = \bigcap_{l\in K} \uparrow_l (X) = \bigcap_{l\in K} Y = Y,$$

$$\downarrow_{\bigvee K}(Y) = \bigcap_{l\in K} \downarrow_l (Y) = \bigcap_{l\in K} X = X.$$

Hence $\langle X, Y\rangle \in \mathrm{C}_{\bigvee K}$. □

Next theorem 3 is saying that if some concept exist in two different cuts, then exists in every cut between them.

Theorem 3 (Convexity) *Let $l_1, l_2 \in L$ be an arbitrary truth degrees, and let $\langle X, Y\rangle \in \mathrm{C}_{l_1} \cap \mathrm{C}_{l_2}$. Then for all $l \in L$, such that $l_1 \leq l \leq l_2$, $\langle X, Y\rangle \in \mathrm{C}_l$.*

Proof: The lemma 1 implies, that for every set of object X and any two arbitrary truth degrees $k, m \in L$, such that $k \leq m$ holds

$$\uparrow_k (X) \subseteq \uparrow_k (X) \cap \uparrow_m (X) = \uparrow_{\bigvee\{k,m\}} (X) = \uparrow_m (X)$$

and

$$\downarrow_k (Y) \subseteq \downarrow_k (Y) \cap \downarrow_m (Y) = \downarrow_{\bigvee\{k,m\}} (Y) = \downarrow_m (Y).$$

So

$$Y = \uparrow_{l_1} (X) \supseteq \uparrow_l (X) \supseteq \uparrow_{l_2} (X) = Y,$$

$$X = \downarrow_{l_1} (Y) \supseteq \downarrow_l (Y) \supseteq \downarrow_{l_2} (Y) = X.$$

Hence $\uparrow_l (X) = Y$ and $\downarrow_l (Y) = X$, which implies $\langle X, Y\rangle \in \mathrm{C}_l$. □

Definition 12 *Let $\langle X, Y\rangle \in \bigcup_{k\in L} \mathrm{C}_k(B, A, r)$ be the concept of some cut of the fuzzy context $\langle B, A, r\rangle$. Triple $\langle X, Y, l\rangle$ such that $l = \bigvee\{k \in L : \langle X, Y\rangle \in \mathrm{C}_k(B, A, R)\}$ will be called a* proto-fuzzy concept. *The set of all proto-fuzzy concepts will be denoted by* $\mathrm{PFC}(B, A, r)$.

Definition 13 *Define a mapping* $\mathrm{pd} : {}^B2 \times {}^A2 \to L$ *such that for every set of objects $Z \subseteq B$ and set of attributes $W \subseteq A$ is*

$$\mathrm{pd}(Z, W) = \bigvee\{l \in L : (\exists X \subseteq B)(\exists Y \subseteq A)$$

$$\langle X, Y, l\rangle \in \mathrm{PFC}(B, A, r) Z \subseteq XW \subseteq Y\}.$$

This mapping assigns to every pair of sets of objects and attributes the truth degree of highest proto-fuzzy concept which owns them. Notation pd means proto-degree of input sets.

4 Alternative definition of Bělohlávek's mappings

Lets go back to introduction.

Definition 14 *A triple $\langle B, A, r\rangle$ is L-fuzzy context where $r : B \times A \to L$ and L is the complete residuated lattice. Define mappings $\uparrow: {}^BL \to {}^AL$ and $\downarrow: {}^AL \to {}^BL$ such that for every $f \in {}^BL$ and $g \in {}^AL$*

$$\uparrow (f)(a) = \bigwedge_{o\in B} \big(f(o) \to r(o, a)\big)$$

$$\downarrow (g)(o) = \bigwedge_{a\in A} \big(g(a) \to r(o, a)\big).$$

ISBN: 978-989-95079-6-8

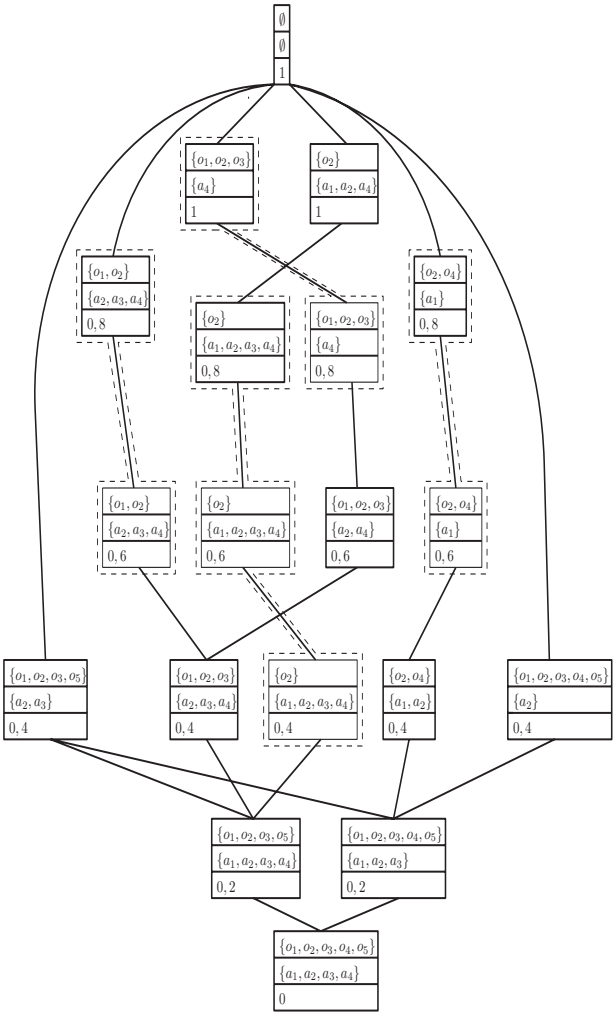


Figure 3: Lattice of all l-concepts of example for all $l \in L$, $\left\langle \cup_{l\in L} \mathbf{C}_l(B, A, r), \preceq \right\rangle$

Lets define new mappings and show that they are equivalent to the mappings above.

Definition 15 *A triple $\langle B, A, r\rangle$ is L-fuzzy context where $r : B \times A \to L$ and L is the complete residuated lattice. Define mappings $\Uparrow$: ${}^B L \to {}^A L$ and $\Downarrow$: ${}^A L \to {}^B L$ such that for every $f \in {}^B L$ and $g \in {}^A L$*

$$\Uparrow(f)(a) = \bigwedge_{l\in rng(f)} \big(l \to \mathrm{pd}(f_l, \{a\})\big)$$

$$\Downarrow(g)(o) = \bigwedge_{l\in rng(g)} \big(l \to \mathrm{pd}(\{o\}, g_l)\big).$$

Theorem 4 *For every set Z of objects, an every set W of attributes,*

$$\bigwedge_{(o,a)\in Z\times W} r(o, a) = \mathrm{pd}(Z, W).$$

Proof:

$\boxed{\leq}$ Let $l = \bigwedge_{(o,a)\in Z\times W} r(o, a)$. If $a \in W$ then $(\forall o \in Z) r(o, a) \geq l$, i.e. $a \in{\uparrow_l}(Z)$, so $W \subseteq{\uparrow_l}(Z)$. Take $Y = {\uparrow_l}(Z)$ and $X = {\downarrow_l}(Y)$ clearly $\langle X, Y\rangle \in \mathbf{C}_l(B, A, r)$. Hence $\langle X, Y, m\rangle \in \mathrm{PFC}(B, A, r)$ where $m = \bigvee\{k \in L : \langle X, Y\rangle \in \mathbf{C}_k(B, A, r)\}$.

Because $W \subseteq{\uparrow_l}(Y)$ and $Z \subseteq{\downarrow_l}{\uparrow_l}(Z) = {\downarrow_l}(Y) = X$, we have

$$\mathrm{pd}(Z, W) \geq m \geq l = \bigwedge_{(o,a)\in Z\times W} r(o, a).$$

$\boxed{\geq}$ Let $\langle X, Y, l\rangle \in \mathrm{PFC}(B, A, r)$, $Z \subseteq X$ and $W \subseteq Y$. Then $l = \bigvee\{k \in L : \langle X, Y\rangle \in \mathbf{C}_k(B, A, r)\}$ and it follows from Theorem 2 that $\langle X, Y\rangle \in \mathbf{C}_l(B, A, r)$. It means that, for all $o \in X$ and $a \in Y$ is $r(o, a) \geq l$
Hence $\bigwedge_{(o,a)\in Z\times W} r(o, a) \geq \bigwedge_{(o,a)\in X\times Y} \geq l$. It follows that $\bigwedge_{(o,a)\in Z\times W} r(o, a) \geq \bigvee\{l \in L : (\exists X \subseteq B)(\exists Y \subseteq A)\langle X, Y, l\rangle \in \mathrm{PFC}(B, A, r), Z \subseteq X, W \subseteq Y\} = \mathrm{pd}(Z, W)$. □

Theorem 5 *For above defined mappings holds*

$$\uparrow = \Uparrow \text{ and } \downarrow = \Downarrow.$$

Proof: Note these three facts:

- Because $\to$ is antitone in the first argument, $f(o) \geq l$ implies
$$f(o) \to r(o, a) \leq l \to r(o, a)$$
- Because $\{o \in B : f(o) \geq l\} \supseteq \{o \in B : f(o) = l\}$, we have
$$\bigwedge_{o\in B: f(o)\geq l} (l \to r(o, a)) \leq \bigwedge_{o\in B: f(o)=l} (l \to r(o, a))$$
- $$\bigcup_{l\in \mathrm{rng}(f)} \{o \in B : f(o) \geq l\} =$$
$$= \bigcup_{l\in \mathrm{rng}(f)} \bigcup_{m\in \mathrm{rng}(f): m\geq l} \{o \in B : f(o) = m\} =$$
$$= \bigcup_{l\in \mathrm{rng}(f)} \{o \in B : f(o) = l\}$$

Using the previous facts we obtain

$$\bigwedge_{l\in \mathrm{rng}(f)} \bigwedge_{o\in B: f(o)=l} (f(o) \to r(o, a)) =$$
$$= \bigwedge_{l\in \mathrm{rng}(f)} \bigwedge_{o\in B: f(o)\geq l} (f(o) \to r(o, a)) \leq$$
$$\leq \bigwedge_{l\in \mathrm{rng}(f)} \bigwedge_{o\in B: f(o)\geq l} (l \to r(o, a)) \leq$$
$$\leq \bigwedge_{l\in \mathrm{rng}(f)} \bigwedge_{o\in B: f(o)=l} (l \to r(o, a)) =$$
$$= \bigwedge_{l\in \mathrm{rng}(f)} \bigwedge_{o\in B: f(o)=l} (f(o) \to r(o, a))$$

It follows that both inequalities are in fact equalities, hence (using Theorem 4)

$$\uparrow(f)(a) = \bigwedge_{o\in B} (f(o) \to r(o, a)) =$$

ISBN: 978-989-95079-6-8

$$= \bigwedge_{l\in \mathrm{rng}(f)} \bigwedge_{o\in B: f(o)=l} (f(o) \to r(o,a)) =$$

$$= \bigwedge_{l\in \mathrm{rng}(f)} \bigwedge_{o\in B: f(o)\geq l} (l \to r(o,a)) =$$

$$= \bigwedge_{l\in \mathrm{rng}(f)} (l \to \bigwedge_{o\in B: f(o)\geq l} r(o,a)) =$$

$$= \bigwedge_{l\in \mathrm{rng}(f)} (l \to \bigwedge_{o\in f_l} r(o,a)) =$$

$$= \bigwedge_{l\in \mathrm{rng}(f)} (l \to \mathrm{pd}(f_l, \{a\})) = \Uparrow (f)(a)$$

The second part can be proved dually. □

4.1 *One-sided fuzzy concepts*

In [8] we defined so-called one-sided fuzzy concepts, the pairs consisting of one classical set of objects and fuzzy set of attributes and defined mappings which are creating them.

Definition 16 *For L-context* $\langle B, A, r\rangle$ *define mappings* $\uparrow\uparrow$: ${}^B2 \to {}^AL$ *and* $\downarrow\downarrow$: ${}^AL \to {}^B2$. *For an arbitrary set of objects* $X \in {}^B2$ *and an L-fuzzy set of attributes* $g \in {}^AL$ *put*

$$\uparrow\uparrow (X)(a) = \bigwedge_{o\in X} r(o,a)$$

$$\downarrow\downarrow (g) = \{o \in B : (\forall a \in A) g(a) \leq r(o,a)\}.$$

By the theorem 4 we can write $\uparrow\uparrow (X)(a) = \mathrm{pd}(X, \{a\})$. Bělohlávek's theorem ([2]) say:

Theorem 6 *For* $X \in {}^B2$ *and* $g \in {}^AL$ *and mappings* $\uparrow, \downarrow, \uparrow\uparrow, \downarrow\downarrow$ *we have*

$$\uparrow\uparrow (X) = \uparrow (X') \text{ and } \downarrow\downarrow (g) = (\downarrow (g))_1$$

and X' *means L-fuzzy set of objects corresponding to* X, *and* $(\downarrow (g))_1$ *means* 1*-cut of* $\downarrow (g)$.

With the theorem 5 we can in previous theorem change the mappings $\uparrow, \downarrow$ by $\Uparrow, \Downarrow$.

5 Sketch of algorithm for generating all fuzzy concepts

```
Set<Object> B;
Set<Attribute> A;
Set<TrueDegree> L;
LFuzzy binary relation r;
Set<PFConcept> PC(B,A,r)
Set<FConcept> fcs;

public Set<FuzzyConcept> generateAllFC(
        Set<Proto-fuzzy concept> PC(B,A,r)
 ){

// generating all basic L-fuzzy concepts

 for ( PFConcept pc : PC(B,A,r) ){
  Set<Object> objs = pc.getObjects();
  Set<Attribute> atrs = pc.getAttributes();
  True Degree deg = pc.getTruthDegree();
   for ( TruthDegree m : L )
    for ( TruthDegree k : L )
     if ( m== k --> deg ){
      //new L-fuzzy set of Objects
      LFSObjs f = new LFSObjs(
        <obj,m> if obj : objs,
        <obj,0> if obj : B-objs
      );
      //new L-fuzzy set of Attributes
      LFSAtrbs g = new LFSAtrbs(
        <atr,k> if atr : atrs,
        <atr,0> if atr : A-atrs
      );
//every fuzzy concept will remeber of which
//of proto fuzzy concepts was created
      fcs.add(
       new FConcept(
                f ,
                g ,
                new Set<PFConcept>{ pfc } )
             );
    }
  }

// creating connected L-fuzzy concepts

 for ( LFuzzyConcept fc1 : fcs )
  for ( LFConcept fc2 : fcs ){
   Set<PFConcept> pfcs1 = fc1.getPFConcepts();
   Set<PFConcept> pfcs2 = fc2.getPFConcepts();
   boolean ordered = true;
   for ( PFConcept pfc1 : pfcs1 )
    for ( PFConcept pfc2 : pfcs2 )
     if ( !pfc1 <= pfc2 AND !pfc2 <= pfc2 )
      ordered = false;
   if ( ordered )
    fcs.add( createNewFC( fc1 , fc2 ) );
  }

  return fcs;
}

public FConcept createNewFConcept(
   FConcept fc1 ,
   FConcept fc2 )
{
//new L-fuzzy set of Objects
LFSObjs f = new LFSObjs(
 <obj,fc1.getDeg(obj)>
   if fc1.getDeg(obj)>=fc2.getDeg(obj),
 <obj,fc2.getDeg(obj)>
   if fc2.getDeg(obj)>=fc1.getDeg(obj)
);
//new L-fuzzy set of Attributes
LFSAtrbs g = new LFSAtrbs(
 <atr,fc1.getDeg(atr)>
   if fc1.getDeg(atr)>=fc2.getDeg(atr),
 <atr,fc2.getDeg(atr)>
   if fc2.getDeg(atr)>=fc1.getDeg(atr)
);
//Union of sets of proto-fuzzy concepts
Set<PFConcept> pfc =
  unite(
     fc1.getPFConcepts,
```

ISBN: 978-989-95079-6-8

```
      fc2.getPFConcepts
        )

return new FConcept( f , g , pfc );
}
```

6 Future work

Our future work will be to finish the sketched algorithm, to proove his good working and apply it.

We are grateful for precious comments of our colleague and friend Jozef Pócs.

Paper was created with support of grant 1/3129/06 Slovak grant agency VEGA.

References

[1] Bělohlávek R. Lattices Generated By Binary Fuzzy Relations Tatra Mountains Mathematical Publications 16 (1999),11-19

[2] Bělohlávek R., Vychodil V. What is fuzzy concept lattice? Proceedings of 3rd international conference on Concept Lattices and Their Applications 2005 (CLA 2005), 34 - 45

[3] Ganter, B. and Wille, R. 1997 Formal Concept Analysis: Mathematical Foundations. 1st. Springer-Verlag New York, Inc.

[4] Krídlo O., Krajči S. Proto-fuzzy concepts, their retrieval and usage Proceedings of Sixth International Conference on Concept Lattices and Their Applications (CLA 2008), 83 - 96

[5] Krajči, S., A Generalized Concept Lattice, Logic Journal of the IGPL, 2005, 543-550

[6] Krajči S., The Basic Theorem on Generalized Concept Lattice, CLA 2004,

[7] Snášel V., Ďuráková D., Krajči S., Vojtáš P.: Merging Concept Lattices of α-cuts Of Fuzzy Contexts Contributions To General Algebra 14 Proceedings of the Olomouc Conference 2002 (AAA 64) and the Postdam Conference 2003 (AAA 65) Verlag Johanes Heyn, Klagenfurt 2004

[8] Krajči, S.: Cluster Based Efficient Generation Of Fuzzy Concepts Neural Network World 5/03 521–530

ISBN: 978-989-95079-6-8

A Preliminary Analysis for Improving Model Structure of Fuzzy Habitat Preference Model for Japanese Medaka (*Oryzias latipes*)

Shinji Fukuda[1]

1. Institute of Tropical Agriculture, Kyushu University
Fukuoka, Japan
Email: shinji-fkd@agr.kyushu-u.ac.jp

Abstract—*The present study examined a preliminary analysis for improving model structure of fuzzy habitat preference model for Japanese medaka (Oryzias latipes) dwelling in agricultural canals in Japan. The present model employed a simplified fuzzy reasoning method for evaluating habitat preference of the fish based on the relationship with physical habitat characteristics observed in the field survey. The model parameter was optimized by using a simple genetic algorithm, in which number of fuzzy membership function was fixed. In the present analysis, number of fuzzy membership function was changed while the other methods were fixed as the original model. The model performance was evaluated based on mean square error between observed and predicted fish population density, and by using two different data sets. As a result, there was no clear tradeoff between number of fuzzy membership functions and prediction accuracy. By contrast, calibration and validation results showed a slight tendency of tradeoff. Further studies on clarifying the tradeoffs would be necessary for improving the model structure in an effective way.*

Keywords— fish habitat, genetic algorithm, habitat modelling, simplified fuzzy reasoning

1 Introduction

Since anthropogenic impacts largely affect and degrade aquatic environments, assessment based on geomorphologic and ecological data is the basis for evaluating the state of the environment and its management planning [1]. In the evaluation, a habitat preference approach, such as habitat evaluation procedure (HEP) [2, 3] and Instream Flow Incremental Methodology (IFIM) [4], is often applied for its simplicity and intuitiveness. Habitat suitability index (HSI), which was originally proposed for HEP and IFIM, is usually classified into three categories: Category I—professional judgements; Category II—habitat use indices; Category III—habitat preference indices [5]. Despite its generality, some studies concluded that HSI cannot describe habitat uses [6, 7, 8, 9]. This failure would be ascribed to non-linear and complex interactions between environmental factors affecting the habitat use and to uncertainty in ecology.

The uncertainty is increasingly recognized and gaining public interests in natural ecosystem research. In ecological modelling, the uncertainty of the model is caused by both the lack of knowledge (i.e. data imperfection) and the variability of models and parameters (models' sensitivity) [10]. For coping with the data imperfection, an approach using fuzzy logic was proposed and widely applied to expressing expert knowledge and dealing with incomplete and/or subjective information [11, 12, 13, 14, 15, 16]. This approach enables qualitative information consisting of linguistic terms to be used for quantitative evaluation of target systems that are required for decision-making at an ecosystem level. By contrast, [17] introduced fuzziness as a reliability bound so as to consider the subjective uncertainty in habitat evaluation by different interpretation. The author also applied fuzzy set theory to habitat evaluation for considering the vagueness of fish behaviour, measurement errors and dispersions of physical environment, in which models showed high ability to represent habitat use of the target fish by integrating fuzzy rule-based model with model optimization techniques of genetic algorithm (GA) [18, 19, 20, 21, 22, 23, 24] or artificial neural network [25, 26]. By applying GA, all the model parameters can be simultaneously optimized, which enables us to assess the nonlinearity in fish habitat preference [23]. This emerging property clarified from the modelling would make an appeal to researchers for applying artificial intelligence techniques. However, as reported in [22], the fuzzy habitat preference model showed weakness in terms of transferability which might be ascribed to the complexity of the models. Therefore, to assess the relationship between models' complexity and accuracy would be necessary for the model improvement without losing its accuracy. This could also promote the application of genetic fuzzy systems in ecological research.

The present study aims to examine a preliminary analysis for improving model structure of fuzzy habitat preference model (FHPM) for Japanese medaka (*Oryzias latipes*) dwelling in agricultural canals in Japan. Field surveys were conducted in the agricultural canals to establish a relationship between fish habitat preference and physical environments of water depth, current velocity, lateral cover ratio and percent vegetation coverage. All the FHPMs of different model structure were compared focusing on the tradeoffs between complexity of the model and accuracy of habitat prediction.

2 Methods

In this section, the author first gives a brief review on the application of fuzzy systems to ecological research. Second, an overview of the study site and target fish is described. And then, the author explains the field surveys and the modelling procedure of habitat prediction models together with the explanation on assessment and comparison of the models.

2.1 Fuzzy Modelling in Ecological Research

The pioneer work of fuzzy logic application to ecological study would be [27] in which fuzzy set theory was introduced to habitat quality evaluation, classification of wetlands, and formation of compartments of ecosystem components. Since

ISBN: 978-989-95079-6-8

then, many researchers have employed fuzzy logic-based approaches in ecological research (e.g., population dynamics [28, 29, 30, 31], cluster analysis [32], elicitation of expert knowledge [33], human decision behaviour [34]). Reference [35] reviewed applications of fuzzy logic for decision support in ecosystem management with focus on the identification, optimization, validation, interpretability and uncertainty aspects of rule-based models. As mentioned earlier, fuzzy modelling has also been applied to habitat assessment. This can be classified into two categories of knowledge-based modelling [12, 13, 14, 15] and data-driven modelling [16, 18, 19, 20, 21, 22, 23, 24, 25, 26, 36]. The former cannot optimize model parameters, while the latter can do it. Several approaches were employed for data-driven optimization of model parameters. For instance, [16] employed a nearest ascent hill climbing algorithm, [22, 25, 26] incorporated fuzzy membership function into artificial neural network, and the others employed GA [18, 19, 20, 21, 22, 23, 24, 36]. According to [37], the latter two approaches can be classified into genetic fuzzy systems. The genetic fuzzy systems would gain more attentions in ecological research because of its interpretability and learning ability.

2.2 Study Site and Target Fish

The survey was carried out in an agricultural canal located in Kurume City, Fukuoka, Japan (33˚20’ N, 130˚42’ E; Fig. 1). The spring-fed canal runs through paddy fields, and is used for both irrigation and drainage purposes. It flows into the Kose River, which is a tributary of the Chikugo River.

The target fish, Japanese medaka (*Oryzias latipes*), is one of the most common freshwater fish in Japan. This fish has been considered as one of the symbols of rural environmental conservation and restoration because of the vulnerability to alterations of physical environment such as concrete lining of the earthen canal. For instance, Japanese medaka generally grows up to approximately 2 cm in length and thus is vulnerable to fast flowing water.

2.3 Field Survey

The field surveys were conducted on 14 October, and 5 and 9 November 2004. The surveys were conducted during a non-irrigation period, thus the discharge in the study reaches remained stable. Therefore, the habitat uses of Japanese medaka were not affected by any agricultural activities or agricultural chemicals. The water temperature remained stable (16.1–20.3˚C) during the surveys.

The habitat uses of Japanese medaka and the physical habitat characteristics of water depth (henceforth referred to as depth), current velocity (velocity), lateral cover ratio (cover), and percent vegetation coverage (vegetation) in the study reach were surveyed on sunny days. The four physical environments are found to be the primary factors affecting spatial distribution, i.e., habitat use, of the fish [20]. After mapping the reach, habitat use of the fish was first observed (11:00–14:30) and then the physical habitat characteristics within the reach were surveyed.

The habitat use of Japanese medaka was observed visually from the bank; the observer moved slowly and carefully to avoid disturbing the fish. The number of the fish was counted in units of five to take into consideration the habit of school formation, i.e., fish in a small school (less than five) were not counted. Observations were repeated eight times and the results were averaged to reduce observational error.

Immediately after completing the fish observation, the four physical habitat characteristics of depth, velocity, cover, and vegetation were surveyed to establish a relationship between physical environment and habitat preferences of Japanese medaka. First, depth and velocity were measured to divide the study reach into small water bodies having similar condition with regard to these two physical parameters. Depth was measured with a stainless steel ruler, and velocity with a portable propeller current meter (KENEK, V-303) at three lateral points comprising a midpoint and two near-shore points at longitudinal intervals of 1 m. By using the measurements of depth and velocity, the reach was divided into water bodies. Next, the other two factors of cover and vegetation were calculated from the schematic diagrams of the water bodies. The lateral cover ratio is defined as a function of the presence of lateral cover, which comprises the water’s edge, a dike, or anything that emerges from the water surface and surrounds the water body. The cover thus consists of four components (four lateral sides). The maximum cover ratio is 100%, and each of the cover components is assigned a score of 25%. In the definition of the cover, objects attached to more than 90% of the boundary between water bodies were regarded as cover components. That is, only instream and undersurface cover structures were considered as cover component because they may have had the same effects as the margin of the stream. Percent vegetation coverage is defined as the percentage of the area covered with aquatic vegetation in each water body. Both submerged and emergent vegetation were pooled because of their same roles in providing food and shelter from predators and fast-flowing currents.



Figure 1: Study reach

The habitat use data used in the following analyses were the observed fish population density obtained for the i^{th} water body $\rho_{o,i}$ (individuals per square metre), where i (i=1, 2, …, n) denotes the index of the water body and n the total number of water bodies.

2.4 Fuzzy Habitat Preference Model

In fuzzy habitat preference model (FHPM), a simplified fuzzy reasoning was introduced to relate physical habitat characteristic to habitat preference with the consideration of the uncertainties (Fig. 2), and a simple GA to optimize the

ISBN: 978-989-95079-6-8

model structure. GA was employed because it enables us to simultaneously optimize all the parameters in FHPM even under the nonlinear, complex interactions between physical environment and fish habitat preference. The premise part of FHPM (Fig. 2(i)), i.e. fuzzy membership functions, is defined for the purpose of reflecting the ecology of Japanese medaka. For instance, because the body length of an adult medaka is approximately 2 cm, the critical requirements for Japanese medaka with regard to depth and velocity thus would be relatively shallower and slowly flowing water. The habitat preference is calculated by taking the weighted mean of singletons in the consequent part (Fig. 2(ii)) by the membership value μ_i. That is,

$$P_j = \frac{\sum_{i=1}^{n_m} \mu_{j,i} \cdot c_{j,i}}{\sum_{i=1}^{n_m} \mu_{j,i}} \quad (1)$$

where P_j denotes the habitat preference to the environmental factor j (=depth, velocity, cover, and vegetation), c_i is the value assigned to each singleton (i.e., degree of habitat preference assigned to the corresponding membership function in the premise part), μ_i is the membership value (i.e., degree of fitness to the membership functions in the premise part), i is the index of the membership functions, and n_m is the number of membership functions (Fig. 2). By taking the weighted mean (1), the results are given in a simple linear form. The singletons in the consequence part are determined by using GA so as to minimize mean square error (MSE) between predicted and observed fish population density. The optimization procedure is summarized as follows.

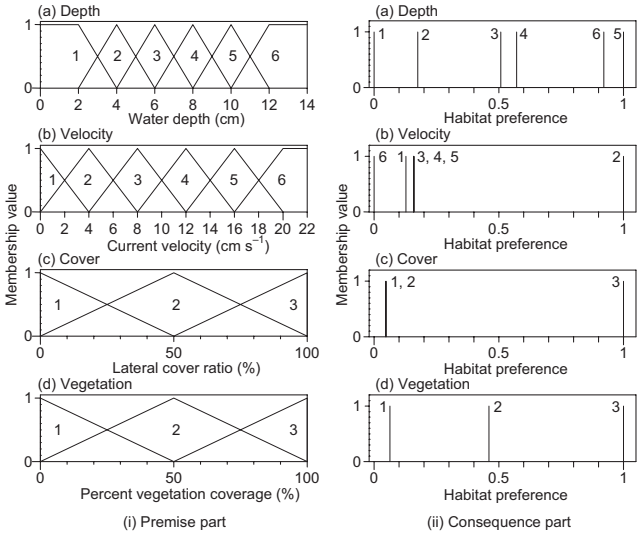


Figure 2: Membership functions of FHPM

First, GA proposed a set of initial model structures (i.e., consequence part) of FHPM. Second, the environmental measurements were given as input values. The input values, i.e., four physical environmental variables, are represented by symmetric triangular fuzzy numbers for considering the uncertainty originated from measurement errors and spatial dispersions. The fuzzy inputs are expressed by its centre a^c_j and spread a^s_j as (a^c_j, a^s_j), where j denotes the environmental factors of depth, velocity, cover, and vegetation. The spread a^s_j is determined from allowable variance when dividing the reach into water bodies: 1 cm is given for the spread of depth a^s_d, 2 cm s^{-1} for velocity a^s_v, 10% each for cover a^s_c and vegetation a^s_{veg}. Of these, the abbreviations of d, v, c, and veg indicate depth, velocity, cover, and vegetation, respectively. Third, habitat preferences for each environmental factor were calculated. Fourth, the habitat preferences to the four environmental factors were combined using (2).

$$P_i = P_{d,i} \times P_{v,i} \times P_{c,i} \times P_{veg,i} \quad (2)$$

Fifth, the habitat use of Japanese medaka was predicted by using (3).

$$\rho_{c,i} = \left(P_i \Big/ \sum_{i=1}^{n} P_i \right) \cdot \sum_{i=1}^{n} \rho_{o,i} \quad (3)$$

Table 1: Condition of GA optimization

Operation	Condition & Remarks
Selection	Elite preservation strategy
Crossover	Uniform
Mutation	0.05%
Number of individuals	100
Number of iteration	5000
Length of binary strings	6 bits per parameter

Table 2: Condition of model structure modification, in which filled circles indicate the factors modified.

Factor	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Depth	●				●	●	●				●	●	●		●
Velocity		●			●			●	●		●	●		●	●
Cover			●			●		●		●	●		●	●	●
Vegetation				●			●		●	●		●	●	●	●
Number*	1	1	1	1	2	2	2	2	2	2	3	3	3	3	4

*Number of membership functions modified.

Next, MSE between the predicted and observed fish population density was calculated, and then GA repeatedly modified the singletons so as to minimize MSE. Finally, the optimized FHPM was obtained. The basic condition of GA optimization is summarised in Table 1. In the original setting, totalled 17 parameters were optimised (5 for depth, 6 for velocity, and 3 each for cover and vegetation). The number of parameters to be modified differed by the condition of model structure modification (see Table 2).

2.5 Model Development and Analysis

In the present study, a data set collected on 14 October and 5 November 2004 was used for calibration while the other data set collected on 9 November was used for validation. The condition given for comparing the results of model structure modification is summarised in Table 2. Totalled 31 sets of the condition were given in which one is given for original model, and 15 each for reducing and adding membership functions to the original model. Since model structures of the models vary in accordance with their initial conditions in the optimization, 20 different initial conditions

ISBN: 978-989-95079-6-8

were thus given to each model so as to evaluate the variance of model structure developed. For the comparison, MSE between predicted and observed fish population densities was calculated. Average and standard deviation of MSE on all the models were then used for the comparison. The relationship between MSE and number of fuzzy membership functions was compared in order to clarify the accuracy-complexity tradeoffs in FHPM.

3 Results

The results of field surveys on 14 October and 5 November 2004 were pooled and used for model calibration (Fig. 3(i)), while the result of survey on 9 November was used for model validation (Fig. 3(ii)). All the models were developed by using raw data shown in Fig. 3.

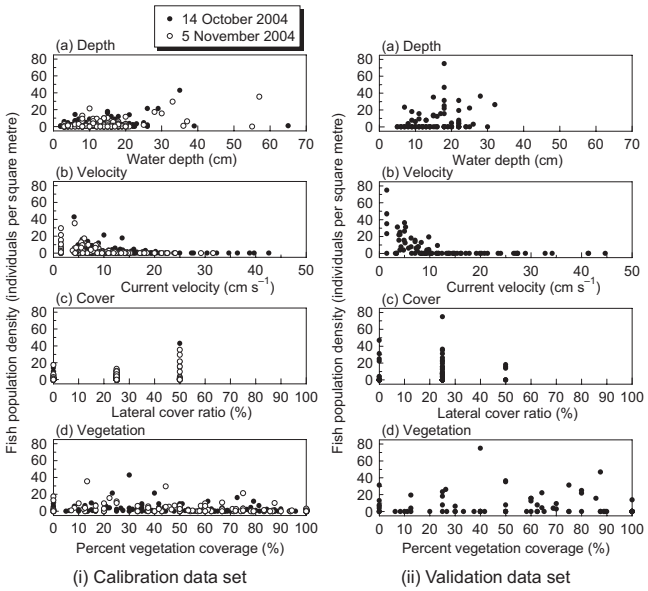


Figure 3: Result of field survey

Habitat preferences evaluated by all the models reflected their model structure, in which the variance in habitat preference curves would be ascribed to the difference in model structure among 20 different initial conditions. Of these, three habitat preference curves evaluated by the models of no modification (original), four-factor reduced and four-factor added are shown as an example (Fig. 4). Despite of the similar tendency in habitat preference, the larger variance in model structure was found in the case of four-factor added model.

By using the habitat preference models together with (2) and (3), habitat uses of Japanese medaka were predicted, of which MSE between predicted and observed fish population density (i.e., habitat use) were used for the comparison. Because some calibration turned out to be failure (especially in the case of adding fuzzy membership function), the best result (i.e., model) achieved at each condition was used in the following analysis. Fig. 5 shows the scatter diagrams of the averaged MSE and number of fuzzy membership functions in both calibration and validation, in which the number of 17 indicates the original model. The model structure could not be improved by modifying the number of fuzzy membership functions in calibration. By contrast, reducing number of fuzzy membership functions resulted in deterioration of prediction ability of the models. The degree of deterioration differed by factors that was modified. As a result, no tradeoff was found between prediction ability of the models (i.e., MSE) and number of fuzzy membership function (i.e., model complexity).

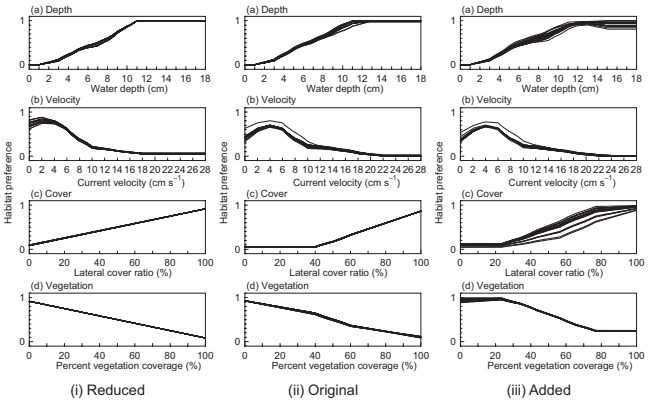


Figure 4: Habitat preference curves evaluated by three models of four-factor reduced (i), no modification (original) (ii), and four-factor added (iii).

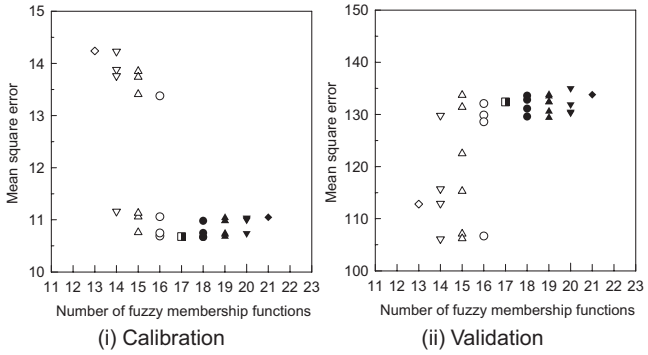


Figure 5: Scatter diagram between minimum mean square error and number of fuzzy membership functions in calibration (i) and validation (ii), in which open marks represent the case of reduced number of fuzzy membership function while filled marks represent that of added.

The MSEs of calibration and validation were compared for the deeper understandings of the relationship between prediction ability and model structure of FHPM. In Fig. 6, there were slight tradeoffs between calibration and validation results. For instance, models with smaller MSE in calibration resulted in larger MSE in validation. Fig. 6 also indicated that no improvement was achieved by adding fuzzy membership functions to the original one.

4 Discussion

The complex interactions between physical environment and habitat preference of the fish made it difficult to clarify the tradeoffs in FHPM. All the habitat preference curves of 31 different conditions showed similar trend of larger water depth, slower velocity, larger cover and smaller vegetation (Fig. 4). However, the prediction result differed between models (Figs. 5 & 6). In Fig. 5, reducing fuzzy membership

ISBN: 978-989-95079-6-8

function resulted in better prediction in the validation data set while prediction in the calibration data set resulted in deterioration. On one hand, this clearly indicates the effect of over-fitting to the data used in model calibration. This could be improved by modifying the conditions of GA optimization such as number of iteration and number of data used in model development. On the other hand, this would be ascribed to the different distribution of the physical environment and habitat use of the fish (Fig. 3). This is one of the major problems in habitat modelling and is often discussed as "transferability" of the habitat models [8, 9]. Adding fuzzy membership function thus could not improve prediction ability of the model. To achieve the model transferability, it is necessary to develop habitat models by using sufficient data which satisfy generality of the habitat use of the fish. In the present case, an approach to achieve generality would be the use of different sub-sets of data in model optimization. For instance, [38] employed different subsets of data to cope with the problems of overtraining and high sensitivity to outliers in the application of artificial neural network model. Reference [16] also applied similar approach in the optimization of fuzzy rule-based habitat model. Another approach would be to modify the fuzzy membership function into less precise definition as often applied in habitat modelling of riverine fish [12, 13, 14, 15, 16]. In this case, the advantage of consideration of critical requirement into model structure would however be lost at the same time. Besides the discussion on complexity and improvement of the model, the present result may suggest the significance of cover in the habitat use of the fish. In [20], the significance of cover was also supported by an approach using Akaike Information Criterion (AIC).

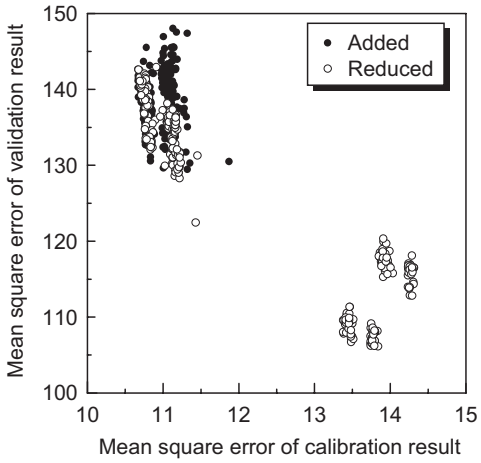


Figure 6: Scatter diagram between mean square errors of calibration and validation results, in which open circles represent the case of reduced number of fuzzy membership function while filled circles represent that of added.

As a concept, adding fuzzy membership function would improve model performance by precisely describing habitat preference of the fish. By contrast, increasing number of parameters to be optimised would put searching loads on GA. Although no clear trend was found between accuracy of the habitat prediction and model structure (i.e., number of fuzzy membership functions), further analyses should be carried out to assess the complex combination such as reducing and adding fuzzy membership functions at the same time. The use of AIC would be appropriate for the quantification of the complex relationship between model structure and prediction ability. In addition, different performance measures such as correctly fuzzy classified instances (%CCFI) and average deviation (AD), and interpretability-preserving optimization of the fuzzy models as presented in [36], can be a pathway for the accuracy improvement of the present model. Further studies would be needed to quantitatively understand the complexity-accuracy tradeoff of FHPM .

Acknowledgment

The author wishes his thanks to Prof. Dr. K. Hiramatsu (Faculty of Agriculture, Kyushu University) for his generous and capable assistance in all phases of this study. This study was partly supported by Grant-in-aid for Young Scientists (start-up) of Japan Society for the Promotion of Science (JSPS), Japan.

References

[1] P. W. Downs and K.J. Gregory, River Channel Management: Toward Sustainable Catchment Hydorosystems. Arnold, London. 2004.

[2] U.S. Fish and Wildlife Service, Habitat Evaluation Procedures (HEP): Ecological Service Manual 102. Washington, DC, USA. 1980.

[3] U.S. Fish and Wildlife Service, Standards for the Development of Habitat Suitability Index Models: Ecological Service Manual 103.Washington, DC, USA. 1980.

[4] K.D. Bovee et al., Stream habitat analysis using the instream flow incremental methodology. U.S. Geological Survey, Biological Resources Division Information and Technology Report. USGS/BRD-1998-0004. 1998.

[5] B. Ahmadi-Nedushan et al., A review of statistical methods for the evaluation of aquatic habitat suitability for instream flow assessment. River Res. Appl. 2006.

[6] D. J. Orth and O. E Maughan, 1982. Evaluation of the incremental methodology for recommending instream flows for fishes. Trans. Am. Fish. Soc. 111, 413–445., 1982.

[7] J.C. Guay et al., Development and validation of numerical habitat models for juveniles of Atlantic salmon (*Salmo salar*). Can. J. Fish. Aquat. Sci. 57, 2065–2075., 2000.

[8] J.C. Guay et al., Assessment of the transferability of biological habitat models for Atlantic salmon parr (*Salmo salar*). Can. J. Fish. Aquat. Sci. 60, 1398–1408., 2003

[9] M. Nykänen and A. Huusko, Transferability of habitat preference criteria for larval European grayling (*Thymallus thymallus*). Can. J. Fish. Aquat. Sci. 61, 185–192, 2004.

[10] S. Lek, Uncertainty in ecological models. Ecol. Model. 207, 1–2., 2007

[11] W. Silvert, Fuzzy indices of environmental conditions. Ecol. Model. 130, 111–119., 2000.

[12] K. Jorde et al., Fuzzy based models for the evaluation of fish habitat quality and instream flow assessment. In: Proceedings of the 2001 International Symposium on Environmental Hydraulics., 2001;

[13] V. Adriaenssens et al., Fuzzy knowledge-based models for prediction of *Asellus* and *Gammarus* in watercourses in Flanders (Belgium). Ecol. Model. 195, 3–10, 2006.

[14] E. Van Broekhoven et al., Fuzzy rule-based macroinvertebrate habitat suitability models for running waters. Ecol. Model. 198, 71–84., 2006.

[15] A. S. Mouton et al., Fish habitat modelling as a tool for river management. Ecol. Eng. 29, 305–315, 2007.

ISBN: 978-989-95079-6-8

[16] A. S. Mouton et al., Optimisation of a fuzzy physical habitat model for spawning European grayling (*Thymallus thymallus* L.) in the Aare river (Thun, Switzerland). Ecol. Model. 215, 122–132, 2008.

[17] M. A. Burgman et al., 2001. Setting reliability bounds on habitat suitability indices. Ecol. Appl. 11 (1), 70–78., 2001

[18] K. Hiramatsu et al., Mathematical modeling of habitat preference of Japanese medaka for instream water environment using fuzzy inference. Trans. JSIDRE 228, 65–72 (in Japanese with English abstract)., 2003.

[19] S. Fukuda et al., Mathematical characterization of fuzziness in fish habitat preference of Japanese medaka (*Oryzias latipes*) in agricultural canal. Trans. JSIDRE 239, 43–49 (in Japanese with English abstract)., 2005.

[20] S. Fukuda et al., An information-theoretic approach for the model selection in habitat preference evaluation of Japanese medaka (*Oryzias latipes*). Bull. Inst. Trop. Agric. Kyushu Univ. 29, 105–118., 2006.

[21] S. Fukuda et al., Numerical quantification of the significance of aquatic vegetation affecting spatial distribution of Japanese medaka (*Oryzias latipes*) in an agricultural canal. Landscape Ecol. Eng. 2, 65–80, 2006.

[22] S. Fukuda and K. Hiramatsu, Prediction ability and sensitivity of artificial intelligence-based habitat preference models for predicting spatial distribution of Japanese medaka (*Oryzias latipes*). Ecol. Model. 215, 301-313., 2008.

[23] S. Fukuda and S. Okushima, Assessing nonlinearity in fish habitat preference of Japanese medaka (*Oryzias latipes*) using genetic algorithm-optimized habitat prediction models. JARQ-Jpn. Agric. Res. Quart. 42 (2), 97–107., 2008.

[24] S. Fukuda, Consideration of fuzziness: Is it necessary in modelling fish habitat preference of Japanese medaka (*Oryzias latipes*)? Ecol. Model., in press, 2009

[25] S. Fukuda et al., Fuzzy neural network model for habitat prediction and HEP for habitat quality estimation focusing on Japanese medaka (*Oryzias latipes*) in agricultural canals. Paddy Water Environ. 4 (3), 119–124, 2006.

[26] S. Fukuda, Uncertainty analysis on fuzzy neural network model for evaluating fish habitat preference of Japanese medaka (*Oryzias latipes*) in agricultural canals in Japan. Proceedings of 7th International Symposium on Ecohydraulics. Paper ID: conf187a145., 2009.

[27] R.W. Bosserman and R.K. Ragade, Ecosystem analysis using fuzzy set theory. Ecol. Model. 16, 191–208. 1982

[28] A. Salski, Fuzzy knowledge-based models in ecological research. Ecol. Model. 63, 103–112. 1992.

[29] W. Bock and A. Salski, A fuzzy knowledge-based model of population dynamics of the Yellow-necked mouse (*Apodemus flavicollis*) in a beech forest. Ecol. Model. 108, 155–161. 1998.

[30] D.G. Chen and S.R. Hare, Neural network and fuzzy logic models for pacific halibut recruitment analysis. Ecol. Model. 195, 11–19. 2006.

[31] S. Mackinson, An adaptive fuzzy expert system for predicting structure, dynamics and distribution of herring shoals. Ecol. Model. 126, 155–178. 2000.

[32] A. Salski, Fuzzy clustering of fuzzy ecological data. Ecol. Inform. 2, 262–269. 2007.

[33] A.M.F. Cornelissen et al., Elicitation of expert knowledge for fuzzy evaluation of agricultural production systems. Agric. Ecosyst. Environ. 95, 1–18. 2003.

[34] R. Bosma, et al., Fuzzy modelling of farmers motivations for integrated farming in the Vietnamese Mekong Delta. The 2005 IEEE International Conference on Fuzzy Systems 827–832. 2005.

[35] V. Adriaenssens et al., Fuzzy rule-based models for decision support in ecosystem management. Sci. Total Environ. 319, 1–12. 2004.

[36] E. Van Broekhoven et al., Interpretability-preserving genetic optimization of linguistic terms in fuzzy models for fuzzy ordered classification: An ecological case study. Int. J. Approx. Reasoning. 44, 65–90. 2007.

[37] F. Herrera, Genetic fuzzy systems: taxonomy, current research trends and prospects. Evol. Intel. 1, 27–46. 2008.

[38] S. C. Dekker et al., Analysing forest transpiration model errors with artificial neural networks. J. Hydrol. 246, 197–208., 2001

ISBN: 978-989-95079-6-8

Towards A Neural-Based Understanding of the Cauchy Deviate Method for Processing Interval and Fuzzy Uncertainty

Vladik Kreinovich[1] and Hung T. Nguyen[2]

[1]Department of Computer Science, University in the Texas at El Paso
500 W. University, El Paso, TX 79968, USA, vladik@utep.edu
[2]Department of Mathematical Sciences, New Mexico State University
Las Cruces, NM 88003, USA, hunguyen@nmsu.edu

Abstract— *One of the most efficient techniques for processing interval and fuzzy data is a Monte-Carlo type technique of Cauchy deviates that uses Cauchy distributions. This technique is mathematically valid, but somewhat counterintuitive. In this paper, following the ideas of Paul Werbos, we provide a natural neural network explanation for this technique.*

Keywords— Cauchy deviate method, fuzzy uncertainty, interval uncertainty, Monte-Carlo simulations, neural networks

1 Formulation of the Problem: Cauchy Deviate Method and Need for Intuitive Explanation

1.1 Practical Need for Uncertainty Propagation

In many practical situations, we are interested in the value of a quantity y which is difficult or even impossible to measure directly. To estimate this difficult-to-measure quantity y, we measure or estimate related easier-to-measure quantities $x_1, \ldots, x_n$ which are related to the desired quantity y by a known relation $y = f(x_1, \ldots, x_n)$. Then, we apply the relation f to the estimates $\widetilde{x}_1, \ldots, \widetilde{x}_n$ for x_i and produce an estimate $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ for the desired quantity y.

In the simplest cases, the relation $f(x_1, \ldots, x_n)$ may be an explicit expression: e.g., if we know the current x_1 and the resistance x_2, then we can measure the voltage y by using Ohm's law $y = x_1 \cdot x_2$. In many practical situations, the relation between x_i and y is much more complicated: the corresponding algorithm $f(x_1, \ldots, x_n)$ is not an explicit expression, but a complex algorithm for solving an appropriate non-linear equation (or system of equations).

Estimates are never absolutely accurate:

- measurements are never absolutely precise, and
- expert estimates can only provide approximate values of the directly measured quantities $x_1, \ldots, x_n$.

In both cases, the resulting estimates $\widetilde{x}_i$ are, in general, different from the actual (unknown) values x_i. Due to these estimation errors $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i$, even if the relation $f(x_1, \ldots, x_n)$ is exact, the estimate $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ is different from the actual value $y = f(x_1, \ldots, x_n)$: $\Delta y \stackrel{\text{def}}{=} \widetilde{y} - y \neq 0$.

(In many situations, when the relation $f(x_1, \ldots, x_n)$ is only known approximately, there is an additional source of the approximation error in y caused by the uncertainty in knowing this relation.)

It is therefore desirable to find out how the uncertainty Δx_i in estimating x_i affects the uncertainty Δy in the desired quantity, i.e., how the uncertainties Δx_i *propagate* via the algorithm $f(x_1, \ldots, x_n)$.

1.2 Propagation of Probabilistic Uncertainty

Often, we know the probabilities of different values of Δx_i. For example, in many cases, we know that the approximation errors Δx_i are independent normally distributed with zero mean and known standard deviations σ_i; see, e.g., [16].

In this case, we can use known statistical techniques to estimate the resulting uncertainty Δy in y. For example, since we know the probability distributions, we can simulate them in the computer, i.e., use the Monte-Carlo simulation techniques to get a sample population $\Delta y^{(1)}, \ldots, \Delta y^{(N)}$ of the corresponding errors Δy. Based on this sample, we can then estimate the desired statistical characteristics of the desired approximation error Δy.

1.3 Propagation of Interval Uncertainty

In many other practical situations, we do not know these probabilities, we only know the upper bounds Δ_i on the (absolute values of) the corresponding measurement errors Δx_i: $|\Delta x_i| \leq \Delta$.

In this case, based on the known approximation $\widetilde{x}_i$, we can conclude that the actual (unknown) value of i-th auxiliary quantity x_i can take any value from the interval

$$\mathbf{x}_i = [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]. \tag{1}$$

To find the resulting uncertainty in y, we must therefore find the range $\mathbf{y} = [\underline{y}, \overline{y}]$ of possible values of y when $x_i \in \mathbf{x}_i$:

$$\mathbf{y} = f(\mathbf{x}_1, \ldots, \mathbf{x}_n) \stackrel{\text{def}}{=}$$

$$\{f(x_1, \ldots, x_n) \,|\, x_1 \in \mathbf{x}_1, \ldots, x_n \in \mathbf{x}_n\}. \tag{2}$$

Computations of this range under interval uncertainty is called *interval computations*; see, e.g., [4, 5].

The corresponding computational problems are, in general, NP-hard [9]. Crudely speaking, this means that, in general, such problems require a large amount of computation time – and that therefore faster methods are needed.

1.4 Propagation of Fuzzy Uncertainty

In many practical situations, the estimates $\widetilde{x}_i$ come from experts. Experts often describe the inaccuracy of their estimates

ISBN: 978-989-95079-6-8

in terms of imprecise words from natural language, such as "approximately 0.1", etc. A natural way to formalize such words is to use special techniques developed for formalizing this type of estimates – specifically, the technique of fuzzy logic; see, e.g., [6, 15].

In this technique, for each possible value of $x_i \in \mathbf{x}_i$, we describe the degree $\mu_i(x_i)$ to which this value is possible. For each degree of certainty α, we can determine the set of values of x_i that are possible with at least this degree of certainty – the *α-cut* $\mathbf{x}_i(\alpha) = \{x \,|\, \mu(x) \geq \alpha\}$ of the original fuzzy set. Vice versa, if we know α-cuts for every α, then, for each object x, we can determine the degree of possibility that x belongs to the original fuzzy set [3, 6, 12, 13, 15]. A fuzzy set can be thus viewed as a nested family of its (interval) α-cuts.

We already know how to propagate interval uncertainty. Thus, to propagate this fuzzy uncertainty, we can therefore consider, for each α, the fuzzy set y with the α-cuts

$$\mathbf{y}(\alpha) = f(\mathbf{x}_1(\alpha), \ldots, \mathbf{x}_1(\alpha)); \tag{3}$$

see, e.g., [3, 6, 12, 13, 15]. So, from the computational viewpoint, the problem of propagating fuzzy uncertainty can be reduced to several interval propagation problems.

1.5 Need for Faster Algorithms for Uncertainty Propagation

Summarizing the above analysis, we can conclude that in principle, we need to consider three basic types of uncertainty propagation: situations when we propagate probabilistic, interval, and fuzzy uncertainty. It is also possible that some quantities are represented by fuzzy sets, while others may be represented by probabilities.

For probabilistic uncertainty, there exist reasonable efficient uncertainty propagation algorithms such as Monte-Carlo simulations. In contrast, the problems of propagating interval and fuzzy uncertainty are, in general, computationally difficult. It is therefore desirable to design faster algorithms for propagating interval and fuzzy uncertainty.

Once such methods are developed, we can then use these methods to propagate interval and fuzzy uncertainty components, and Monte-Carlo simulations to propagate the probabilistic uncertainty.

The computational problem of propagating fuzzy uncertainty can be naturally reduced to the problem of propagating interval uncertainty. Because of this reduction, in the following text, we will mainly concentrate on faster algorithms for propagating interval uncertainty.

1.6 Linearization Situations: Description

Due to the approximation errors $\Delta x_i = \widetilde{x}_i - x_i$, the unknown (actual) values $x_i = \widetilde{x}_i - \Delta x_i$ of the input quantities x_i are, in general, different from the approximate estimates $\widetilde{x}_i$. In many practical situations, the approximation errors Δx_i are small – e.g., when the approximations are obtained by reasonably accurate measurements. In such situations, we can ignore terms which are quadratic (and of higher order) in Δx_i.

1.7 Linearization Situations: Analysis

In the above situations, we can expand the expression for

$$\Delta y = \widetilde{y} - y = f(\widetilde{x}_1, \ldots, \widetilde{x}_n) - f(x_1, \ldots, x_n) =$$

$$f(\widetilde{x}_1, \ldots, \widetilde{x}_n) - f(\widetilde{x}_1 - \Delta x_1, \ldots, \widetilde{x}_n - \Delta x_n) \tag{4}$$

in Taylor series in Δx_i and keep only the linear terms in this expansion. In this case, we get

$$\Delta y = c_1 \cdot \Delta x_1 + \ldots + c_n \cdot \Delta x_n, \tag{5}$$

where we denoted

$$c_i \stackrel{\text{def}}{=} \frac{\partial f}{\partial x_i}(\widetilde{x}_1, \ldots, \widetilde{x}_n).$$

For a linear function, the largest possible value of Δy is obtained when each of the variables $\Delta x_i \in [-\Delta_i, \Delta_i]$ attains:

- either its largest value Δ_i (when $c_i \geq 0$)
- or its smallest value $-\Delta_i$ (when $c_i < 0$).

In both cases, the largest possible value of the corresponding term in Δy is equal to $|c_i| \cdot \Delta_i$. Thus, the largest possible value of Δy is equal to

$$\Delta = |c_1| \cdot \Delta_1 + \ldots + |c_n| \cdot \Delta_n. \tag{6}$$

Similarly, the smallest possible value of Δy is obtained when each of the variables $\Delta x_i \in [-\Delta_i, \Delta_i]$ attains

- either its smallest value $-\Delta_i$ (when $c_i \geq 0$)
- or its largest value Δ_i (when $c_i < 0$).

In both cases, the smallest possible value of the corresponding term in Δy is equal to $-|c_i| \cdot \Delta_i$. Thus, the smallest possible value of Δy is equal to

$$-\Delta = -|c_1| \cdot \Delta_1 - \ldots - |c_n| \cdot \Delta_n. \tag{7}$$

Can we transform these natural formulas into an algorithm? Due to the linearization assumption, we can estimate each partial derivative c_i as

$$c_i \approx \frac{f(\widetilde{x}_1, \ldots, \widetilde{x}_{i-1}, \widetilde{x}_i + h_i, \widetilde{x}_{i+1}, \ldots, \widetilde{x}_n) - \widetilde{y}}{h_i} \tag{8}$$

for some small values h_i. So, we arrive at the following algorithm.

1.8 Linearization Situations: Algorithm

To compute the range $\mathbf{y}$ of y, we do the following.

- First, we apply the algorithm f to the original estimates $\widetilde{x}_1, \ldots, \widetilde{x}_n$, resulting in the value $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$.
- Second, for all i from 1 to n, we compute $f(\widetilde{x}_1, \ldots, \widetilde{x}_{i-1}, \widetilde{x}_i + h_i, \widetilde{x}_{i+1}, \ldots, \widetilde{x}_n)$ for some small h_i and then compute

$$c_i = \frac{f(\widetilde{x}_1, \ldots, \widetilde{x}_{i-1}, \widetilde{x}_i + h_i, \widetilde{x}_{i+1}, \ldots, \widetilde{x}_n) - \widetilde{y}}{h_i}. \tag{9}$$

- Finally, we compute

$$\Delta = |c_1| \cdot \Delta_1 + \ldots + |c_n| \cdot \Delta_n$$

and the desired range $\mathbf{y} = [\widetilde{y} - \Delta, \widetilde{y} + \Delta]$.

ISBN: 978-989-95079-6-8

1.9 Linearization Situations: Computational Complexity

The main computation time is spent on calling the time-consuming algorithm f. In the above uncertainty propagation algorithm, after one call to f to compute $\widetilde{y}$, we need n calls to f to compute the corresponding partial derivatives c_i and then, we can estimate the desired uncertainty Δ in y by using the above simple formula.

Overall, we thus need $n+1$ calls to the algorithm f.

1.10 Cauchy Deviate Method

For large n, we can further reduce the number of calls to f if we use a special technique of Cauchy-based Monte-Carlo simulations, which enables us to use a fixed number of calls to f (≈ 200) for all possible values n; see, e.g., [7, 8].

1.11 Mathematics Behind the Cauchy Method

In our simulations, we use *Cauchy distribution* – i.e., probability distributions with the probability density

$$\rho(z) = \frac{\Delta}{\pi \cdot (z^2 + \Delta^2)}; \tag{10}$$

the value Δ is called the *(scale) parameter* of this distribution.

Cauchy distribution has the following property that we will use: if $z_1, \ldots, z_n$ are independent random variables, and each of z_i is distributed according to the Cauchy law with parameter Δ_i, then their linear combination

$$z = c_1 \cdot z_1 + \ldots + c_n \cdot z_n \tag{11}$$

is also distributed according to a Cauchy law, with a scale parameter $\Delta = |c_1| \cdot \Delta_1 + \ldots + |c_n| \cdot \Delta_n$.

Therefore, if we take random variables δ_i which are Cauchy distributed with parameters Δ_i, then the value

$$\delta \stackrel{\text{def}}{=} f(\widetilde{x}_1, \ldots, \widetilde{x}_n) - f(\widetilde{x}_1 - \delta_1, \ldots, \widetilde{x}_n - \delta_n) =$$

$$c_1 \cdot \delta_1 + \ldots + c_n \cdot \delta_n \tag{12}$$

is Cauchy distributed with the desired parameter

$$\Delta = \sum_{i=1}^{n} |c_i| \cdot \Delta_i. \tag{13}$$

So, repeating this experiment N times, we get N values $\delta^{(1)}, \ldots, \delta^{(N)}$ which are Cauchy distributed with the unknown parameter, and from them we can estimate Δ.

The bigger N, the better estimates we get.

1.12 Cauchy Method: Towards Implementation

To implement this idea, we must answer the following two questions:

- how to simulate the Cauchy distribution; and
- how to estimate the parameter Δ of this distribution from a finite sample.

Simulation can be based on the functional transformation of uniformly distributed sample values:

$$\delta_i = \Delta_i \cdot \tan(\pi \cdot (r_i - 0.5)), \tag{14}$$

where r_i is uniformly distributed on the interval $[0, 1]$.

In order to estimate Δ, we can apply the Maximum Likelihood Method

$$\rho(\delta^{(1)}) \cdot \rho(\delta^{(2)}) \cdot \ldots \cdot \rho(\delta^{(N)}) \to \max, \tag{15}$$

where $\rho(z)$ is a Cauchy distribution density with the unknown Δ. When we substitute the above-given formula for $\rho(z)$ and equate the derivative of the product with respect to Δ to 0 (since it is a maximum), we get an equation

$$\frac{1}{1 + \left(\dfrac{\delta^{(1)}}{\Delta}\right)^2} + \ldots + \frac{1}{1 + \left(\dfrac{\delta^{(N)}}{\Delta}\right)^2} = \frac{N}{2}. \tag{16}$$

The left-hand side of (16) is an increasing function that is equal to $0 (< N/2)$ for $\Delta = 0$ and $> N/2$ for $\Delta = \max\left|\delta^{(k)}\right|$; therefore the solution to the equation (16) can be found by applying a bisection method to the interval $\left[0, \max\left|\delta^{(k)}\right|\right]$.

It is important to mention that we assumed that the function f is reasonably linear within the box

$$[\widetilde{x}_1 - \Delta_1, \widetilde{x}_1 + \Delta_1] \times \ldots \times [\widetilde{x}_n - \Delta_n, \widetilde{x}_n + \Delta_n]. \tag{17}$$

However, the simulated values δ_i may be outside the box. When we get such values, we do not use the function f for them, we use a normalized function that is equal to f within the box, and that is extended linearly for all other values (we will see, in the description of an algorithm, how this is done).

As a result, we arrive at the following algorithm.

1.13 Cauchy Deviates Method: Algorithm

- Apply f to the results of direct measurements:

$$\widetilde{y} := f(\widetilde{x}_1, \ldots, \widetilde{x}_n); \tag{18}$$

- For $k = 1, 2, \ldots, N$, repeat the following:
  - use the standard random number generator to compute n numbers $r_i^{(k)}$, $i = 1, 2, \ldots, n$, that are uniformly distributed on the interval $[0, 1]$;
  - compute Cauchy distributed values

$$c_i^{(k)} := \tan(\pi \cdot (r_i^{(k)} - 0.5)); \tag{19}$$

  - compute the largest value of $|c_i^{(k)}|$ so that we will be able to normalize the simulated measurement errors and apply f to the values that are within the box of possible values: $K := \max_i |c_i^{(k)}|$;
  - compute the simulated measurement errors

$$\delta_i^{(k)} := \Delta_i \cdot c_i^{(k)} / K; \tag{20}$$

  - compute the simulated "actual values"

$$x_i^{(k)} := \widetilde{x}_i - \delta_i^{(k)}; \tag{21}$$

  - apply the program f to the simulated "actual values" and compute the simulated error of the indirect measurement:

$$\delta^{(k)} := K \cdot \left(\widetilde{y} - f\left(x_1^{(k)}, \ldots, x_n^{(k)}\right)\right); \tag{22}$$

ISBN: 978-989-95079-6-8

- Compute Δ by applying the bisection method to solve the equation (16).

Comment. To avoid confusion, we should emphasize that, in contrast to the Monte-Carlo solution for the probabilistic case, the use of Cauchy distribution in the interval case is a computational trick and *not* a truthful simulation of the actual measurement error Δx_i: indeed, we know that the actual value of Δx_i is always inside the interval $[-\Delta_i, \Delta_i]$, but a Cauchy distributed random attains values outside this interval as well.

1.14 Cauchy Deviate Method: Need for Intuitive Explanation

The above Cauchy deviate method is one of the most efficient techniques for processing interval and fuzzy data. However, this method has a serious drawback: while the corresponding technique is mathematically valid, it is somewhat counterintuitive – we want to analyze errors which are located *instead* a given interval $[-\Delta, \Delta]$, but this analysis use Cauchy simulated errors which are located, with a high probability, *outside* this interval.

It is therefore desirable to come up with an intuitive explanation for this technique. In this paper, we show that such an explanation can be obtained from neural networks. (For a general introduction to neural networks, see, e.g., [2, 11].)

2 Solution: Neural Explanation

2.1 Werbos's Idea: Use Neurons

Our explanation comes from the idea promoted by Paul Werbos, the author of the backpropagation algorithm for training neural networks. Traditionally, neural networks are used to simulate a deterministic dependence; Paul Werbos suggested that the same neural networks can be used to describe stochastic dependencies as well – if as one of the inputs, we take a standard random number r uniformly distributed on the interval $[0, 1]$; see, e.g., [18] and references therein.

In view of this idea, as a natural probability distribution, we can take the result of applying a neural network to this random number. The simplest case is when we have a single neuron. In this case, we apply the activation (input-output) function $f(y)$ corresponding to this neuron to the random number r.

So, let us see what will happen if we apply a neuron to the standard random number and get a value $f(r)$.

2.2 What is the Activation Function of a Neuron: Reminder

To answer the above question, let us recall what are the optimal choices of an activation function of a neuron. This problem was analyzed in detail in [14]; see also [10].

2.3 We Must Choose a Family of Functions, Not a Single Function

We talk about choosing f, but the expression for $f(y)$ will change if we change the units in which we measure all the signals (input, output and intermediate), so in mathematical terms, it is better to speak about choosing a *family* of functions f.

It is reasonable to suggest that if an f belongs to this family, then this family must contain $k \cdot f$ for positive real numbers k. This corresponds to changing units.

Also, it must contain $f + c$, where c is a constant. This is equivalent to adding a constant bias and therefore does not change the abilities of the resulting network.

Since we are talking about non-linear phenomena, we can also assume that some non-linear "rescaling" transformations $x \to g(x)$ are also applicable, i.e., the family must include the composition $g(f(y))$ for each of functions f.

This family must not be too big, therefore, it must be determined by finitely many parameters and should ideally be obtained from one function $f(y)$ by applying all these transformations. Without loss of generality, we can assume that this set of transformations is closed under composition and under inverse, i.e., if $z \to g_1(z)$ and $z \to g_2(z)$ are possible transformations, then $z \to g_1(g_2(z))$ and $z \to g_1^{-1}(z)$ are possible transformations, where by g_1^{-1} we denoted an inverse function $g_1^{-1}(z) = w$ if and only if $g_1(w) = z$. In mathematical terms this means that these transformations form a *group*, and therefore a family is obtained by applying to some function $f(y)$ all transformations from some finite-dimensional transformation group G that includes all linear transformations (and maybe some non-linear ones).

All these transformations correspond to appropriate "rescalings". Rescaling is something that is smoothly changing the initial scale. This means that if we have two different transformations, there must be a smooth transition between them. In mathematical terms, the existence of this continuous transition is expressed by saying that the group is *connected*, and the fact that both the transformations and the transitions are smooth is expressed by saying that this is a *Lie group*).

2.4 Which Family is the Best?

Among all such families, we want to choose the best one. In formalizing what "the best" means we follow the general idea described in [14].

The criteria to choose may be computational simplicity, efficiency of training, or something else. In mathematical optimization problems, numeric criteria are most frequently used, when to every family we assign some value expressing its performance, and choose a family for which this value is maximal. However, it is not necessary to restrict ourselves to such numeric criteria only. For example, if we have several different families that have the same training ability A, we can choose between them the one that has the minimal computational complexity C. In this case, the actual criterion that we use to compare two families is not numeric, but more complicated: *a family F_1 is better than the family F_2 if and only if either $A(F_1) > A(F_2)$ or $A(F_1) = A(F_2)$ and $C(F_1) < C(F_2)$*. A criterion can be even more complicated. What a criterion *must* do is to allow us for every pair of families to tell whether the first family is better with respect to this criterion (we'll denote it by $F_1 > F_2$), or the second is better ($F_1 < F_2$) or these families have the same quality in the sense of this criterion (we'll denote it by $F_1 \sim F_2$). Of course, it is necessary to demand that these choices be consistent, e.g., if $F_1 > F_2$ and $F_2 > F_3$ then $F_1 > F_3$.

Another natural demand is that this criterion must choose a *unique* optimal family (i.e., a family that is better with respect to this criterion than any other family). The reason for this demand is very simple. If a criterion does not choose any family at all, then it is of no use. If several different families are "the

ISBN: 978-989-95079-6-8

best" according to this criterion, then we still have a problem to choose among those "best". Therefore, we need some additional criterion for that choice. For example, if several families turn out to have the same training ability, we can choose among them a family with minimal computational complexity. So what we actually do in this case is abandon that criterion for which there were several "best" families, and consider a new "composite" criterion instead: F_1 is better than F_2 according to this new criterion if either it was better according to the old criterion or according to the old criterion they had the same quality and F_1 is better than F_2 according to the additional criterion. In other words, if a criterion does not allow us to choose a unique best family it means that this criterion is not ultimate; we have to modify it until we come to a final criterion that will have that property.

The next natural condition that the criterion must satisfy is connected with the following. Suppose that instead of a neuron with the transformation function $f(y)$ we consider a neuron with a function $\bar{f}(y) = f(y + a)$, where a is a constant. This new neuron can be easily simulated by the old ones: namely, the output of this new neuron is $\bar{f}(y) = f(y + a)$, so it is equivalent to an old neuron with an additional constant input a. Likewise, the old neuron is equivalent to the new neuron with an additional constant input $-a$. Therefore, the networks that are formed by these new neurons have precisely the same abilities as those that are built from the old ones.

We cannot claim that the new neurons have the same quality as the old ones, because adding a can increase computational complexity and thus slightly worsen the overall quality. But it is natural to demand that adding a does not change the *relative* quality of the neurons, i.e., if a family $\{f(y)\}$ is better that a family of $\{g(y)\}$, then for every a the family $\{f(y+a)\}$ must be still better than the family $\{g(y + a)\}$.

Now, we are ready for the formal definitions.

2.5 Definitions

By a *transformation* we mean a smooth (differentiable) function from real numbers into real numbers. By an *appropriate transformation group* G we mean a finite-dimensional connected Lie group of transformations. By a *family* of functions we mean the set of functions that is obtained from a smooth (everywhere defined) non-constant function $f(y)$ by applying all the transformations from some appropriate transformation group G. Let us denote the set of all the families by F.

A pair of relations $(>, \sim)$ is called *consistent* if it satisfies the following conditions: (1) if $a > b$ and $b > c$ then $a > c$; (2) $a \sim a$; (3) if $a \sim b$ then $b \sim a$; (4) if $a \sim b$ and $b \sim c$ then $a \sim c$; (5) if $a > b$ and $b \sim c$ then $a > c$; (6) if $a \sim b$ and $b > c$ then $a > c$; (7) if $a > b$ then $b > a$ or $a \sim b$ are impossible.

Assume a set A is given. Its elements will be called *alternatives*. By an *optimality criterion* we mean a consistent pair $(>, \sim)$ of relations on the set A of all alternatives. If $a > b$, we say that a is *better* than b; if $a \sim b$, we say that the alternatives a and b are *equivalent* with respect to this criterion. We say that an alternative a is *optimal* (or *best*) with respect to a criterion $(>, \sim)$ if for every other alternative b either $a > b$ or $a \sim b$.

We say that a criterion is *final* if there exists an optimal alternative, and this optimal alternative is unique.

In the present section we consider optimality criteria on the set F of all families.

By the *result of adding* a to a function $f(y)$ we mean a function $\bar{f}(y) = f(y + a)$. By the *result of adding* a to a family F we mean the set of the functions that are obtained from $f \in F$ by adding a. This result will be denoted by $F+a$. We say that an optimality criterion on F is *shift-invariant* if for every two families F and G and for every number a, the following two conditions are true:

i) if F is better than G in the sense of this criterion (i.e., $F > G$), then $F + a > G + a$;

ii) if F is equivalent to G in the sense of this criterion (i.e., $F \sim G$), then $F + a \sim G + a$.

2.6 Main Result

As we have already remarked, the demands that the optimality criterion is final and shift-invariant are quite reasonable. The only problem with them is that at first glance they may seem rather weak. However, they are not, as the following theorem shows:

Theorem. *If a family F is optimal in the sense of some optimality criterion that is final and shift-invariant, then every function f from F has the form $a + b \cdot s_0(K \cdot y + l)$ for some a, b, K and l, where $s_0(y)$ is either a linear function, or a fractional-linear function, or $s_0(y) = \exp(y)$, or the logistic (sigmoid) function $s_0(y) = 1/(1 + \exp(-y))$, or $s_0(y) = \tan(y)$.*

Comment. The logistic function is indeed the most popular activation function for actual neural networks, but others are also used. For our purpose, we will use the tangent function. As we have mentioned earlier, the application of the tangent function to the standard random number r indeed leads to the desired Cauchy distribution.

2.7 Proof: Main Idea

The idea of this proof is as follows: first we prove that the appropriate transformation group consists of fractionally-linear functions (in Part 1), then we prove that the optimal family is shift-invariant (in Part 2), and from that in Part 3 we conclude that any function f from F satisfies some functional equations, whose solutions are known.

2.8 Proof: Part 1

By an *appropriate group* we meant a connected finite-dimensional Lie group of transformations of the set of real numbers R onto itself that contains all linear transformations. Norbert Wiener asked [19] to classify such groups for an n-dimensional space with arbitrary n, and this classification was obtained in [17]. In our case (when $n = 1$) the only possible groups are the group of all linear transformations and the group of all fractionally-linear transformations $x \to (a \cdot x + b)/(c \cdot x + d)$. In both cases the group consists only of fractionally linear transformations.

2.9 Proof: Part 2

Let us now prove that the optimal family F_{opt} exists and is *shift-invariant* in the sense that $F_{opt} = F_{opt} + a$ for all real numbers a. Indeed, we assumed that the optimality criterion

ISBN: 978-989-95079-6-8

is final, therefore there exists a unique optimal family F_{opt}. Let's now prove that this optimal family is shift-invariant.

The fact that F_{opt} is optimal means that for every other F, either $F_{opt} > F$ or $F_{opt} \sim F$. If $F_{opt} \sim F$ for some $F \neq F_{opt}$, then from the definition of the optimality criterion we can easily deduce that F is also optimal, which contradicts the fact that there is only one optimal family. So for every F either $F_{opt} > F$ or $F_{opt} = F$.

Take an arbitrary a and let $F = F_{opt} + a$. If $F_{opt} > F = F_{opt} + a$, then from the invariance of the optimality criterion (condition ii) we conclude that $F_{opt} - a > F_{opt}$, and that conclusion contradicts the choice of F_{opt} as the optimal family. So $F_{opt} > F = F_{opt} + a$ is impossible, and therefore $F_{opt} = F = F_{opt} + a$, i.e., the optimal family is really shift-invariant.

2.10 Proof: Part 3

Let us now deduce the actual form of the functions f from the optimal family. If $f(y)$ is such a function, then the result $f(y + a)$ of adding a to this function f belongs to $F + a$, and so, due to 2., it belongs to F. But all the functions from f can be obtained from each other by fractionally linear transformations, so $f(y + a) = (A + B \cdot f(y))/(C + D \cdot f(y))$ for some A, B, C and D. So we arrive at a functional equation for f. Let us reduce this equation to a one with a known solution. For that purpose, let us use the fact that fractionally linear transformations are projective transformations of a line, and for such transformations the cross ratio is preserved ([1], Section 2.3), i.e., if $g(y) = (A + B \cdot f(y))/(C + D \cdot f(y))$, then

$$\frac{g(y_1) - g(y_3)}{g(y_2) - g(y_3)} \cdot \frac{g(y_2) - g(y_4)}{g(y_1) - g(y_4)} = \frac{f(y_1) - f(y_3)}{f(y_2) - f(y_3)} \cdot \frac{f(y_2) - f(y_4)}{f(y_1) - f(y_4)} \tag{23}$$

for all y_i. In our case this is true for $g(y) = f(y+a)$, therefore for all a the following equality is true:

$$\frac{f(y_1 + a) - f(y_3 + a)}{f(y_2 + a) - f(y_3 + a)} \cdot \frac{f(y_2 + a) - f(y_4 + a)}{f(y_1 + a) - f(y_4 + a)} = \frac{f(y_1) - f(y_3)}{f(y_2) - f(y_3)} \cdot \frac{f(y_2) - f(y_4)}{f(y_1) - f(y_4)}. \tag{24}$$

The most general continuous solutions of this functional equation are given by Theorem 2.3.2 from [1]: either f is fractionally linear, or $f(y) = (a+b\cdot\tan(k\cdot y))/(c+d\cdot\tan(k\cdot y))$ for some a, b, c, d, or $f(y) = (a+b\cdot\tanh(k\cdot y))(c+d\cdot\tanh(k\cdot y))$, where

$$\tanh(z) \stackrel{\text{def}}{=} \frac{\sinh(z)}{\cosh(z)}, \quad \sinh(z) \stackrel{\text{def}}{=} \frac{\exp(z) - \exp(-z))}{2},$$

$$\cosh(z) \stackrel{\text{def}}{=} \frac{\exp(z) + \exp(-z)}{2}. \tag{25}$$

The $\tanh(z)$ expression is equivalent to the logistic function.

The theorem is proven.

Acknowledgment

This work was supported in part by NSF grant HRD-0734825 and by Grant 1 T36 GM078000-01 from the National Institutes of Health. The authors are thankful to Paul Werbos for valuable discussions, and to the anonymous referees for valuable suggestions.

References

[1] J. Aczel, *Lectures on functional equations and their applications,* Dover, New York, 2006.

[2] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, 2006.

[3] D. Dubois and H. Prade, Operations on fuzzy numbers, *International Journal of Systems Science*, 1978, Vol. 9, pp. 613–626.

[4] Interval computations website http://www.cs.utep.edu/interval-comp

[5] L. Jaulin, M. Kieffer, O. Didrit, and E. Walter, *Applied Interval Analysis, with Examples in Parameter and State Estimation, Robust Control and Robotics*, Springer-Verlag, London, 2001.

[6] G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic*, Prentice Hall, New Jersey, 1995.

[7] V. Kreinovich, J. Beck, C. Ferregut, A. Sanchez, G. R. Keller, M. Averill, and S. A. Starks, "Monte-Carlo-type techniques for processing interval uncertainty, and their potential engineering applications", *Reliable Computing*, 2007, Vol. 13, No. 1, pp. 25–69.

[8] V. Kreinovich and S. Ferson, "A New Cauchy-Based Black-Box Technique for Uncertainty in Risk Analysis", *Reliability Engineering and Systems Safety*, 2004, Vol. 85, No. 1–3, pp. 267–279.

[9] V. Kreinovich, A. Lakeyev, J. Rohn, and P. Kahl, *Computational complexity and feasibility of data processing and interval computations*, Kluwer, Dordrecht, 1998.

[10] V. Kreinovich and C. Quintana. "Neural networks: what non- linearity to choose?," *Proceedings of the 4th University of New Brunswick Artificial Intelligence Workshop*, Fredericton, N.B., Canada, 1991, pp. 627–637.

[11] J. Lawrence, *Introduction to Neural Networks*, California Scientific Software Press, 1994.

[12] R. E. Moore and W. Lodwick, Interval Analysis and Fuzzy Set Theory, *Fuzzy Sets and Systems*, 2003, Vol. 135, No. 1, pp. 5–9.

[13] H. T. Nguyen and V. Kreinovich, Nested Intervals and Sets: Concepts, Relations to Fuzzy Sets, and Applications, In: R. B. Kearfott and V. Kreinovich, eds., *Applications of Interval Computations*, Kluwer, Dordrecht, 1996, pp. 245–290.

[14] H. T. Nguyen and V. Kreinovich, *Applications of Continuous Mathematics to Computer Science*, Kluwer, Dordrecht, 1997.

[15] H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, CRC Press, Boca Raton, Florida, 2006.

[16] S. Rabinovich, *Measurement Errors and Uncertainties: Theory and Practice*, American Institute of Physics, New York, NY, 2005.

[17] I. M. Singer and S. Sternberg, "Infinite groups of Lie and Cartan, Part 1", *Journal d'Analyse Mathematique*, 1965, Vol. XV, pp. 1–113.

[18] P. J. Werbos, "Intelligence in the brain: A theory of how it works and how to build it", *Neural Networks*, 2009, to appear.

[19] N. Wiener, *Cybernetics, or Control and Communication in the animal and the machine*, MIT Press, Cambridge, MA, 1962.

Fuzzy and PSO Based Algorithm for driver's behavior modeling

S. Khanmohammadi [1] M.A. Tinati[2] S. Ghaemi[2] M.A. Badamchizadeh[2]

1 Islamic Azad University, Science & Research Branch,
Poonak sq. Ashrafi Esfahani Blv., Hesarak, Tehran, Iran
2 Faculty of Electrical and Computer Engineering,
University of Tabriz, Tabriz, Iran
Email: khan@iaucss.org, {Tinati, ghaemi, mbadamchi}@tabrizu.ac.ir

Abstract— *The study of human behavior during driving is of primary importance for the improvement of drivers' security. This study is complex because of numerous situations in which the driver may be involved. In this paper, we propose a hierarchical fuzzy system for human in a driver-vehicle-environment system to model takeover by different drivers. The driver's behavior is affected by the environment. We include climate, road and car conditions in a fuzzy mode. For obtaining fuzzy rules, we have provided three separate questionnaires on the effects of climate; road and car condition on driver's performance. The number of fuzzy rules is optimized by Particle Swarm Optimization algorithm. Also the precision, age and driving individuality are used to model the driver's behavior under difference environments. We investigate the behavior of different drivers when a driver intends to pass the leading car. The comparative study showed that the simulation result is in good agreement with the real situations.*

Keywords— Fuzzy, PSO, Driver's behaviour.

1 Introduction

In recent years, by using new control ideas, safety in driving has been improved. MacAdam (1980) developed an optimal preview control algorithm; however, this algorithm could only be applied to single input single output systems. Fenton applied Linear Quadratic algorithm (1988) to design a controller for steering. Ackermann and Sienel (1990) used parameter space robust control to design the automatic steering controller. These models did not take the driver's preview behavior into consideration. Lee (1989) developed a discrete time preview control algorithm for four-wheel steering passenger vehicles and found that the control accuracy was improved substantially by taking the preview behavior into account. The fuzzy logic has proven to be a very effective tool for handling imprecision and uncertainty, which are both very important characteristics of driving environments. This makes fuzzy logic a powerful candidate tool in most traffic engineering studies [1]. In [2] Kamada et al. proposed a fuzzy logic lateral controller. Hessburg and Tomizuka [6] developed a fuzzy logic controller for vehicle lateral guidance which consisted of three sub-controllers: preview, feedback and gain scheduling. Cai, Lin and Mourant investigated the influence of driver emotion on performance through platoon driving simulated with multiple simulators [4]. They induced two kinds of emotion states (anger and excitation) through realistic driver-driver interaction by using networked driving simulators. Drivers' psycho physiological parameters changes were the indicators of emotions. In the anger and excitation states, drivers showed poorer lane control capability. In [4] the author investigated emotional behavior (anger, neutral, and excitation) of drivers by collecting driving performance data. The results demonstrated the feasibility and efficiency of using multiple networked driving simulators to study driver emotional behavior, e.g., road rage.

In this paper, we use new parameters to identify uniqueness of driving maneuver of each driver under different environment, the precision, age and driving individuality. In modeling, the decision making process is based on three positions, one position in equal lane and two positions in opposite lane. By considering three positions, the speed, direction of car and the steering angle, a fuzzy model is presented for steering angle and speed control. We used two levels for modeling. The low level control model is responsible for modeling the steering angle and the speed variations enforced by the driver. The high level control model, models the decision making process of the driver. For this purpose, first the fuzzified low level control models are illustrated in section 2. The high level decision making model is designed in section 3 and the simulation results are presented in section 4. Finally section 5 describes conclusions.

2 Low level control

In order to implement the low-level control model, a simple car's model is required [5]. The car states include the Cartesian position (x,y) centered mid-way between the rear wheels and the car's orientation denoted by equations 1 to 3, where φ is the steering angle, V is the car's speed and θ is the angle of the car with respect to the X axis and L is the distance between center of rear and front wheels.

$$x_{n+1} = x_n + V_n \times \cos(\theta_n) \times \Delta t \quad (1)$$

$$y_{n+1} = y_n + V_n \times \sin(\theta_n) \times \Delta t \quad (2)$$

$$\theta_{n+1} = \theta_n + \frac{V_n \times \tan(\varphi_n) \times \Delta t}{L} \quad (3)$$

The driver controls V and φ. In order to model the complete low level control procedure, we have assumed that apart from the information from the environment that the driver perceives, other information such as age and driver's individuality are influential in the driver's control procedure. This can be formulated as:

$$\Delta V = f(x_n, y_n, \varphi_n, \theta_n, Environment_n, Driver_n) \quad (4)$$

ISBN: 978-989-95079-6-8

$$\Delta\varphi = f(x_n, y_n, \varphi_n, \theta_n, Environment_n, Driver_n) \quad (5)$$

Where $(x_n, y_n, \theta_n, \varphi_n)$ determine car's state, *Environment* is the information of the condition of *Climate*, *Road* and *Car* and *Driver* is the individual driver influence.

2-1) Environmental condition

The *Climate* condition is the information of luminosity, field of view, rain, temperature and humidity. The *Road* condition is about the traffic surrounding the driver, road safety, quality of road materials, moving obstacle and having enough signs and finally the *Car* condition is the information of car's ergonomic, safety equipments performance and agent of distraction in car. The information is considered as the input variables of fuzzy system (Figure 1).

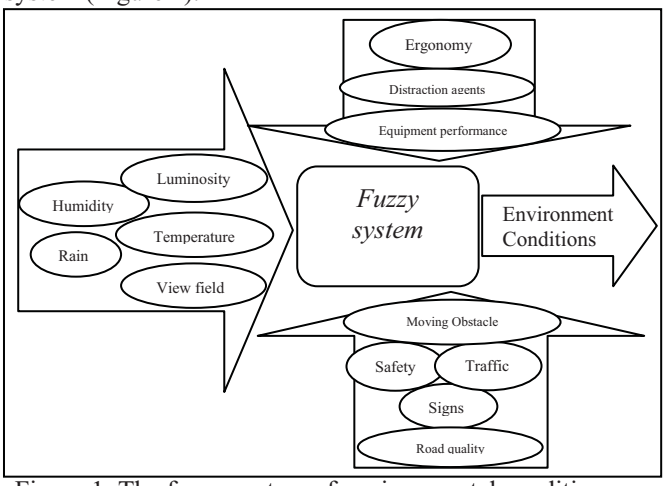


Figure 1. The fuzzy system of environmental condition

2-1-1) *Fuzzy rule base and membership functions*

Questionnaires have been provided to obtain the fuzzy rules, where the effects of climate, road and car condition in driving have been questioned separately. Each expert driver answered each question in three manners, percent value, graphic representation and linguistic term. We allocated a weight for each of them. For percent value $w_{per} = 1$, graphic representation $w_{gr} = 2$ and finally for linguistic term $w_{ling} = 0.5$ have been considered.

The membership functions of fuzzy values are supposed to be bell shape as follows:

$$\mu(x) = \frac{1}{1 + \frac{(x-c)^2}{d}} \quad (6)$$

where x is the member of universe, c is the median and d represents the shape factor.

By using collected data, the median value and shape factor for each expert driver is calculated as follows:

$$c_{ij} = \frac{w_{per} \times val_{per,j} + w_{gr} \times val_{gr,j} + w_{ling} \times val_{ling,j}}{w_{per} + w_{gr} + w_{ling}} \quad (7)$$

$$d_{ij} = \frac{\left(\left|val_{per,j} - val_{gr,j}\right|\right) + \left(\left|val_{per,j} - val_{ling,j}\right|\right) + \left(\left|val_{gr,j} - val_{ling,j}\right|\right)}{3} \quad (8)$$

where c_{ij} is median and d_{ij} is the shape factor suggested by i^{th} driver for the j^{th} parameter. So we have n values of c_{ij} and d_{ij} for n drivers.

Now we can obtain the overall membership function for each of environmental parameters by:

$$\mu_{P_j}(x) = \frac{1}{1 + \frac{(x - c_{P_j})^2}{d_{P_j}}} \quad (9)$$

where c_{P_j} and d_{P_j} are the proposed median and shape factor for the j^{th} parameter. c_{P_j} is obtained based on the number of linguistic terms which are optimized by PSO algorithm (section 2-1-2). For example, suppose the number of linguistic terms and range of universe for a variable are 5 and [0 1] respectively. Then values of medians for five verbal are: 0, 0.25, 0.5, 0.75 and 1 for *very bad*, *bad*, *medium*, *good* and *very good* respectively.

Also d_{P_j} is defined as:

$$d_{Pj} = \frac{1}{2} \times \frac{W_{univ_j} \times d_{ref_j}}{W_{ref_univ_j} \times (N_{ling_term_j} - 1)} \quad (10)$$

where W_{univ_j} is the width of universes for j^{th} parameter, $N_{ling_term_j}$ is the number of linguistic terms for j^{th} parameter (obtained from PSO algorithm), $W_{ref_univ_j}$ and d_{ref_j} are the width and the shape factor of reference universe for j^{th} parameter. The average of n value of the shape factors (obtained from eq. 8) is considered as d_{ref}. For all of parameters, $W_{ref_univ_j}$ is fixed to 10.

The importance of the parameters of climate, road and car condition varies for different people. So we use a pair wise comparison table for determining the weights of different parameters. Table1 shows part of pair wise comparison table.

Table1. A part of compared table

Parameter	A	B	C	D
A		√	√	-

The above table shows that parameter A is more important than parameters B and C. But parameter D is more important than parameter A. The proposed method for determining weights of parameters in fuzzy rules is as follows:

First arrange the parameters. First parameter is the least important. Then a_i is assigned to i^{th} parameter such that $a_i = i \times w$. where i is the priority index of parameter and w is the priority weight which is considered as 2 in this paper. When k parameters have equal importance, $a = w \times \frac{\sum i}{k}$ for all of k parameters. After that, we form the priority table as follows:

- If i^{th} parameter is more important than the j^{th} parameter, value of i^{th} row and j^{th} column in table is equal to $T_{ij} = a_i - a_j$.

ISBN: 978-989-95079-6-8

- If the *jth* parameter is more important than the i^{th} one then $T_{ij} = \frac{1}{a_j - a_i}$
- For diagonal elements $T_{ii} = 1$

Finally the weight of each parameter is obtained as follow:

$$M_i = (\prod_{j=1}^{m} T_{ij})^{\frac{1}{m}} \quad , \quad W_i = \frac{M_i}{\sum_{j=1}^{m} M_j} \qquad (11)$$

where *m* is the number of parameters. The obtained weights for parameters are used to generate valid fuzzy rules in section 2-1-2.

The general form of fuzzy rules for climate condition is as follows:

If field of View is … *and* Luminosity is … *and* *Rain* is … *and* Temperature is … *and* Humidity is … *Then* climate condition is …

For road condition:

If Traffic is … *and* Road quality is … *and* Sign is … *and* Moving obstacle is … *and* Safety is … *Then* Road condition is …

Finally for car condition, we have:

If Safety equipment operation is … *and* Ergonomic is … *and* Distract agent is … *Then* Car condition is …

2-1-2) *Particle Swarm Optimization*

Particle Swarm Optimization (PSO) is an optimization algorithm which uses properties of a swarm to find an optimal solution [4]. In this work, the swarm is represented by 100 individuals (or *particles*) whose values change at each iteration. The performance of each particle is measured at each position using a "fitness" function. This function increases as the optimality of the solution increases; in this way, a particle with a higher fitness is considered to fit better than the one with a lower fitness. Also, a record of the best position (*pbest*) for each particle, as well as the best overall position (*gbest*) for all particles, is kept in memory. The entire swarm then searches around the gbest solution and each of the pbest solutions, all the while trying to find even better solutions.

a) particle representation

The number of fuzzy rules depends on the number of linguistic terms of input variables. So, in order to optimize number of fuzzy rules, the number of linguistic terms of each parameter of climate, road and car condition is optimized using PSO. The number of linguistic terms for each input parameter is randomly selected 3, 5 or 7, so maximum number of fuzzy rules will be 7^m (*m* is the number of input variables). The number of linguistic terms of output fuzzy variable is set to 5.

To optimize the number of fuzzy rules and linguistic terms, the following particle is considered for PSO:

Particle=[*par_set par_rules*]

In the presented particle, the first part *par*_set is related to number of linguistic terms and is defined as follows:

$par_set = (N_{in1}, N_{in2}, \ldots, N_{in_j}, \ldots, N_{in_m})$

Where *m* is the number of input variables and N_{in_j} shows the number of linguistic terms of j^{th} input variable. The number of linguistic terms of output variable is always equal to 5. So the *par_set* has *m* elements.

The second part of particle *par_rules* is:

$par_rules = (O_{R1}, O_{R2}, \ldots, O_{Ri}, \ldots, O_{R7^m})$

The allele value at each location in the second part of particle contains either zero or the label of an output linguistic value to be used for a given rule. In other words, if O_{Ri} represents the allele at position *i*, its nonzero value gives the consequent part (i.e., the label of the corresponding fuzzy set on the output variable) of the rule which corresponds to the i^{th} location of the rules particle. A particle containing a zero allele value at the i^{th} position (i.e., R_i=0) indicates that the rule set represented by the rule particle has not selected any rule with the i^{th} antecedent clause.

As mentioned before, maximum number of fuzzy rules is 7^m. So in each rule_par, when the number of fuzzy rules is less than 7^m, other elements are considered to be zero. A simple example is considered which has 2 input variables. Both of variables have 3 linguistic terms (named Lt1, Lt, Lt3). So the maximum of fuzzy rules is nine.

All fuzzy rules are as follows:

Rule1: *If* $input_1$ is Lt_1 *and* $input_2$ is Lt_1 *Then* output is O_{R1}
Rule2: *If* $input_1$ is Lt_1 *and* $input_2$ is Lt_2 *Then* output is O_{R2}
Rule3: *If* $input_1$ is Lt_1 *and* $input_2$ is Lt_3 *Then* output is O_{R3}
Rule4: *If* $input_1$ is Lt_2 *and* $input_2$ is Lt_1 *Then* output is O_{R4}
Rule5: *If* $input_1$ is Lt_2 *and* $input_2$ is Lt_2 *Then* output is O_{R5}
Rule6: *If* $input_1$ is Lt_2 *and* $input_2$ is Lt_3 *Then* output is O_{R6}
Rule7: *If* $input_1$ is Lt_3 *and* $input_2$ is Lt_1 *Then* output is O_{R7}
Rule8: *If* $input_1$ is Lt_3 *and* $input_2$ is Lt_2 *Then* output is O_{R8}
Rule9: *If* $input_1$ is Lt_3 *and* $input_2$ is Lt_3 *Then* output is O_{R9}

Since the number of linguistic terms of output variable is 5, then $O_{Ri} \in \{0, 1, 2, 3, 4, 5\}$. Suppose linguistic terms of output variable are {*very bad*, *bad*, *medium*, *good* and *very good*}, therefore, the allele value is 0 when the corresponding rule hasn't been selected as a fuzzy rule, 1 when output variable is *very bad*, 2 when it is *bad*, and so on.

The optimized *par_rules* is obtained by PSO algorithm as follows:

par_rules =(0, 1, 2, 0, 3, 0, 0, 3, 5, 0, … , 0)

where first nine values are related to fuzzy rules. By considering the optimized *par_rules* , We have five rules: Rule2, Rule3, Rule5, Rule8 and Rule9. So optimized fuzzy rules are:

If $input_1$ is Lt_1 *and* $input_2$ is Lt_2 *Then* output is *very bad*
If $input_1$ is Lt_1 *and* $input_2$ is Lt_3 *Then* output is *bad*
If $input_1$ is Lt_2 *and* $input_2$ is Lt_2 *Then* output is *medium*
If $input_1$ is Lt_3 *and* $input_2$ is Lt_2 *Then* output is *medium*
If $input_1$ is Lt_3 *and* $input_2$ is Lt_3 *Then* output is *very good*

b) The fitness function

In the fitness function, we need to use some valid fuzzy rules. In order to generate valid fuzzy rules, we select the worse (value=0) and the best (value=1) state (verbal value) of each input variable. The linguistic terms of them are *very bad* and *very good* respectively. For the middle state (*medium*), value of 0.5 is assigned. If the input variable is defined by 5 linguistic terms, the state between the worse

ISBN: 978-989-95079-6-8

and middle state (*bad*) is set to 0.25 and the state between the best and middle (*good*) is set to 0.75. Also the center of gravity defuzzifier is applied as follows:

$$z^* = \frac{\sum_{i=1}^{m} z_i \times w_i}{\sum_{i=1}^{m} w_i} \quad (12)$$

where z_i *and* w_i are the value and weight (section 2-1-1, eq.11) of i^{th} input variable respectively. All of the rules where z^* gets one of values 0, 0.25, 0.5, 0.75 and 1, corresponding to linguistic terms of input variables, are considered as valid rules.

For example if linguistic term of traffic, road quality, sign, moving obstacle and safety are considered to be medium, then values of all of them is 0.5. Then by considering appropriate weight for each input variable (section 2-1-1, eq.11), z^* will be obtained as follows:

$$z^* = \frac{0.5\times0.21+0.5\times0.13+0.5\times0.11+0.5\times0.21+0.5\times0.34}{0.21+0.13+0.11+0.21+0.34} = 0.5$$

The obtained value is corresponding to *medium* linguistic term for z^*. By this way the following valid rule is obtained:

Rule: If Traffic is medium *and* Road quality is medium *and* Sign is medium *and* Moving obstacle is medium *and* Safety is medium *Then* Road condition is medium

Finally the fitness function used for the optimization of the membership functions is defined by:

$$\text{Fitness} = W_b \times n_c + \frac{W_r}{n_R} \quad (13)$$

Where n_R is the number of rules and W_r is its weighting factor, n_c is the number of input data set which have same output for valid fuzzy rules and the rules generated by PSO and W_r is its weighting factor.

2-2) Driver behavior

The personality (individuality) of each person is very important in the task performance. Because it affects on driver's decisions and acts. We consider drivers personality into three groups, risky person, normal person and attentive person. Also, the people have different reactions in different ages. So we include age of driver in modeling of driver's behavior. In addition, the precision of each person affect on calculating of distance. So we include it by considering graduating diploma and the measure of necessary precision in his work. Figure 2 shows the hierarchical fuzzy system which models driver behavior in decision making. For obtaining fuzzy rules in hierarchical fuzzy system, we use the applied method to generating valid rules in PSO algorithm. In proposed model, only by applying one coefficient in the universes of speed change and steering angle change also the ideal low and high speed for each person, we can easily model the different behavior under take over conditions.

3 High Level Control

3-1) Structure of road and existence cars

Two lanes road is considered as the movement trajectory of cars. In the simulation, four cars are considered (Figure 3). The *A* car is the car which we intend to control it by fuzzy rules, it is called *Controlled car*. The *B* car is the back car which intends to pass *Controlled car* and it is called *BEDOL*. The *C* car is the front car which moves in opposite lane. If *C* car moves in the same direction as *Controlled car,* then it is called *FEDOL* and if direction of movement is in opposite direction then it is called *FODOL*. The *D* car is the front car which moves in the same lane of *Controlled car*. If *D* car moves in the same direction as controlled car then it is called *FEDEL* and if it is in opposite direction then it is called *FODEL*.

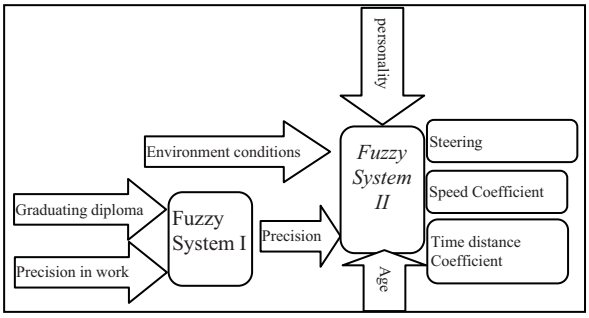


Figure 2. The driver's behavior hierarchy fuzzy system

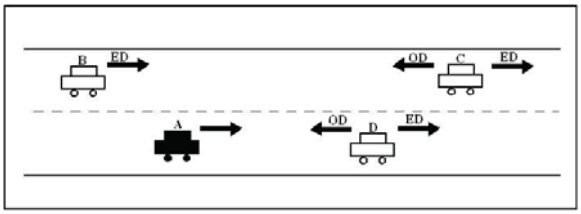
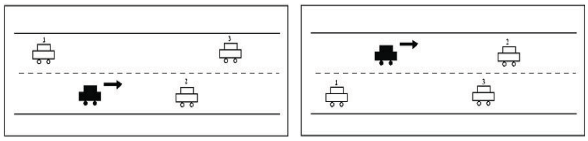


Figure 3. The movement trajectory of cars

3-2) Effective cars in the decision making of driver

When a person drives a car on the road, movements of other cars affect his/her decision making. For example, *FODEL* car is more important than *FEDEL* car and *FODOL* is more important than *FEDOL* car. When driver decides to pass *FEDEL* car, location of *BEDOL* and *FODOL* is very important. The three important cars in decision making of driver are shown in Figure 4.



(a) Controlled car is on the right lane

(b) Controlled car is on the left lane

Figure 4. The situations of important cars

3-3) Decision making process

By considering three positions, the speed and direction of car and the steering angle, the fuzzy variables for steering angle and speed control are presented as follows:

Td_{BOL}: (input variable) time distance of *Controlled car* with the back car in other lane, moving in the same direction.

Td_{FEL}: (input variable) Time distance with the front car, moving in the same lane

ISBN: 978-989-95079-6-8

Td_{FOL} : (input variable) Time distance with the car moving in opposite lane
V: (input variable) Speed of *Controlled car*
Dir: (input variable) *Controlled car*'s direction
Δv : (output variable) Change of speed
$\Delta\varphi$: (output variable) Direction change of steering.

Each of the input variables, Td_{BOL} , Td_{FEL} and Td_{FOL} has four linguistic values or fuzzy sets: The speed has three fuzzy sets. The car direction and the steering angle direction have seven fuzzy sets. Each of output variables is defined by using seven linguistic variables. We define a vector for Td_{BOL} , Td_{FEL} and Td_{FOL} separately. First element of each time distance vector represents the minimum time distance for collision with corresponding car. It is called very low. The second element is the lowest value of the middle time with corresponding car. We call it medium time. The last element is the lowest value as the highest time with corresponding car which is called high time. The above time distance vectors are defined for a *normal driver* who drives according to driving laws. But for modeling different behaviors of driver, we use the time distance coefficient which is obtained by hierarchical fuzzy system. The time distance coefficient is called *K_TD*. The purposed time distance vectors for different drivers are shown in table 2.
In table 2 we have

$$b=\begin{cases}(\sqrt{k_TD})^{-1}, & K_TD\geq 1\\ \sqrt{k_TD}, & K_TD<1\end{cases}$$

$$Td_{FEL}=\begin{cases}Td_{FEDEL}, & \text{if } FEDEL \text{ exict}\\ Td_{FODEL}, & \text{oherwise}\end{cases}$$

$$Td_{FOL}=\begin{cases}Td_{FEDOL}, & \text{if } FEDOL \text{ exict}\\ Td_{FODOL}, & \text{oherwise}\end{cases}$$

Table 2: Time distance vector for different drivers

Left Lane:	Right Lane:
Td_{BOL}=*K_TD**[0.5/b 2/b 7]	Td_{BOL}=*K_TD**[3 10 15]
Td_{FEDEL}=*K_TD**[1 3 5]	Td_{FEDEL}=*K_TD**[4/b 8/b 12]
Td_{FEDOL}=*K_TD**[3 5 8]	Td_{FEDOL}=*K_TD**[1 3 7]
Td_{FODEL}=*K_TD**[2 4 6]	Td_{FODEL}=*K_TD**[4 7 20]
Td_{FODOL}=*K_TD**[3 7 20]	Td_{FODOL}=*K_TD**[3 7 15]

3-3-1) Road partitions in decision making of driver
Drivers drive in different positions of road. When a driver intends to pass the leading car, he/she moves toward the middle of road while in low speed and dangerous conditions, the car moves toward the shoulder of road. So in decision making, we divide the road to four divisions: Left lane, Middle Lane, Right lane and Shoulder of lane. Among four decision lanes, the middle of lane and the shoulder of road lane are the transient lanes. A driver is driving in the right lane and only when he/she is passing the leading car, he/she drives in the left lane.

3-3-2) Fuzzy rule base for decision making of driver
The decision making process is categorized into four scenarios:
1. Right lane
- Staying in the same lane and continuing the path

a) No car is in front. The car continues its path.
b) A car is in front and the time distance with it is low, the driver starts to make a decision based on his/her desired speed, safety priorities but *FODOL* car is near or there is *BEDOL* car, then the driver decreases the car's speed and continues the path.

- Going to the shoulder of road

A car is in front and the time distance with it is very low; but *FODOL* car is near and *BEDOL* car is near. So the takeover from the left lane is impossible. In order to avoid collision, the driver must move the car to the shoulder of road.

- Going to the middle lane to passing

When the driver decides to change lane and perform a takeover from the left lane. So he/she moves the car to the middle lane.
2. Middle lane
As mentioned before, this lane is a transient lane. If the takeover conditions are satisfied, the driver increases the speed and continues with the same direction to enter the left lane. Otherwise he/she decides to stay in middle lane or to return to right lane.
After taking over, driver with the same speed from this lane enters to right lane
3. Left lane
- Before taking over

If the takeover conditions are satisfied, the driver continues the path with high speed. Otherwise he/she goes to middle lane and decides to stay there until the takeover conditions are provided or he/she returns to right lane.

- After taking over

The driver with the same speed from this lane goes to middle lane in order to go to the right lane.
4. Shoulder of road lane
As mentioned before, a driver in order to avoid collision goes to this lane. So he/she must stay in this lane until he/she could return.

4 Results

The obtained weights of each of environmental parameters from eq. (11) are summarized in Table 3. This table shows that among climate parameters, view field is the most important parameter. Safety is the most important among the road parameters and safety equipment operation is the most important among the car parameters.

Table 3: The weights of environmental parameters in fuzzy rules

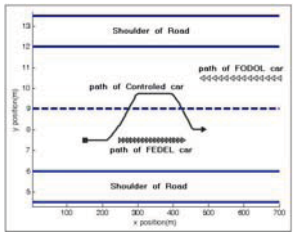
Climate parameter	*W*	Road parameter	*W*	Car parameter	*W*
View	0.37	Traff.	0.21	Eq.o.	0.47
Lum.	0.297	R.q.	0.13		
Rain	0.178	Sign	0.11	Er.	0.25
Temp.	0.098	M.obs	0.21	Dis.a.	0.28
Hum.	0.057	Safety	0.34		

ISBN: 978-989-95079-6-8

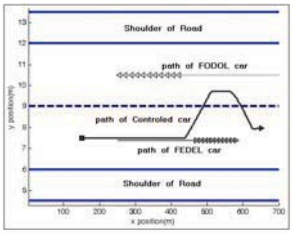
The obtained optimum number of fuzzy sets for environmental parameters input variables by PSO algorithm is shown in Table 4. The number of fuzzy sets of output variables (climate, road and car condition) is fixed and is equal to 5 sets.

Table 4. The optimum number of fuzzy sets for environmental parameters

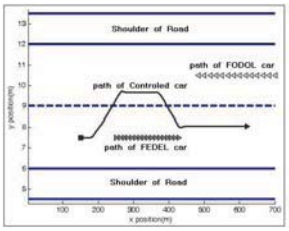
Climate parameter	number	Road parameter	number	Car parameter	number
View	5	Traff.	3	Eq.o.	5
Lum.	3	R.q.	5		
Rain	5	Sign	3	Er.	5
Temp.	3	M.obs.	5	Dis.a.	5
Hum.	3	Safety	3		



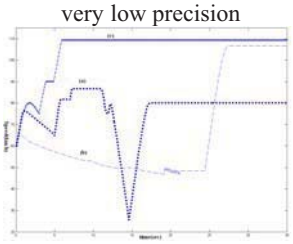
(a) Normal driver



(b) Attentive old driver with very low precision



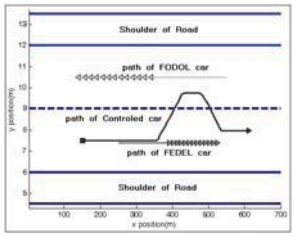
(c) Attentive old driver with very high precision



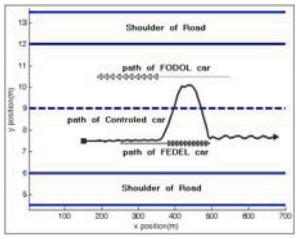
(d) Car's speed by (a), (b), (c) drivers

Figure 5: Comparison of different drivers in the same conditions

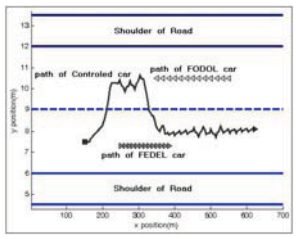
We investigate the behavior of different drivers under take over condition. The comparison of behavior of three drivers under equal passing conditions and the best environmental conditions is shown in Figures 5 and 6. As it is seen in Figure 5, *normal driver* and attentive old driver with very high precision pass the leading car. But attentive old driver with very low precision waits until the *FODOL* passes and then he/she passes the leading car. The paths of cars before passing are shown by solid lines. The very low precision person's error is higher than the very high precision person in calculating distance with other car. Because of being attentive, the driver waits for passing the leading car to collision avoidance. The car's speed of each driver shows the driver's behavior and decision under equal conditions. The *normal driver* is decided to passing and increases the speed. The movement of car is smooth. The old person with very high precision has more steering direction change because he/she is attentive and old and needs more time of passing than normal driver. The old person with very low precision is late in decision making and waits. After that he/she is passing the leading car and returns the right lane as soon as because of being attentive. Figure 6 compares a *normal driver* with a young driver. In equal conditions, the risky young driver passes the leading car. But the attentive young driver and normal driver waits until the *FODOL* car is passed. The speeds of cars are shown in figure (11- d). The results are obtained in good mental status.



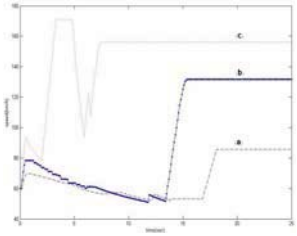
(a) Normal driver



(b) Attentive young driver with very low precision



(c) Risky young driver with very low precision



(d) Car's speed by (a), (b), (c) drivers

Figure 6: The comparison of different driving behavior in equal conditions

5 Conclusion

In this paper a fuzzy three positions model has been presented for the low-level control model and a fuzzy hierarchical system have been proposed for the high-level Control Model. The obtained results show that in equal conditions an attentive old driver with very high precision passes the leading car, but attentive old driver with very low precision waits until *FODOL* car passes and then he/she passes the leading car. The very low precision person's error is higher than that of the very high precision person in calculating distance with other car. This method also provides a basis for modeling individual driver behavior characteristics that may be tuned and used in automatically guided vehicles. It also provides a reference of natural driver behavior of each individual.

References

[1] Hoogendoom, S., Hoogendoom-lanser,S., Shouurman, H. "Perspectives of fuzzy logic in traffic engineering," TRB 78th Annual meeting, USA, 1999.

[2] M. Yoshida and H. Kamada, "A Visual Control System Using Image processing and fuzzy theory," IEEE round table discussion on vision based vehicle guidance, pp. 111-128, 1990.

[3] T. Hessburg and M. Tomizuka, "Fuzzy Logic Control for Lateral Vehicle Guidance," Proc. of IEEE Conference on Control Applications, Vancouver, pp. 581-586, Sept. 1993.

[4] H. Cai, Y. Lin and R. R. Mourant ,"Study on Driver Emotion in Driver-Vehicle-Environment Systems Using Multiple Networked Driving Simulators", DSC 2007 North America – Iowa City – September 2007.

[5] Mark W. Spong, Seth Hutchinson, M. Vidyasagar, "Robot Modeling and Control", John Wiley & Sons, Inc., 2006

[6] J, Kennedy and R. Eberhart, *Swarm Intelligence*, Morgan Kauffman Publishers, San Franscisco, CA. ISBN 1-55860-595-9, 2001.

ISBN: 978-989-95079-6-8

Evaluating Firms' Gender Equity by Fuzzy Logic

Tindara Addabbo[1] Gisella Facchinetti[1] Giovanni Mastroleo[2] Tiziana Lang[3]

1 University of Modena e Reggio Emilia, Department of Economia Politica, Modena, Italia
2 University of Calabria, Cosenza, Italia
3 Isfol, Roma, Italia
Email: tindara.addabbo@unimore.it, gisella.facchinetti@unimore.it, mastroleo@unical.it, t.lang@isfol.it

Abstract—This paper uses a Fuzzy Expert System (FES) in order to evaluate firms in terms of Gender Equity (GE). The scope for using FES is connected not only to the multidimensional nature of GE and to the need of providing a synthetic indicator of firm's GE without losing the its complexity, but even with the composite group of experts involved. The presence of sociologists, economists trade unions members, not used to a mathematics language, carried us to propose an instrument more user friendly like a FES.

Keywords— Fuzzy logic, Fuzzy expert system, gender equity, equal opportunity, discrimination.

1 Introduction

The Italian labour market shows remarkable gender inequalities notwithstanding the advanced labour market regulation in terms of Equal Opportunities. Italian women still experience many inequalities at their workplaces such as: wage differences, lower career paths, higher percentages of fixed-term and short-term contracts, etc. Women's employment rate on average in Italy was 46.6% in year 2007 (Istat, Labour Force Survey data) far below the 60% Lisbon target and far below men's employment rate (70.7%) Inequalities at the workplace are reinforced by and interact with the unequal distribution of unpaid work at home (care and domestic work). Italian women bear a higher share of unpaid work than their partners and when they enter the labour market this is going to produce an unequal share of total (paid and unpaid) work. [3],[10]. Though the Italian situation in terms of gender equity and access to paid labour is particularly weak gender segregation in employment is a feature that characterizes EU27 on average as the results by Burchell, Fagan, O'Brien and Smith [6], based on the European Foundation for the Improvement of living and working conditions 2005 Survey shows. The problem of firms' evaluation with regards to gender equity has been differently tackled across Europe in connection also to the different level of enforcement of equal opportunity laws and a different level of integration of gender equality policies in employment policies [5]. In order to improve women employment rates and to reduce the existing gender gaps in the workplaces the former Italian Ministry of Labour promoted the project "Bollino Rosa" together with Isfol (Istituto per la Formazione Professionale dei Lavoratori). The project started in June 2007 in the framework of the European Year of Equal Opportunities for All and has been co-financed by the European Commission. The test was carried out from November 2007 until February 2008. This study presents the fuzzy expert system (Section 2.1-3) proposed within the Bollino Rosa project and applied to a sample (section 2.4) of firms. Section 3 comments on the results of the application of the system to firms involved in the project while section 4 contains concluding remarks and proposals for the extension of the project.

2 A fuzzy expert system for evaluation

2.1 Why a fuzzy expert system

To face this complex problem and to reach an aggregated value of the certification level, we propose a Fuzzy Expert System (FES), which utilizes fuzzy sets and fuzzy logic to overcome some of the problems that occur when the data provided by the user are vague or incomplete.

This is not the natural framework of a FES, in fact engineering problems are more typical for FES, but recently economic and management researches have found in this instrument interesting applications, [2], [4], [7], [8], [9], [15]. In a multidisciplinary research, like the one we present, the power of FES shows its ability to describe linguistically a particular phenomenon or process, and then to represent that description with a small number of very flexible rules. In a FES, the knowledge is contained both in its rules and in fuzzy sets, which hold general description of the properties of the phenomenon under consideration. FES provides all possible solutions whose truth is above a certain threshold, and the user or the application program can then choose the appropriate solution depending on the particular situation. This fact adds flexibility to the system and makes it more powerful. FES uses fuzzy data, fuzzy rules, and fuzzy inference, in addition to the standard ones implemented in the ordinary Expert Systems. From the mathematical point of view a fuzzy system can be described as a function approximator. [16] More specifically it aims at performing an approximate implementation of an unknown mapping $f : A \subseteq R^n \to R^m$ where A is a compact of R^n. By means of

ISBN: 978-989-95079-6-8

variable knowledge relevant to the unknown mapping, it is possible to prove that that fuzzy systems are dense in the space of continuous functions on a compact domain and so can approximate arbitrarily well any continuous function on a compact domain. The following are the main phases of a FES design ([11], [17]):

- Identification of the problem and choice of the type of FES, which best suits the problem requirement. A modular system can be designed. It consists of several fuzzy modules linked together. A modular approach may greatly simplify the design of the whole system, dramatically reducing its complexity and making it more comprehensible. Fixed input and output variables, it is necessary to describe their linguistic attributes (fuzzy values) and their membership function (fuzzification of input and output);
- definition of the set of heuristic fuzzy rules. (IF-THEN rules);
- choice of the fuzzy inference method (selection of aggregation operators for precondition and conclusion);
- translation of the fuzzy output in a crisp value (defuzzification methods);
- test of the fuzzy system prototype, drawing of the goal function between input and output fuzzy variables, change of membership functions and fuzzy rules if necessary, tuning of the fuzzy system, validation of results.

2.2 System structure

The structure of the Fuzzy Expert System (FES) we present is very complex and contains sixty initial inputs. We try in the space we have to explain how it is born. Preparing the questionnaire, the economics and sociology experts fix the macro-indicators that produce the final evaluation called "certification". The two are "Gender Equity"(GE) and "Gender SusTainability" (GT). The two are macro-aggregations may be split in other sub macro-indicators: GE sums up two: Equity in the Firm (EF) and Equal Opportunity (EO). They include information on what firm makes to treat woman at the same level of man in the policy of entrance, haw firm respects welfare laws and so on. The EF intermediate variable, we present in detail, contains information on the firm's gender fairness in terms of wages and career opportunities. Firms are assessed with respect to employment equity (measured by horizontal segregation in terms of area or type of job), equity in earnings (measured by the gender wage gap and the share of women and men in top positions as a gross index of vertical segregation) and the participation of women as trade unions representatives. The other macro-indicator GT contains several types of information: how firm takes into account work life balance policy (WLB) that is how the way working time is scheduled and its balance with other use of time related to the family. The Human Resource Management (HRM) variables contain information on firm's recruitment methods for entrance into work and on employees leaving or being fired by the firm. The third Safety (S) produces information on security policies that the firm adopts and on the awareness of gender specific health risks at work. Due to limits of space we are going to present more in depth the "Equity in the Firm (EF)". Women tend to be severely underrepresented in Italy in apical positions [13], [14] and their low presence in apical positions in the firm is considered in this system as a negative element in firm's assessment with regards to gender equity both because:

1) it can lead to a higher wage gap at the disadvantage of women;

2) it can be a sign of employment discrimination by gender; 3) it can be a signal of an environment not favourable to women progression of career;

4) it can be a sign of a human resources evaluation system that rewards less women's competences than men's.

The wage gap has been computed by the ratio of women total cost of labour to men's. This is a gross measure of wage gap that, in further research, can be computed as a net wage gap by taking into account employee's characteristics (like work experience, tenure, education, hours of work). We have then considered the bargaining power of women inside the firm by computing how many women are representatives in trade unions with respect to women employed in the firm. An underrepresentation of women in unions representatives contributes to decrease the value of the final outcome since we deem that their voice cannot be enough listened in bargaining on job conditions if they are underrepresented amongst employees representatives. Equal opportunity, EO, takes into account the existence of agreements with unions and public institutions in terms of equal opportunities (the highest mark has been given to enacted agreements and the lowest if there are no agreements and the firm does not

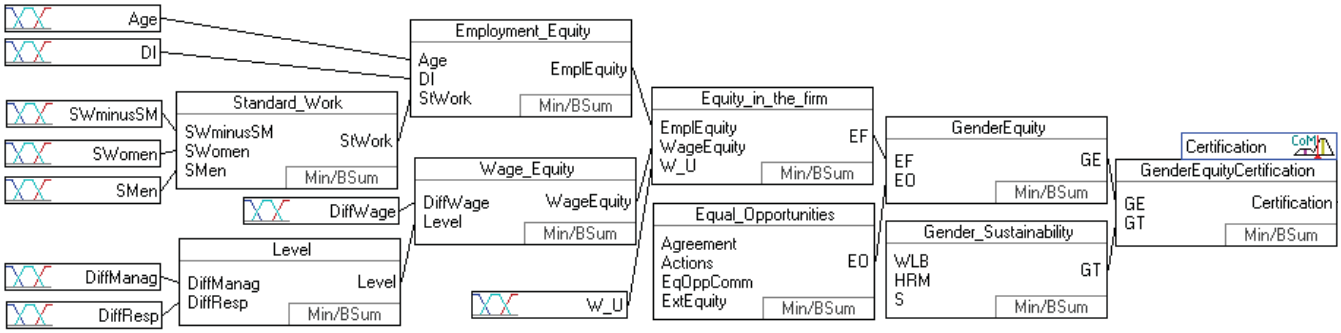


Figure 1: Reduced system layout

ISBN: 978-989-95079-6-8

intend to sign any in the future), of affirmative actions, of incentives to hire women,of the use of Italian Law 53/2000- that contains measures to increase work life balance in the firm - and equal opportunities and the presence of an Equal Opportunity Commission in the firm. We aim at improving the measurement of this last variable by collecting information on the type of activities performed by the Commission. This dimension contains also all strategies enacted by the firm to inform and make aware personnel about gender differences and all the actions that lead to an increase (in qualitative and quantitative terms) of women's participation in the firm. The equal opportunity dimension is also extended to include firm's awareness of the other firms of the network in which it operates.

2.3 Reduced system layout

We present the inputs involved in EF macro-indicator and the sub-system in which the variable GE is split in the original inputs.

Age. Age

SWminusSM. Gender gap in standard employment inside the firm

DI. Dissimilarity index defined as $\frac{1}{2}\sum_i \left|\frac{w_i}{w} - \frac{m_i}{m}\right|$ where m_i and w_i are the numbers of men and women in the i-th area.

SWomen. Women in standard jobs

SMen. Men in standard jobs

DiffMang. Gender gap in managerial position

DiffResp. Gender gap in high responsibility white-collar positions

DiffWage. Gender wage gap

W_U. Women share in union's representatives

Intermediate variables:

EmplEquity. Employment equity

EF. Equity in the firm

EO. Equal opportunities

GE. Gender equity

GT. Gender Sustainability

HRM. Human resources management

Level. Level in the job ladder as percentage of women in apical positions

S. Safety

StWork. Standard work

WageEquity. Wage equity

WLB. Work Life Balance

Now we present the fuzzyfied input variable DiffWage and a rule block in which this variable is involved. DiffWage measures the gross wage gap in the firm as the ratio of women's average labour cost with respect to men's. As it is possible to see, this variable enters directly at the second step and is aggregated with the intermediate variable Level. Level measures the share of women in apical positions with respect to men and can be considered as a measure of vertical segregation against women. Wage equity is then evaluated with respect to both variables to evaluate current wage gap and a cause of wage gap at disadvantage of women. All the cut points that define the fuzzifcation and the rules have been discussed in depth inside the research group where also the researchers who interviewed firm's representatives were involved.

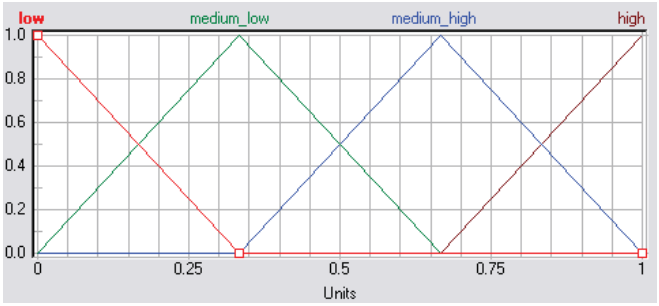


Figure 2: **DiffWage** variable layout

This research step can be deemed very important both in order to state the rules that are at the basis of evaluation in FES and in order to detect problems in the evaluation process and in the tools that have been used.

The following Rule Table is referred to Wage Equity: this intermediate variable is built with the input DiffWage and the intermediate variable Level.

Table 1: rule table of **WageEquity**

#	DiffRetrib	Level	WageEquity
1	low v medium_low	low	very_low
2	low	medium_low	very_low
3	low	medium	low
4	low	medium_high	medium_low
5	low	high	medium_high
6	medium_low	medium_low	low
7	medium_low	medium	medium_low
8	medium_low	medium_high	medium_high
9	medium_low v medium_high	high	high
10	medium_high	low	low
11	medium_high	medium_low	medium_low
12	medium_high	medium	medium_high
13	medium_high	medium_high	high
14	high	low	medium_low
15	high	medium_low	medium_high
16	high	medium	high
17	high	medium_high	very_high

The experts decided that if women's wage cost with respect to men's is low or medium and women's share in managerial and high responsibility white-collar positions are low the wage equity dimension assumes value 'very low'. The experts have suggested this rule that combines wage differentials and gender inequalities in access to apical positions since though the existence of a wage gap at the disadvantage of women can be related to vertical employment segregation one can observe wage differentials also in the same job position. Moreover the measure of the wage gap obtained from the sample of firms, given the structure of the questionnaire, is a gross wage gap (not corrected for employee's level of education or work experience). Inequalities in the access to apical positions can be considered as a proxy of vertical employment segregation by gender to which the experts' group assigns a high weigh in affecting firms' gender equity. This rule block shows that there is a function $f(DiffWage, Level): R^2 \rightarrow R$ increasing

ISBN: 978-989-95079-6-8

in all the two variables but they are weighted in a different way. Looking at rules 5 and 14 we have that: 5) *f(Low, High)→ medium_high;14) f(High, Low)→ medium_low* . This table translates in a simple way what the experts consider correct. This way to translate the experts' position is simpler than asking them numerical weights for the single variables and what type of function better represents their idea.

2.4 The sample of firms to which the system has been applied

The system introduced in this Section has been applied to the gender evaluation of 34 private firms of different size belonging to the different productive sectors that have been selected by the Italian Ministry of Labour through a public call. The sample was formed by: 14 joint-stock companies, 2 Ltds, 3 cooperatives Ltd, 10 social cooperatives and 5 others form of enterprise. According to the number of employees those firms can be classified as follows: 1 micro-enterprise (<10 employees), 12 small and medium enterprises (>10 employees <250), 21 big enterprises of which 4 employ more than 30.000 employees. 21 firms have got a turnover of max 50 millions euro (SMEs) while 5 companies exceed a turnover of 1 billion euro per year. Concerning the economic sectors, 17 firms belong to the service industry (research and training, welfare services, company services), 5 are utility companies (telecommunications and public transports), 3 of them belong to the pharmaceuticals industry, 3 to the retail trade, 1 is an energy producing industry and 1 is a bank. The last 4 firms of the sample belong to other sectors like: edutainment, food production and airport management. The selected firms filled in the self-assessment questionnaires and the sociologists conducted in-depth interviews to the relevant union representatives. The system has been presented and discussed inside a group of experts to whom the results of evaluation has also been submitted to validate the fuzzy outcome.

3 Results

We can start by observing that none of the 34 firms that applied to be evaluated and were selected in the Bollino Rosa project obtained a high result in terms of "Certification value", the final output of the FES presented in the previous section. The firm at the top position in terms reached a final evaluation of 0.68 against the maximum of 1 (Table 2) and the firm at the bottom has a value of 0.25, with 9 firms out of 34 with a value for the final outcome of the system lower than 0.40 (Table 3).At this stage of the project we cannot analyse in depth the situation of each firm in the sample and, for privacy reasons, we cannot provide more details about the firms involved in this experimental phase of the project. In the two tables we call them Ai and Ci if the firm is "not corporate" or "corporate" respectively. Table 2 shows the value of the intermediate and input variables of the first ten firms (in terms of the value of the final output of the system that we call certification) and Table 3 shows the last 10 firms. The tables show how similar output values can be derived by different situation in terms of the dimensions of gendered equity condensed in our intermediate variables. For instance the corporate firm C08 that performs as the not corporate firm A01 In terms of the final value of the FES output (0.62) shows a higher evaluation in terms of gender equity (GE=0.57) than in terms of gender sustainability (GT=0.65), whereas firm A01 performs better in terms of gender sustainability (GT=0.71) than in terms of gender equity (GE=0.53). The same result in terms of certification can therefore mean different outcomes in terms of policies suggestions to the firm in order to improve its 'Bollino Rosa' outcome. A further investigation on the reasons of a lower evaluation of firm A01 in terms of gender equity with respect to firm C08 reveals that this is connected to the high degree of vertical segregation in the firm (with no woman in apical positions) whereas the presence of women in apical position in firm C08 though lower (0.32) is higher than in firm A01. On the other hand both perform rather well in terms of gross wage gap.

Table 2: results the first ten best

Firm	DiffWage	W_U	GE	EF	EmplEquity	WageEquity	EO	Level	GT	Certification
A07	0,90	0,00	0,718	0,576	0,777	0,742	0,800	0,500	0,643	**0,67812**
A04	0,82	0,05	0,593	0,390	0,652	0,291	0,800	0,000	0,714	**0,65588**
A08	1,17	0,38	0,500	0,659	0,535	0,800	0,300	0,500	0,786	**0,65000**
A06	1,39	0,00	0,646	0,469	0,792	0,400	0,800	0,000	0,643	**0,63774**
C05	0,99	0,01	0,635	0,578	0,847	0,796	0,700	0,500	0,625	**0,63042**
A02	1,86	0,00	0,627	0,690	0,805	1,000	0,500	1,000	0,643	**0,62616**
A01	0,79	0,08	0,534	0,365	0,588	0,274	0,700	0,000	0,714	**0,62058**
C08	0,83	0,01	0,570	0,480	0,761	0,568	0,700	0,324	0,650	**0,62004**
C07	0,00	0,72	0,639	0,584	0,842	0,058	0,700	0,322	0,559	**0,60646**
A13	0,00	0,51	0,583	0,500	0,500	0,200	0,700	0,500	0,643	**0,60000**

Table 3: of results the last ten

Firm	DiffWage	W_U	GE	EF	EmplEquity	WageEquity	EO	Level	GT	Certification
A12	0,00	0,00	0,250	0,250	0,573	0,000	0,300	0,000	0,571	**0,40000**
A10	0,00	0,00	0,247	0,245	0,494	0,000	0,300	0,000	0,571	**0,39816**
A19	1,03	0,00	0,250	0,441	0,755	0,400	0,100	0,000	0,500	**0,35000**
A11	0,53	0,03	0,219	0,329	0,620	0,116	0,200	0,000	0,429	**0,32890**
A03	0,00	0,41	0,119	0,304	0,357	0,000	0,000	0,000	0,571	**0,32146**
A23	0,90	0,00	0,166	0,374	0,665	0,339	0,000	0,000	0,500	**0,31022**
A09	0,00	0,00	0,167	0,250	0,500	0,200	0,200	0,500	0,429	**0,30000**
A15	0,00	0,14	0,167	0,250	0,500	0,000	0,100	0,000	0,429	**0,30000**
A21	0,00	0,00	0,167	0,132	0,233	0,000	0,200	0,000	0,484	**0,30000**
A20	0,00	0,12	0,201	0,301	0,575	0,000	0,200	0,000	0,332	**0,24554**

Turning to the last ten firms in terms of the final assessment of gender certification, we can see that the bottom one performs very poorly both in terms of wage equity (0.00) and in terms of the equal opportunities (0.20) and gender sustainability (0.33) dimensions while it scores better (0.575) in terms of employment equity denoting a more equal presence of women and men in the areas of the firm and with regards to the type of contract. This firm is very far from a sufficient level and has to work a lot in all the dimensions to

ISBN: 978-989-95079-6-8

be certified as a 'Bollino Rosa' Firm.
The results shown in the previous section can be considered as a step in the whole process of certification that can use FES tools for assessing firms and addressing specific policies for each firm to reach a final assessment. The results of Fuzzy evaluation will be transmitted to the firms and commented upon by the relevant expert/auditor. According to the results obtained each firm will introduce the needed changes for correcting the gender discriminations. At the end of the evaluation path and only when the organizational situation within the firm will be improved, it will receive a Label that will identify it as a "Bollino Rosa firm", i.e. a gender friendly workplace in terms of equity and sustainability. The process towards certification of gender equity and sustainability involves also an assessment of the steps towards a higher values in the critical dimensions by means of the FES application to evaluate the changes occurred over time. Firms can therefore assess their level of 'Bollino Rosa' connected to the strategies enacted to improve its value.

4 Pursuing in the process of certification

The results shown in the previous section can be considered as a step in the whole process of certification that can use FES tools for assessing firms and addressing specific policies for each firm to reach a final assessment. The results of Fuzzy evaluation will be transmitted to the firms and commented upon by the relevant expert/auditor. According to the results obtained each firm will introduce the needed changes for correcting the gender discriminations. At the end of the evaluation path and only when the organizational situation within the firm will be improved, it will receive a Label that will identify it as a "Bollino Rosa firm", i.e. a gender friendly workplace in terms of equity and sustainability. The process towards certification of gender equity and sustainability involves also an assessment of the steps towards a higher values in the critical dimensions by means of the FES application to evaluate the changes occurred over time. Firms can therefore assess their level of 'Bollino Rosa' connected to the strategies enacted to improve its value. Further research involve considering how the level of gender equity and sustainability in the firm relates to the socioeconomic environment the firm is inserted in. As shown in the following chart the firm operates in connection to local, national government and European Commission that can set rules or adopt policies that are going to affect the firm's situation with regard to Gender Equity or Gender Sustainability. For instance the National Law 53/00 introduces the possibility for firms to receive funds dedicated to the implementation of family friendly policies that can improve gender sustainability in the firm. On the other hand the firm can supply childcare services for their employees but open to other families that can improve gender sustainability of employment in the area. Equal Opportunity Counsellor can take an action to remove discrimination in the firm. The firms that are suppliers or clients can be characterized by a different degree of gender equity and the workers in the firm are going to be affected by the equality or gender sustainability policies of the firm. Trade unions at local or at national level can bargain over working hours schedule to obtain a working time more that balances more with private and family life. The network of relations amongst private and public agents leading to the obtained results can itself be evaluated calling for an extension of the evaluation process to subcontractors, suppliers, clients. The system now includes firms' evaluation on the situation of clients and suppliers with regards to gender equity and this awareness is accounted for as a part of the gender equity dimension. However to have a more precise information on the interactions amongst firms in a gender perspective the whole network of firms should be reconstructed and subject to the evaluation process. Italy is characterized by a highly segmented labour market (by area and type of occupation) and by a different level of diffusion of public care services that can have an impact on gender equity and sustainability of work. Local labour market and institutional variables can be inserted to provide a comparative analysis upon different areas in Italy producing a more complex indicator of gender equity that can be inserted in the evaluation of gender human development.

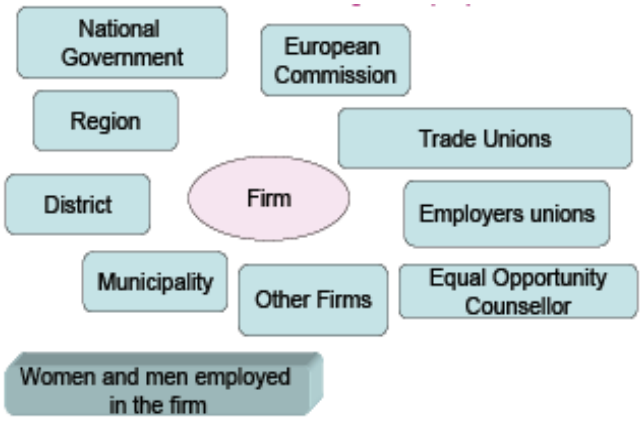


Figure 3: Toward firm's evaluation in a gendered perspective

5 Conclusions

The process of firm's evaluation in terms of gender equity pursued by the project 'Bollino Rosa' promoted by the Italian Ministry of Labour in year 2007 requires collecting and analysing indicators on different dimensions of gender equity. To provide a synthetic indicator on the firm's situation without loosing the complexity of the different dimensions of gender equity we have modelled and applied a Fuzzy Expert System. To our knowledge this is the first attempt of using Fuzzy Expert System to assess firms with respect to gender equity and gender sustainability, though broader issue of the gender perspective in the evaluation of quality of work using fuzzy expert system has been considered in [8]. Other experiences carried out in European countries implies self-evaluation like Total Equity Prize in Germany) or evaluation on the basis of the policies enacted by the firms in terms of equal opportunities, on their human resources management and on the presence of work life balance policies (like Label Egalité in France) however they do not use a system of evaluation structured as the one used

ISBN: 978-989-95079-6-8

in this application [5]. The micro data used have been collected through a structured questionnaire submitted to a sample of firms in the Private Sector that have proposed themselves by answering to a public competition on participation to the experiment, further extension to the public sector and a calibration of the questionnaire should follow its experimental phase. In this line we plan to let firms construct measures of net gender wage gap (by connecting wages to employees' individual characteristics on their work experience and education) that can improve the system of gender equity evaluation (currently based on gross wage gap). With respect to other systems of evaluation the proposed method entails an analytical model of evaluation that confers numerical indicators useful for the firm to assess its position as far as gender equity and sustainability are concerned allowing a backward process to detect causes of poor grades in terms of gender certification. Another advantage of the system is the transparency of the evaluation system reached by a clear statement of input variables and the rules used by the experts. Sometimes the transparency is not an advantage especially in situations in which politics is involved, but we think that a clear and transparent method is better than others in which the choices are expressed in "obscure" words in order to hide real choices from people involved. Public institutions and government can use the results of this model in terms of gender certification of firms to take policy decisions. They can also infer by comparative analyses on different areas how different policies and agreement carried out also at local level may interact with the gender assessment of firms in that area. Policies to consider firms' gender certification in gender equity and sustainability can also be used as an indicator for the assessment of firms when public institutions or firms contract out part of their production. This approach, more typical of engineering problems, finds here an interesting multi-disciplinary application. A very heterogeneous research unit containing mathematics, economics, sociology, political and trade union experts have faced this application. The fuzzy approach has given us all the possibility to work together, to have a common language. This language is due to the linguistic attributes of the variables, the rules given in a verbal way and the final interpretation of the results

Acknowledgment

The Authors thank G.Natoli and V.Cardinali for giving consent to the use of their contributions included in the volume "Strumenti per certificare e promuovere la parità di genere in azienda" [7]. The Authors thank the firms that have been involved in the experimental phase of the project and the Italian Ministry of Labour and other participants to Bollino Rosa project.

References

[1] Addabbo T.- Borghi, V. -Favaro, D."(2006) Differenze di genere nell'accesso a posizioni apicali. Risultati di una ricerca sul campo, in A. Simonazzi (ed.) Questioni di genere, questioni di politica. Trasformazioni economiche e sociali in una prospettiva di genere, Roma, Carocci, 2006.

[2] Addabbo T. – Facchinetti G. - Mastroleo G. (2005). "A fuzzy expert system to measure functionings. An application to child wellbeing". In K. Saeed, R. Mosdorf, J. Pejas, P. Hilmola, Z. Sosnowsky (eds) Proceeding of Image Analysis, Computer Graphics, Security Systems and Artificial Intelligence Applications. Bialystok Vol. 1, 29-42.

[3] Addabbo, T.: Unpaid work by gender in Italy,Ch.2 (2003) in'Unpaid work and the economy', A.Picchio (ed.) Routledge,London and New York (2003)

[4] Addabbo, T. - Di Tommaso, M.L. -Facchinetti, G.(2004): To what extent fuzzy set theory and structural equation modelling can measure functionings? An application to child well being. Materiali di Discussione del Dipartimento di Economia Politica n.468, Modena (2004)

[5] Bassanini, C. (2008) 'Modelli e buone pratiche per la certificazione di genere nelle imprese: un confronto internazionale', in Príncipe G. et al.,(2008). Strumenti per certificare e promuovere la parità di genere in azienda, ISFOL, 2008, Ch.10.

[6] Burchell, B., Fagan, C., O'Brien, C. and Smith, M. (2007) Working conditions in the European Union: The gender perspective, Dublin, European Foundation for the Improvement of living and working conditions.

[7] Addabbo T. – Facchinetti G -Mastroleo G. - Solinas G. (2006). "A fuzzy way to Measure Quality of Work in a multidimensional view" Plenary lecture in the ACS Conference Poland 0ctober 2006. Jerzy Pejas, Imed El Fray, Khalid Saeed Editors Vol I, 13-24. ISBN 83-87362-75-

[8] Addabbo T. – Facchinetti G -Mastroleo G. (2007). "Capability and functiongs: A Fuzzy Way to measure interaction between Father and Child" in Biometrics, Computer Security Systems and Artificial Intelligence Applications. Khalid Saeed, Jerzy Pejas, Romuald Mosdof Editors 185-195, Springer, ISBN 978-0-387-36232-8 (Print) 978-0-387-36503-9 (Online).

[9] Facchinetti G. –Franci F.- Mastroleo -G. Pagliaro V.-Ricci G. (2007) Illogica di un conflitto. La logica fuzzy applicata al conflitto tra Israele e Libano, Eurilink editor ISBN 978-88-95151-04-5

[10] Picchio A. (ed.) (2003) Unpaid work and the economy London and New York: Routledge.

[11] Piegat A. (2001), Fuzzy modelling and control. Springer-Verlag, Heidelberg-New York.

[12] Príncipe G. et al. (2008). Strumenti per certificare e promuovere la parità di genere in azienda, ISFOL, 2008.

[13] Rosti, L.(2006) La segregazione occupazionale in Italia, in A. Simonazzi (ed.) Questioni di genere, questioni di politica. Trasformazioni economiche e sociali in una prospettiva di genere, Roma, Carocci, 2006.

[14] Sabbadini L.L. (2004) Come cambia la vita delle donne, Ministero per le Pari Opportunità, 2004.

[15] Shun-Hsien L. (2005) Expert System Methodologies and application- a decade review from 1995 to 2004. Expert Systems with application. Vol 28 Isuue 1 93-103.

[16] Wang L., (1992), Fuzzy systems are universal approximators Proc. Of Int. Conf. On Fuzzy Engineering, 471-496, 1992.

[17] Von Altrock C. (1997), Fuzzy Logic and NeuroFuzzy applications in Business and Finance, Prentice-Hall Inc.

Morphisms in categories of sets with similarity relations

Jiří Močkoř

Institute for Research and Applications of Fuzzy Modelling, University of Ostrava
Ostrava, Czech Republic
Email: Jiri.Mockor@osu.cz

Abstract— *Morphisms of some categories of sets with similarity relations (Ω-sets) are investigated, where Ω is a complete residuated lattice. Namely a category $\mathbf{SetF}(\Omega)$ with morphisms $(A,\delta)\to(B,\gamma)$ defined as special maps $A\to B$ and a category $\mathbf{SetR}(\Omega)$ with morphisms defined as a special relations $A\times B\to\Omega$. It is proved that arbitrary maps $A\to\Omega$ and $A\times B\to\Omega$ can be extended onto morphisms $(A,\delta)\to(\Omega,\leftrightarrow)$ in $\mathbf{SetF}(\Omega)$ and morphisms $(A,\delta)\to(B,\gamma)$ in $\mathbf{SetR}(\Omega)$, respectively and that these extension processes are from categorical point of view special reflections. Moreover, if Ω is a complete Heyting algebra we also investigate morphisms in another category $\mathbf{Set}(\Omega)$ which consists of classical fuzzy sets $f:A\to\Omega$.*

Keywords— residuated lattice, Heyting algebra, similarity relations, fuzzy sets.

1 Introduction

In fuzzy set theory the category $\mathbf{SetF}(\Omega)$ of sets with similarity relations defined over a complete residuated lattice $\Omega=(\Omega,\to,\otimes,\vee,\wedge)$ is of principal importance ([5]-[10]). This category consists of objects (A,δ) (called Ω-sets), where A is a set and δ is a similarity relation, i.e. a map $\delta:A\times A\to\Omega$ such that

(a) $\delta(x,x)=1$,

(b) $\delta(x,y)=\delta(y,x)$,

(c) $\delta(x,y)\otimes(\delta(y,y)\to\delta(y,z))\le\delta(x,z)$.

A morphism $f:(A,\delta)\to(B,\gamma)$ in $\mathbf{SetF}(\Omega)$ is a map $f:A\to B$ such that $\gamma(f(x),f(y))\ge\delta(x,y)$ for all $x,y\in A$.

From historical point of view there is another category consisting of sets with similarity relations defined. This category $\mathbf{SetR}(\Omega)$ is an analogy of a category of sets with relations between sets as morphisms. Objects of this category $\mathbf{SetR}(\Omega)$ are the same as in the category $\mathbf{SetF}(\Omega)$ and morphism $f:(A,\delta)\to(B,\gamma)$ are maps $f:A\times B\to\Omega$ such that

(a) $(\forall x,z\in A)(\forall y\in B)\quad\delta(z,x)\otimes f(x,y)\le f(z,y)$,

(b) $(\forall x\in A)(\forall y,z\in B)\quad f(x,y)\otimes\gamma(y,z)\le f(x,z)$,

If $f:(A,\delta)\to(B,\gamma)$ and $g:(B,\gamma)\to(C,\omega)$ are two morphisms then their composition is a relation $g\circ f:A\times C\to\Omega$ such that

$$g\circ f(x,z)=\bigvee_{y\in B}(f(x,y)\otimes g(y,z)).$$

Finally we can consider a classical category $\mathbf{Set}(\Omega)$ of fuzzy sets over Ω with objects couples (A,f), where A is a set and f is a map $A\to\Omega$ and with morphisms $u:(A,f)\to(B,g)$ such that $u:A\to B$ is map and $f(a)\le g\circ u(a)$ for all $a\in A$.

With categories $\mathbf{SetF}(\Omega)$ and $\mathbf{SetR}(\Omega)$ we can consider new sets of objects:

(a) the set $\mathbf{Mor}(\mathbf{SetR}(\Omega))$ of all morphisms in $\mathbf{SetR}(\Omega)$,

(b) the set $\mathbf{Rel}(\mathbf{SetR}(\Omega))$ of all *weak morphisms* in $\mathbf{SetR}(\Omega)$, where $f:(A,\delta)\to(B,\gamma)$ is a weak morphism in $\mathbf{SetR}(\Omega)$ if $f:A\times B\to\Omega$ is an Ω-valued relation,

(c) the set $\mathbf{Map}(\mathbf{SetF}(\Omega))$ of all *weak fuzzy sets* in $\mathbf{SetF}(\Omega)$, where $f:(A,\delta)\to(\Omega,\leftrightarrow)$ is a weak fuzzy set in $\mathbf{SetF}(\Omega)$ if $f:A\to\Omega$ is a map and where $\leftrightarrow$ is a biresiduation in a lattice Ω.

(d) the set $\mathbf{Fuz}(\mathbf{SetF}(\Omega))$ of all *fuzzy sets* in $\mathbf{SetF}(\Omega)$, i.e. morphisms $(A,\delta)\to(\Omega,\leftrightarrow)$ in the category $\mathbf{SetF}(\Omega)$, where (A,δ) is an Ω-set.

It is clear that we have

$$\mathbf{Mor}(\mathbf{SetR}(\Omega))\hookrightarrow\mathbf{Rel}(\mathbf{SetR}(\Omega)),$$
$$\mathbf{Fuz}(\mathbf{SetF}(\Omega))\hookrightarrow\mathbf{Map}(\mathbf{SetF}(\Omega)).$$

In this paper we want to consider some relationships between these sets. We show that new categories (denoted by the same symbols as the above sets) can be defined with the above sets as object sets. Then we will receive the following functor commutative diagram:

$$\begin{array}{ccc}\mathbf{Fuz}(\mathbf{SetF}(\Omega)) & \xrightarrow{\ \hookrightarrow\ } & \mathbf{Map}(\mathbf{SetF}(\Omega))\\ {\scriptstyle F}\downarrow & & \downarrow{\scriptstyle F}\\ \mathbf{Mor}(\mathbf{SetR}(\Omega)) & \xrightarrow{\ \hookrightarrow\ } & \mathbf{Rel}(\mathbf{SetR}(\Omega))\end{array}$$

Finally, we show that

(a) the category $\mathbf{Mor}(\mathbf{SetR}(\Omega))$ is a full reflective subcategory in the category $\mathbf{Rel}(\mathbf{SetR}(\Omega))$,

(b) the category $\mathbf{Map}(\mathbf{SetF}(\Omega))$ is a full reflective subcategory in the category $\mathbf{Fuz}(\mathbf{SetF}(\Omega))$.

The reflections $f\to\tilde{f}$ and $g\to\widehat{g}$ of a weak morphism $f:(A,\delta)\to(B,\gamma)$ and a weak fuzzy set $g:(A,\delta)\to(\Omega,\leftrightarrow)$ can be then used for construction of a fuzzy logic formulas interpretation in corresponding categories $\mathbf{SetF}(\Omega)$ and $\mathbf{SetR}(\Omega)$.

ISBN: 978-989-95079-6-8

2 Categories of maps and relations

Let us firstly define two categories $\mathbf{Rel}(\mathbf{SetR}(\Omega))$ and $\mathbf{Mor}(\mathbf{SetR}(\Omega))$. Objects of the category $\mathbf{Rel}(\mathbf{SetR}(\Omega))$ are all weak morphisms from $\mathbf{SetR}(\Omega)$. A morphism from a weak morphism $f : (A,\delta) \to (B,\gamma)$ to a weak morphism $g : (C,\rho) \to (D,\omega)$ is a couple (u,v) of morphisms from the category $\mathbf{SetR}(\Omega)$ such that

(i) $u : (A,\delta) \to (C,\rho)$ is a morphism in $\mathbf{SetR}(\Omega)$,

(ii) $v : (B,\gamma) \to (D,\omega)$ is a morphism in $\mathbf{SetR}(\Omega)$,

(iii) $(v \circ f)(a,d) \le (g \circ u)(a,d)$ for all $a \in A, d \in D$, where the composition of a morphism and a weak morphism is formally the same as for morphisms in the category $\mathbf{SetR}(\Omega)$.

In that case we say that the diagram

$$\begin{array}{ccc} (A,\delta) & \xrightarrow{f} & (B,\gamma) \\ {\scriptstyle u}\downarrow & & \downarrow{\scriptstyle v} \\ (C,\rho) & \xrightarrow{g} & (D,\omega) \end{array}$$

fuzzy commutes. If $(u,v) : f \to g$ and $(u_1,v_1) : g \to h$ are two morphisms in $\mathbf{Rel}(\mathbf{SetR}(\Omega))$ such that $f : (A,\delta) \to (B,\gamma)$, $g : (C,\rho) \to (D,\omega)$, $h : (E,\sigma) \to (F,\tau)$ are weak morphisms then a composition $(u_1,v_1) \circ (u,v)$ is defined as $(u_1 \circ u, v_1 \circ v)$. Using a composition of weak morphisms we can easily prove that the following diagram fuzzy commutes

$$\begin{array}{ccc} (A,\delta) & \xrightarrow{f} & (B,\gamma) \\ {\scriptstyle u_1\circ u}\downarrow & & \downarrow{\scriptstyle v_1\circ v} \\ (E,\sigma) & \xrightarrow{h} & (F,\tau), \end{array}$$

i.e. the definition is correct.

Objects of the category $\mathbf{Mor}(\mathbf{SetR}(\Omega))$ are all morphisms $f : (A,\delta) \to (B,\gamma)$ in the category $\mathbf{SetR}(\Omega)$ and morphisms between such objects are defined formally in the same way as for the category $\mathbf{Rel}(\mathbf{SetR}(\Omega))$. It is clear that $\mathbf{Mor}(\mathbf{SetR}(\Omega))$ is a full subcategory of the category $\mathbf{Rel}(\mathbf{SetR}(\Omega))$.

Now, objects of the category $\mathbf{Map}(\mathbf{SetF}(\Omega))$ are all weak fuzzy sets $(A,\delta) \to (\Omega,\leftrightarrow)$ in the category $\mathbf{SetF}(\Omega)$, where (A,δ) are Ω-sets. If $f : (A,\delta) \to (\Omega,\leftrightarrow)$ and $g : (B,\gamma) \to (\Omega,\leftrightarrow)$ are objects in $\mathbf{Map}(\mathbf{SetF}(\Omega))$ then $u : f \to g$ is a morphism in $\mathbf{Map}(\mathbf{SetF}(\Omega))$ if $u : (A,\delta) \to (B,\gamma)$ is a morphism in $\mathbf{SetF}(\Omega)$ and $g \circ u(a) \ge f(a)$ for all $a \in A$. A composition of such morphisms is defined as a composition of corresponding morphisms in the category $\mathbf{SetF}(\Omega)$.

Finally objects of the category $\mathbf{Fuz}(\mathbf{SetF}(\Omega))$ are all fuzzy sets in the category $\mathbf{SetF}(\Omega)$, i.e. morphisms $(A,\delta) \to (\Omega,\leftrightarrow)$ in the category $\mathbf{SetF}(\Omega)$. Morphisms in the category $\mathbf{Fuz}(\mathbf{SetF}(\Omega))$ are defines similarly as in the category $\mathbf{Map}(\mathbf{SetF}(\Omega))$.

It is well known that there exists a functor $F : \mathbf{SetF}(\Omega) \to \mathbf{SetR}(\Omega)$ such that F is an identity on objects and if $f : (A,\delta) \to (B,\gamma)$ is a morphism in $\mathbf{SetF}(\Omega)$ then $F(f) : A \times B \to \Omega$ is such that $F(f)(a,b) = \gamma(f(a),b)$.

Lemma 1
Let $g : (A,\delta) \to (B,\gamma)$ be a weak morphism in $\mathbf{SetR}(\Omega)$. Let $\tilde{g} : A \times B \to \Omega$ be defined by the formula

$$\tilde{g}(a,b) = \bigvee_{x\in A}\bigvee_{y\in B} g(x,y) \otimes \delta(a,x) \otimes \gamma(b,y).$$

Then

(a) $\tilde{g}$ is a morphism in $\mathbf{SetR}(\Omega)$,

(b) $\tilde{g} = \bigwedge\{h : h$ is a morphism $(A,\delta) \to (B,\gamma)$ in $\mathbf{SetR}(\Omega), h \ge g\}$,

(c) If g is a morphism in $\mathbf{SetR}(\Omega)$, then $\tilde{g} = g$.

Lemma 2
Let $s : (A,\delta) \to (\Omega,\leftrightarrow)$ be a weak fuzzy set in $\mathbf{SetF}(\Omega)$. Let a map $\widehat{s} : A \to \Omega$ be defined such that $\widehat{s}(a) = \bigvee_{x\in A} \delta(a,x) \otimes s(x)$ for all $a \in A$. Then

(a) $\widehat{s} : (A,\delta) \to (\Omega,\leftrightarrow)$ is a morphism in $\mathbf{SetF}(\Omega)$,

(b) $\widehat{s} = \bigwedge\{t : t$ is a morphism $(A,\delta) \to (\Omega,\leftrightarrow)$ in $\mathbf{SetF}(\Omega), t \ge s\}$.

(c) If $s : (A,\delta) \to (\Omega,\leftrightarrow)$ is a fuzzy set in $\mathbf{SetF}(\Omega)$ then $\widehat{s} = s$,

Now
we define functors $G : \mathbf{Rel}(\mathbf{SetR}(\Omega)) \to \mathbf{Mor}(\mathbf{SetR}(\Omega))$ and $H : \mathbf{Map}(\mathbf{SetF}(\Omega)) \to \mathbf{Fuz}(\mathbf{SetF}(\Omega))$ which will be reflections.

Proposition 1
(a) Let a weak morphism f be an object from $\mathbf{Rel}(\mathbf{SetR}(\Omega))$ and let $(u,v) : f \to g$ be a morphism in $\mathbf{Rel}(\mathbf{SetR}(\Omega))$. Let $G(f) = \tilde{f}$ and $G(u,v) = (u,v)$. Then

$$G : \mathbf{Rel}(\mathbf{SetR}(\Omega)) \to \mathbf{Mor}(\mathbf{SetR}(\Omega))$$

is a functor.

(b) Let a weak fuzzy set f be an object in $\mathbf{Map}(\mathbf{SetF}(\Omega))$ and let $u : f \to g$ be a morphism in $\mathbf{Map}(\mathbf{SetF}(\Omega))$. Let $H(f) = \widehat{f}$ and $H(u) = u$. Then

$$H : \mathbf{Map}(\mathbf{SetF}(\Omega)) \to \mathbf{Fuz}(\mathbf{SetF}(\Omega))$$

is a functor.

Proof. (a) Let $f : (A,\delta) \to (B,\gamma)$ and $g : (C,\rho) \to (D,\tau)$ be weak morphisms. We show that (u,v) is a morphism $\tilde{f} \to \tilde{g}$. Since $(u,v) : f \to g$ is a morphism in $\mathbf{Rel}(\mathbf{SetR}(\Omega))$, we have $v \circ f \le g \circ u$. Then since v and u are morphisms in $\mathbf{SetR}(\Omega)$, we have

$$\begin{aligned} v \circ \tilde{f}(a,d) = \\ \bigvee_{b,b',a'} f(a',b') \otimes \delta(a,a') \otimes \gamma(b,b') \otimes v(b,d) \le \\ \bigvee_{b',a'} f(a',b') \otimes \delta(a,a') \otimes v(b',d) = \\ \bigvee_{a'} v \circ f(a',d) \otimes \delta(a,a') \le \\ \bigvee_{a'} g \circ u(a',d) \otimes \delta(a,a') \le \bigvee_{c} u(a,c) \otimes g(c,d) \le \\ \tilde{g} \circ u(a,d). \end{aligned}$$

ISBN: 978-989-95079-6-8

(b) Let $f : (A,\delta) \to (\Omega,\leftrightarrow)$ and $g : (B,\gamma) \to (\Omega,\leftrightarrow)$ be weak fuzzy sets. We need to show that $\widehat{g} \circ u \geq \widehat{f}$. Since u is a morphism in $\mathbf{SetF}(\Omega)$, we have

$$\widehat{g} \circ u(a) = \bigvee_{b\in B} g(b) \otimes \gamma(b,u(a)) \geq \bigvee_{x\in A} g(u(x)) \otimes \gamma(u(x),u(a)) \geq \bigvee_{x\in A} g \circ u(x) \otimes \delta(x,a) \geq \bigvee_{x\in A} f(x) \otimes \delta(x,a) = \widehat{f}(a).$$

We now introduce a notion of a fuzzy reflective subcategory.

Theorem 1

(a) $\mathbf{Mor}(\mathbf{SetR}(\Omega))$ *is a full reflective subcategory in the category* $\mathbf{Rel}(\mathbf{SetR}(\Omega))$ *and* G *is a reflection.*

(b) $\mathbf{Map}(\mathbf{SetF}(\Omega))$ *is a full reflective subcategory in the category* $\mathbf{Fuz}(\mathbf{SetF}(\Omega))$ *and* H *is a reflection.*

Proof. (a) Let $f : (A,\delta) \to (B,\gamma)$ be an object in $\mathbf{Rel}(\mathbf{SetR}(\Omega))$. It is clear that $(id_A, id_B) : f \to \tilde{f}$ is a morphism in $\mathbf{Rel}(\mathbf{SetR}(\Omega))$, where for an object (X,σ) from $\mathbf{SetR}(\Omega)$ the identity morphism $id_X : (X,\sigma) \to (X,\sigma)$ is defined such that $id_X : X \times X \to \Omega$ is such that $id_X(x,y) = 1_\Omega$ if $x = y$ and 0_Ω, otherwise. Let $g : (C,\rho) \to (D,\tau)$ be an object in $\mathbf{Mor}(\mathbf{SetR}(\Omega))$ and let $(u,v) : f \to g$ be a morphism in $\mathbf{Rel}(\mathbf{SetR}(\Omega))$. Then we show that (u,v) is the unique morphism such that the diagram commutes:

$$\begin{array}{ccc} f & \xrightarrow{(id_A,id_B)} & \tilde{f} \\ \| & & \downarrow{\scriptstyle (u,v)} \\ f & \xrightarrow{(u,v)} & g \end{array}$$

In fact, it suffices to prove that $(u,v) : \widetilde{f} \to g$ is a morphism in $\mathbf{Rel}(\mathbf{SetR}(\Omega))$, i.e. that the diagram

$$\begin{array}{ccc} (A,\delta) & \xrightarrow{\widetilde{f}} & (B,\gamma) \\ {\scriptstyle u}\downarrow & & \downarrow{\scriptstyle v} \\ (C,\rho) & \xrightarrow{g} & (D,\tau) \end{array}$$

fuzzy commutes. We have $v \circ f(a,d) \leq g \circ u(a,d)$ for all $a \in A, d \in D$. Since u and v are morphisms in a category $\mathbf{SetR}(\Omega)$, we have

$$v \circ \tilde{f}(a,d) = \bigvee_{b,b'\in B, a'\in A} f(a',b') \otimes \delta(a,a') \otimes \gamma(b,b') \otimes v(b,d) \leq \bigvee_{b,b'\in B a'\in A} f(a',b') \otimes \delta(a,a') \otimes v(b',d) = \bigvee_{a'\in A} \delta(a,a') \otimes v \circ f(a',d) \leq \bigvee_{a'\in A} \delta(a,a') \otimes g \circ u(a',d) \leq g \circ u(a,d).$$

It is clear that (u,v) is the unique morphism with such property and that the diagram commutes.

(b) Let $f : (A,\delta) \to (\Omega,\leftrightarrow)$ be an object of $\mathbf{Map}(\mathbf{SetF}(\Omega))$. It is clear that $id_A : f \to \widehat{f}$ is a morphism in $\mathbf{Map}(\mathbf{SetF}(\Omega))$. Let $g : (B,\gamma) \to (\Omega,\leftrightarrow)$ be an object in $\mathbf{Fuz}(\mathbf{SetF}(\Omega))$ and let $u : f \to g$ be a morphism in $\mathbf{Map}(\mathbf{SetF}(\Omega))$. Then u is the unique morphism in $\mathbf{Fuz}(\mathbf{SetF}(\Omega))$ such that the diagram commutes:

$$\begin{array}{ccc} f & \xrightarrow{id_A} & \widehat{f} \\ {\scriptstyle u}\downarrow & & \downarrow{\scriptstyle u} \\ g & = & g. \end{array}$$

In fact, since $g \circ u(a) \geq f(a)$ and $g \circ u$ is an object in $\mathbf{Fuz}(\mathbf{SetF}(\Omega))$, we have

$$\widehat{f}(a) = \bigvee_{x\in A} \delta(a,x) \otimes f(x) \leq \bigvee_{x\in A} \delta(a,x) \otimes g \circ u(x) \leq g \circ u(a).$$

3 Reflections and fuzzy logic models

The reflections G, H can be used for a definition of fuzzy logic formulas interpretation in categories $\mathbf{SetR}(\Omega)$ and $\mathbf{SetF}(\Omega)$, respectively (see [10]). Let J be a first order language of a fuzzy logic which consists (as classically) of a set of predicate symbols $P \in \mathscr{P}$, a set of functional symbols $f \in \mathscr{F}$ and a set of classical logical connectives $\{\wedge, \vee, \Rightarrow, \neg, \otimes\}$. Moreover J contains also a set Ω of logical constants.

Definition 1

Let $\mathbf{K}$ *be a category with products and with* Ω*-sets as objects. Then a model of a language* J *in a category* $\mathbf{K}$ *is*

$$\mathscr{D} = ((A,\delta), \{P_{\mathscr{D}} : P \in \mathscr{P}\}, \{f_{\mathscr{D}} : f \in \mathscr{F}\}),$$

where

(a) (A,δ) *is an* Ω*-set from a category* $\mathbf{K}$*,*

(b) $P_{\mathscr{D}} : (A,\delta) \times \cdots \times (A,\delta) \to (\Omega,\leftrightarrow)$ *is a morphism in* $\mathbf{K}$*,*

(c) $f_{\mathscr{D}} : (A,\delta) \times \cdots \times (A,\delta) \to (A,\delta)$ *is a morphism in a category* $\mathbf{K}$*.*

Further, let ψ (t, respectively) be a formula (term, respectively) with free variables contained in a set X of variables. Then an interpretation $\|\psi\|_{\mathscr{D},X}$ ($\|t\|_{\mathscr{D},X}$, respectively) of ψ (t, respectively) in a model $\mathscr{D}$ in a category $\mathbf{K}$ should be defined such that

(a) $\|\psi\|_{\mathscr{D},X} : (A,\delta)^X \to (\Omega,\leftrightarrow)$ is a morphism in $\mathbf{K}$,

(b) $\|t\|_{\mathscr{D},X} : (A,\delta)^X \to (A,\delta)$ is a morphism in $\mathbf{K}$,

where $(A,\delta)^X$ is a product $(A,\delta)^{|X|} = (A^{|X|}, \delta_X)$ in a category $\mathbf{K}$. We show shortly how by using reflections H and G the definition of a formula interpretation can be done in models $\mathscr{D}$ in categories $\mathbf{K} = \mathbf{SetF}(\Omega), \mathbf{SetR}(\Omega)$.

Let $\mathscr{D} = ((A,\delta), \{P_{\mathscr{D}} : P \in \mathscr{P}\}, \{f_{\mathscr{D}} : f \in \mathscr{F}\})$ be a model of a language J in a category $\mathbf{K}$, where $\mathbf{K} = \mathbf{SetF}(\Omega)$ or $\mathbf{K} = \mathbf{SetR}(\Omega)$, i.e.

ISBN: 978-989-95079-6-8

(i) (A,δ) is a Ω-set from $\mathbf{K}$,

(ii) $P_{\mathscr{D}} : (A,\delta)\times\cdots\times(A,\delta)\to(\Omega,\leftrightarrow)$ is a morphism in a category $\mathbf{K}$,

(iii) $f_{\mathscr{D}} : (A,\delta)\times\cdots\times(A,\delta)\to(A,\delta)$ is a morphism in a category $\mathbf{K}$.

Let t be a term with a set of variables contained in a set X. Then $\|t\|_{\mathscr{D},X} = \|t\|_X : (A,\delta)^X \to (A,\delta)$ is a morphism in $\mathbf{K}$ defined as follows.

(i) Let $t = x$, where $x \in X$. Then $\|t\|_X := pr_x : (A,\delta)^X \to (A,\delta)$ is a projection morphism in a category $\mathbf{K}$.

(ii) Let $t = f(t_1,\dots,t_n)$. Then $\|t\|_X$ is a composition (in $\mathbf{K}$) of morphisms

$$(A,\delta)^X \xrightarrow{\prod_i \|t_i\|_X} (A,\delta)^n \xrightarrow{f_{\mathscr{D}}} (A,\delta)\cdot$$

Hence, for $\mathbf{K} = \mathbf{SetR}(\Omega)$ we have $\|t\|_X(\mathbf{a},b) = \bigvee_{\mathbf{x}\in A^n}(\prod_i \|t_i\|_X(\mathbf{a},\mathbf{x})\otimes f_{\mathscr{D}}(\mathbf{x},b)$, where $\mathbf{a}\in A^X$.

Now let ψ be a formula with free variables contained in a set X. In a category $\mathbf{K}$ we will define firstly a weak morphisms (i.e. *weak fuzzy set* for $\mathbf{K} = \mathbf{SetF}(\Omega)$ or a *weak morphism* for $\mathbf{K} = \mathbf{SetR}(\Omega)$) $|\psi|_{\mathscr{D},X} = |\psi|_X : (A,\delta)^X \to (\Omega,\leftrightarrow)$. A definition will be done by induction principle on a structure of ψ.

(a) Let $\psi \equiv P(t_1,\dots,t_n)$. Then (according to induction assumption) $\|t_i\|_X, P_{\mathscr{D}}$ are defined and we define $|\psi|_X$ as a composition of the following morphisms in $\mathbf{K}$:

$$(A,\delta)^X \xrightarrow{\prod_i \|t_i\|_X} (A,\delta)^n \xrightarrow{P_{\mathscr{D}}} (\Omega,\leftrightarrow).$$

(b) Let $\psi \equiv t_1 = t_2$. Then $|\psi|_X$ is a composition of the following weak morphisms:

$$(A,\delta)^X \xrightarrow{\|t_1\|_X\times\|t_2\|_X} (A,\delta)^2 \xrightarrow{\Delta_{\mathbf{K},A}} (\Omega,\leftrightarrow).$$

(c) Let $\psi \equiv \psi_1\wedge\psi_2$. Then $|\psi|_X$ is a composition of the following weak morphisms:

$$(A,\delta)^X \xrightarrow{\|\psi\|_X\times\|\psi_2\|_X} (\Omega,\leftrightarrow)^2 \xrightarrow{\sqcap_{\mathbf{K}}} (\Omega,\leftrightarrow).$$

(d) Let $\psi \equiv \psi_1\vee\psi_2$. Then $|\psi|_X$ is a composition of the following weak morphisms:

$$(A,\delta)^X \xrightarrow{\|\psi\|_X\times\|\psi_2\|_X} (\Omega,\leftrightarrow)^2 \xrightarrow{\sqcup_{\mathbf{K}}} (\Omega,\leftrightarrow).$$

(e) Let $\psi \equiv \sigma\Rightarrow\tau$. Then $|\psi|_X$ is a composition of the following weak morphisms:

$$(A,\delta)^X \xrightarrow{\|\sigma\|_X\times\|\tau\|_X} (\Omega,\leftrightarrow)^2 \xrightarrow{\Rrightarrow_{\mathbf{K}}} (\Omega,\leftrightarrow).$$

(f) Let $\psi \equiv \neg\sigma$. Then $|\psi|_X$ is a composition of the following weak morphisms:

$$(A,\delta)^X \xrightarrow{\|\sigma\|_X} (\Omega,\leftrightarrow) \xrightarrow{\neg_{\mathbf{K}}} (\Omega,\leftrightarrow).$$

(g) Let $\psi \equiv (\exists x)\sigma$. Then $\|\sigma\|_{X\cup\{x\}}$ is already defined as a morphism $(A,\delta)^{X\cup\{x\}} = (A,\delta)^X\times(A,\delta)\to(\Omega,\leftrightarrow)$ in $\mathbf{K}$. Then we set

$$|\psi|_X(\mathbf{a},\alpha) = \bigvee_{x\in A}\|\sigma\|_{X\cup\{x\}}((\mathbf{a},x),\alpha), \quad \text{if } \mathbf{K} = \mathbf{SetR}(\Omega),$$
$$|\psi|_X(\mathbf{a}) = \bigvee_{x\in A}\|\sigma\|_{X\cup\{x\}}(\mathbf{a},x), \quad \text{if } \mathbf{K} = \mathbf{SetR}(\Omega).$$

The maps $\Rrightarrow_{\mathbf{K}}, \neg_{\mathbf{K}}, \sqcup_{\mathbf{K}}, \sqcap_{\mathbf{K}}, \Delta_{\mathbf{K},A}$ are defined as follows: For $\mathbf{K} = \mathbf{SetR}(\Omega)$, we have

(i) $\Delta_{\mathbf{SetR}(\Omega),A}((a,b),\alpha) = \alpha \leftrightarrow \delta(a,b)$ for all $a,b\in A, \alpha\in\Omega$.

(ii) $\sqcap_{\mathbf{SetR}(\Omega)}((\beta,\gamma),\alpha) = \alpha\leftrightarrow(\beta\wedge\gamma)$.

(iii) $\sqcup_{\mathbf{SetR}(\Omega)}((\beta,\gamma),\alpha) = \alpha\leftrightarrow(\beta\vee\gamma)$.

(iv) $\Rrightarrow_{\mathbf{SetR}(\Omega)}((\beta,\gamma),\alpha) = (\beta\otimes\alpha)\to\gamma$.

(v) $\neg_{\mathbf{SetR}(\Omega)}(\alpha,\beta) = \beta\leftrightarrow(\alpha\to 0)$.

For $\mathbf{K} = \mathbf{SetF}(\Omega)$, we have

(a) $\Delta_{\mathbf{SetF}(\Omega),A}(a,b) = \delta(a,b)$ for all $a,b\in A$.

(b) $\sqcap_{\mathbf{SetF}(\Omega)}(\beta,\gamma) = \beta\wedge\gamma$.

(c) $\sqcup_{\mathbf{SetF}(\Omega)}(\beta,\gamma) = \beta\vee\gamma$.

(d) $\Rrightarrow_{\mathbf{SetF}(\Omega)}(\beta,\gamma) = \beta\to\gamma$.

(e) $\neg_{\mathbf{SetF}(\Omega)}(\beta) = \neg\beta$.

Definition 2
Let ψ be a formula in J and let X be a set of variables containing all free variables of ψ. Let $\mathscr{D}$ be a model of J in a category $\mathbf{K} = \mathbf{SetF}(\Omega), \mathbf{SetR}(\Omega)$. Then we set

$$\|\psi\|_{\mathscr{D},X} = G(|\psi|_{\mathscr{D},X}) \quad \textit{if} \quad \mathbf{K} = \mathbf{SetR}(\Omega),$$
$$\|\psi\|_{\mathscr{D},X} = H(|\psi|_{\mathscr{D},X}) \quad \textit{if} \quad \mathbf{K} = \mathbf{SetF}(\Omega).$$

Theorem 2
For any formula ψ, $\|\psi\|_{\mathscr{D},X} : (A,\delta)^X \to (\Omega,\leftrightarrow)$ is a morphism in a category $\mathbf{K} = \mathbf{SetR}(\Omega), \mathbf{SetF}(\Omega)$.

4 Category $\mathbf{Set}(\Omega)$

To investigate a relationship between classical fuzzy sets (A,f), where $f : A\to\Omega$ is a map and Ω-sets (A,δ) it will be necessary to modify a definition of a similarity relation δ. There are several natural ways how to transform a classical fuzzy set into a similarity relation. For example, for a fuzzy set f we can define a similarity relation σ_f such that $\sigma_f(x,y) = f(x)\leftrightarrow f(y)$. We can also introduce another relation, namely $\Delta_f(x,y) = f(x)$ iff $x = y$ and $\Delta_f(x,y) = 0$, otherwise. It is clear that in that case Δ_f does not satisfies the condition $\Delta_f(x,x) = 1$ from the definition of similarity relations. Nevertheless to investigate a relationship between fuzzy sets and such similarity relations, we will introduce in this section a generalization of Ω-sets, namely we will define a new category $\mathbf{SetR}(\Omega)_*$ with objects (A,δ), where A is a set and δ satisfies only the conditions (b) and (c) from definition of a

ISBN: 978-989-95079-6-8

similarity relation. Moreover, we will require that morphisms of $\mathbf{SetR}(\Omega)_*$ are those of morphisms $f : (A,\delta) \to (B,\gamma)$ from $\mathbf{SetR}(\Omega)$ that satisfy the following additional conditions:

(a) $\gamma(x,x) = \bigvee_{y\in B} f(x,y)$, for all $x \in A$,

(b) $f(x,y) \otimes f(x,z) \le \gamma(y,z)$.

In this section we show a deeper relationship between morphisms in $\mathbf{Set}(\Omega)$ and $\mathbf{SetR}(\Omega)_*$ but only for a very special Ω, namely for a totally ordered Heyting algebra Ω.

Proposition 2
Let Ω be a totally ordered complete Heyting algebra and let $(A,\alpha),(B,\beta)$ be objects from a category $\mathbf{Set}(\Omega)$. Let $f : A \times B \to \Omega$ be a map. Then the following statement are equivalent.

(a) $f : (A,\Delta_\alpha) \to (B,\Delta_\beta)$ is a morphism in the category $\mathbf{SetR}(\Omega)_*$.

(b) There exist a morphism $g : (A,\alpha) \to (B,\beta)$ in a category $\mathbf{Set}(\Omega)$ such that

(i) $(\forall a \in A, b \in B) \quad f(a,b) \le \Delta_\beta(g(a),b)$,

(ii) $(\forall a \in A, b \in B) \quad f(a,g(a)) = \alpha(a)$.

(c) There exists a morphism $g : (A,\alpha) \to (B,\beta)$ in a category $\mathbf{Set}(\Omega)$ such that

$$(\forall a \in A, b \in B) \quad f(a,b) = \alpha(a) \wedge \Delta_\beta(g(a),b).$$

Proof. Recall that a *singleton* in $(X,\delta) \in \mathbf{SetR}(\Omega)_*$ is a map $s : X \to \Omega$ such that

$$(\forall x,y \in X) \quad s(x) \wedge \delta(x,y) \le s(y),$$
$$(\forall x,y \in X) \quad s(x) \wedge s(y) \le \delta(x,y).$$

The set $\mathrm{singl}\,(X,\delta)$ of all singletons of (X,δ) with a function $\tau_{(X,\delta)}$ is then an object of $\mathbf{SetR}(\Omega)_*$ if we set

$$(\forall s,t \in \mathrm{singl}\,) \quad \tau_{(X,\delta)}(s,t) = \bigvee_{a\in X} s(a) \wedge t(a).$$

(a)⇒ (b). We construct a map $h : A \to \mathrm{singl}\,(B,\Delta_\beta)$ such that

$$(\forall a \in A)(\forall y \in B) \quad h(a)(y) = \bigvee_{x\in A} (f(x,y) \wedge \Delta_\alpha(a,x)).$$

We have to show that $h(a) \in \mathrm{singl}\,(B,\Delta_\beta)$ for any $a \in A$. In fact, for $y \ne y' \in B$ we have

$$h(a)(y) \wedge h(a)(y') \le f(a,y) \wedge f(a,y') \le \Delta_\beta(y,y') = 0,$$
$$h(a)(y) \le f(a,y) \le \Delta_\alpha(a,a) = \alpha(a).$$

Moreover, $h : (A,\Delta_\alpha) \to (\mathrm{singl}\,(B,\Delta_\beta),\tau)$ is a morphism in a category $\mathbf{SetR}(\Omega)_*$. In fact, for any $a,a' \in A$ we have

$$\tau(h(a),h(a')) = \bigvee_{y\in B}\bigvee_{x,x'\in A} (f(x,y) \wedge \Delta_\alpha(a,x) \wedge f(x',y) \wedge \Delta_\alpha(a',x')) \ge$$
$$\ge \bigvee_{y\in B} (f(a,y) \wedge \Delta_\alpha(a,a) \wedge \Delta_\alpha(a',a)) = \Delta_\alpha(a',a),$$
$$\tau(h(a),h(a)) = \bigvee_{y\in B}\bigvee_{x\in A} f(x,y) \wedge \Delta_\alpha(a,x) \le \bigvee_{y\in B} f(a,y) = \Delta_\alpha(a,a).$$

Since $h(a)(y) \wedge h(a)(y') = 0$ for all $y \ne y' \in B$ and Ω is totally ordered, for any $a \in A$ there exists at most one element $g(a) \in B$ such that $h(a)(g(a)) > 0$. If $h(a)(y) = 0$ for all $y \in B$, let $g(a) \in B$ be an arbitrary element. Then $g : (A,\alpha) \to (B,\beta)$ is a morphism in $\mathbf{Set}(\Omega)$. In fact, if $h(a)(g(a)) > 0$, then we have

$$\alpha(a) = \Delta_\alpha(a,a) = \tau(h(a),h(a)) = \bigvee_{y\in B} h(a)(y)$$
$$= h(a)(g(a)) = \bigvee_{x\in A} f(x,g(a)) \wedge \Delta_\alpha(a,x) \le$$
$$f(a,g(a)) \le \Delta_\beta(g(a),g(a)) = \beta(g(a)).$$

If $h(a)(y) = 0$ for all $y \in B$, then $\alpha(a) = \Delta_\alpha(a,a) = 0 \le \Delta_\beta(g(a),g(a))$. We show that the function g satisfies the conditions (i) and (ii). Let $h(a)(g(a)) > 0$. Then we have

$$0 < h(a)(g(a)) = \bigvee_{x\in A} f(x,g(a)) \wedge \Delta_\alpha(a,x) \le f(a,g(a)).$$

Moreover, since f is a morphism in $\mathbf{SetR}(\Omega)_*$, we have

$$(\forall a,b) \quad f(a,b) \wedge f(a,g(a)) \le \Delta_\beta(b,g(a)),$$

and it follows that $f(a,b) = 0$ for all $b \in B$ such that $b \ne g(a)$. Hence, the condition (i) holds. Since $\alpha(a) = \Delta_\alpha(a,a) = \bigvee_{y\in B} f(a,y) = f(a,g(a))$, the condition (ii) holds. Now, if $h(a)(y) = 0$ for all $y \in B$, it follows that $f(a,y) = 0$ for all $y \in B$ and the conditions (i),(ii) hold as well.

(b)⇒(c). This is a trivial computation only.

(c)⇒(a). Let $g : (A,\alpha) \to (B,\beta)$ be a morphism in $\mathbf{Set}(\Omega)$ and let $f : A \times B \to \Omega$ be a function from the statement (c). Then for any $a \in A$ we have

$$\bigvee_{b\in B} f(a,b) = \bigvee_{b\in B} (\alpha(a) \wedge \Delta_\beta(g(a),b)) =$$
$$= \alpha(a) \wedge \Delta_\beta(g(a),g(a)) = \alpha(a) \wedge \beta(g(a)) = \alpha(a).$$

and $f : (A,\Delta_\alpha) \to (B,\Delta_\beta)$ is a morphism in $\mathbf{SetR}(\Omega)_*$. ■

It is clear that Proposition 2 introduces a map

$$\varphi : \mathbf{Hom}_{\mathbf{Set}(\Omega)}((A,\alpha),(B,\beta)) \to$$
$$\mathbf{Hom}_{\mathbf{SetR}(\Omega)_*}((A,\Delta_\alpha),(B,\Delta_\beta))$$

such that for any morphism $f : (A,\alpha) \to (B,\beta)$ in $\mathbf{Set}(\Omega)$, $\varphi(f) : (A,\Delta_\alpha) \to (B,\Delta_\beta)$ is a morphism in $\mathbf{SetR}(\Omega)_*$ such

ISBN: 978-989-95079-6-8

that $\varphi(f)(a,b) = \alpha(a) \wedge \Delta_\beta(f(a),b)$ for all $a \in A, b \in B$. On the other hand, there exists a relation

$$\psi \subseteq \mathbf{Hom}_{\mathbf{SetR}(\Omega)_*}((A,\Delta_\alpha),(B,\Delta_\beta)) \times \mathbf{Hom}_{\mathbf{Set}(\Omega)}((A,\alpha),(B,\beta))$$

such that $(f,g) \in \psi$ if g is a morphism from (c) (Proposition 2) for a morphism f. It should be observed that (on the contrary to φ) ψ is not a map, in general. In fact, let (A,α) be an object in $\mathbf{Set}(\Omega)$ such that $\alpha(a) = 0$ for some $a \in A$. Let $f : (A,\Delta_\alpha) \to (B,\Delta_\beta)$ be a morphism in $\mathbf{SetR}(\Omega)_*$ and let $g : (A,\alpha) \to (B,\beta)$ be a morphism in $\mathbf{Set}(\Omega)$ satisfying the condition (c) from Proposition 2. Let us define a map $g' : A \to B$ such that $g'(x) = g(x)$ for all $x \in A, x \neq a$ and $g'(a) \neq g(a)$. Then $g' : (A,\alpha) \to (B,\beta)$ is a morphism, $(f,g),(f,g') \in \psi$ and $g' \neq g$. Hence, ψ is not a map.

Acknowledgment

Author express his thanks to reviewers for their valuable remarks and recommendations. The research was supported by the project MSM6198898701 of the MŠMT ČR, grant 201/07/0191 of GAČR and grant 1M0572.

References

[1] Höhle, U., Presheaves over GL-monoids. *Non-Classical Logic and Their Applications to Fuzzy Subsets*, Kluwer Academic Publ., Dordrecht, New York(1995), 127-157.

[2] Höhle, U., M-Valued sets and sheaves over integral, commutative cl-monoids. *Applications of Category Theory to Fuzzy Subsets*, Kluwer Academic Publ., Dordrecht, Boston (1992), 33-72.

[3] Höhle, U., Classification of Subsheaves over GL-algebras. *Proceedings of Logic Colloquium 98 Prague*, Springer Verlag (1999).

[4] Höhle, U., Commutative, residuated l-monoids. *Non-Classical Logics and Their Applications to Fuzzy Subsets*, Kluwer Academic Publ. Dordrecht, New York (1995), 53–106.

[5] Močkoř, J., Complete subobjects of fuzzy sets over MV-algebras. *Czech.Math.J.* **129(54)**(2004), 379–392.

[6] Močkoř, J., Extensional subobjects in categories of Ω-fuzzy sets. *Czech.Math.J.* **57(132)**(2007), 631–645.

[7] Močkoř, J., Characteristic morphisms and models of fuzzy logic in a category of sets with similarities. *Theoretical Advances and Applications of Fuzzy Logic and Soft Computing* Springer-Verlag Berlin (2007),832–840.

[8] Močkoř, J., Covariant functors in categories of fuzzy sets over MV-algebras. *Advances in Fuzzy Sets and Systems***1(2)**(2006), 83–109.

[9] Močkoř, J., Models of fuzzy logic in a category of sets with similarity relations. *International Journal of Innnovative Computing, Information and Control***4(5)**(2008),1063–1068.

[10] Močkoř, J.,Fuzzy logic models in a category of fuzzy relations. *to appear in Soft Computing 2009.*

[11] Novák,V., Perfilijeva, I., Močkoř,J., *Mathematical principles of fuzzy logic* Kluwer Academic Publishers,Boston,Dordrecht, London (1999).

Bipolar Queries: An Approach and its Various Interpretations

Sławomir Zadrożny[1] Janusz Kacprzyk[2]

1.Systems Research Institute, Polish Academy of Sciences
Warszawa, Poland
2.Systems Research Institute, Polish Academy of Sciences
Warszawa, Poland
Email: Slawomir.Zadrozny@ibspan.waw.pl, Janusz.Kacprzyk@ibspan.waw.pl

Abstract *— The concept of a bipolar query is studied in the framework of the flexible fuzzy querying of databases. The focus is on the aggregation of the negative and positive conditions forming a bipolar query. Three formal logical representations of such an aggregation are proposed and analyzed taking into account various possible interpretations of fuzzy logical connectives. The relation to other approaches known in the literature is shown.*

Keywords— aggregation operators, bipolar queries, flexible queries, fuzzy logic, logical connectives modeling.

1 Introduction

A query to a (relational) database may be identified, in a slightly simplified view, with a condition expressing what the user is looking for. A database management system returns in the response to a query a list of tuples satisfying this condition. Such a condition is usually composed of a few, simple, atomic conditions putting some constraints on the values of the attributes characterizing given relation (table). We will adopt such a simplified view, assuming moreover that the user is concerned with just one relation, i.e., using SQL's terminology, joins, subqueries etc. are excluded for the sake of clarity of the presentation of the main idea.

Atomic conditions are connected using the classical logical connectives of the conjunction, disjunction and negation. In many standard application scenarios this is exactly what is needed to retrieve required data from a database. However in some other scenarios more sophisticated forms of the queries seem to be useful. Thus, in the literature further extensions of this basic setting are proposed. Firstly, many authors advocate the convenience of using *linguistic terms* in queries (cf., e.g., [1, 2]). For example, it is more natural and comfortable for a user of a real-estate agency to state that she or he is looking for a "not very expensive house" rather than to use a precise interval of acceptable prices. The benefits of using fuzzy logic to model linguistic terms in queries are widely advocated and there is a large number of of both theoretical and practical results obtained by the researchers in this area (cf., e.g., [3, 1, 2]). Secondly, some (atomic) conditions may be for the user more important than the others. For example, a customer of a real-estate agency may be looking for a "not very expensive house located in a nice city district" and to treat the former condition as a much more important than the latter. Thus in the overall matching degree of a tuple against such a query the satisfaction of the former condition is crucial, while of the latter is to some extent secondary. Proper modeling of importance weights is a subject of many papers (cf., e.g., [4]).

The essence of the concept of a *bipolar query* is in a new way of differentiating the conditions than by assigning them with some *fixed* importance weights. Namely, a bipolar query is defined by two conditions: one – *negative (required)* – expresses the constraints that *have to be* satisfied while the second – *positive (preferred)* – expresses only what is *desired* and its violation does not necessarily lead to the rejection of given tuple. This combination of conditions may be meant and studied in a few different ways. Here we are interested in the ways the satisfaction of both types of conditions should be combined to obtain an overall matching degree of the whole query. From this point of view the most important is the question if there is a conflict between the conditions or they can be satisfied simultaneously. If there is a total conflict, i.e., satisfying one condition means totally failing to satisfy another, the bipolar query reduces itself to the required condition. On the other hand, if both conditions may be totally satisfied simultaneously then the bipolar query reduces to a simple conjunction. Thus, the most interesting are intermediate cases which may be characterized by a degree of conflict between the conditions.

In this paper we study three logical formulas which may be used as a representation of a bipolar query. They are equivalent in case of the classical predicate logic but as soon as we adopt the fuzzy (multivalued) logic context they become distinct and exhibit different properties depending on the assumed set of fuzzy operators used to model logical connectives. Our study may be seen as an attempt at justifying particular choices in this respect.

The structure of this paper is as follows. In Section 2 we briefly remind the concept of the bipolar query and introduce the notation used later on. We also discuss the concept of the *winnow* operator which is used to derive logical representations of bipolar queries. In Section 3 we compare particular representations under different choices of logical operators, i.e., operators used to model particular logical connectives.

2 The concept of a bipolar query

The very concept of bipolar queries has been introduced by Dubois and Prade [5]. Its basic idea is to distinguish two types of query conditions, which are related to the *negative* and *positive* preferences of a user. The former coincide with the traditional understanding of a condition as a constraint, which defines a set of *feasible* tuples or, equivalently, excludes all tuples that do not satisfy it. The latter, on the other hand, characterizes those tuples which are really desired, with such an un-

ISBN: 978-989-95079-6-8

derstanding that violating such a condition by a tuple does not necessarily exclude it from the consideration. Bipolar queries may be exemplified by the following one:

"Find a *non-expensive* house *and possibly* located *near* a railway station (1)

where the condition referring to the price is a negative one, excluding expensive houses, and the condition referring to the location is a positive one, expressing just a desire to get a house conveniently located, *if possible*.

Now there is a crucial question of the interplay between these two types of conditions which distinguishes different lines of research in this area. Namely, Dubois and Prade basically assume that these conditions should be consistent in such a sense that the set of desired tuples should be a subset of the set of feasible tuples. Then the main question is how to take into account the sets of negative and positive conditions. These are aggregated separately and if the resulting overall negative and positive conditions are not consistent some measures are undertaken to make them so. The answer to such a bipolar query is generated according to the strategy "first select (with respect to the negative condition) and then order (with respect to the positive condition)". This strategy requires a precisiation in case the first condition is fuzzy, i.e., is satisfied by tuples to a degree – what is, of course, the most interesting case, anyway. Then it is not that clear what does it mean to select tuples satisfying a fuzzy condition as they form, in fact, a fuzzy set. Dubois and Prade [6] propose to employ here a lexicographic order of the tuples represented by vectors of the degrees of matching of particular conditions. They propose also a comprehensive representation of bipolar preferences in the framework of the possibility theory [7].

Another line of research explicitly takes into account the conflict between the constraints and the desires and looks for the aggregation of both conditions directly referring to the degree of this conflict. This interpretation is emphasized by the use of the "and possibly" operator in (1). Thus, in general describing a bipolar query we will use the following notation:

$$C \textit{ and possibly } P \tag{2}$$

or, equivalently, an answer to a bipolar query may be defined as the set of tuples: $\{t : C(t) \textit{ and possibly } P(t)\}$. The above form puts emphasis on the question of a proper modeling of the aggregation of both types of conditions, which is expressed here with the use of the "and possibly" operator.

The interest of the database community in this type of queries dates back to the paper by Lacroix and Lavency [8]. They were the first to propose the use of a query comprising two categories of conditions: one which is mandatory (C) and another which expresses just mere preferences (desires) (P). The bipolarity of these conditions becomes evident as soon as one adopts the following interpretation. The former condition C may be seen as expressing the *negative* preferences: the tuples which do not satisfy it are definitely not matching the whole query. The latter condition P, on the other hand, has a *positive* character: a tuple satisfying it is preferred over another tuple not satisfying it, provided both tuples satisfy the mandatory condition C.

For the purposes of a further discussion we will use the following notation. We assume the queries are addressed against a set of tuples $T = \{t_j\}$ comprising a relation. We will identify the negative and positive conditions of a bipolar query with the predicates that represent them and denote them as C and P, respectively. For a tuple $t \in T$, $C(t)$ and $P(t)$ will denote that the tuple t satisfies respective condition (in crisp case) or the degrees of satisfaction, if the conditions are fuzzy. We will also denote the whole bipolar query as (C, P).

According to the original (crisp) approach by Lacroix and Lavency if there are no tuples meeting both conditions then the result of the aggregation is determined by the negative condition C alone. Otherwise the aggregation becomes a regular conjunction of both conditions. Thus the answer to such a query depends not only on the explicit arguments, i.e., $C(t)$ and $P(t)$, but also on the content of the database. This dependence is best expressed by the following logical formula [8]:

$$C(t) \textit{ and possibly } P(t) \equiv C(t) \wedge \exists s(C(s) \wedge P(s)) \Rightarrow P(t) \tag{3}$$

The characteristic feature of such an interpretation of bipolar queries is that if there is no conflict between the conditions P and C, i.e., there are tuples satisfying both of them, then the query turns into a conjunction of the conditions. On the other hand if there are no tuples satisfying both conditions then only condition C is used to select tuples.

Such an aggregation operator has been later proposed independently by Dubois and Prade [9] in the context of default reasoning and by Yager [10, 11] in the context of the multicriteria decision making for the case of so-called *possibilistically qualified criteria*. Yager [11] intuitively characterizes a possibilistically quantified criterion as such which should be satisfied unless it interferes with satisfaction of other criteria. This is in fact the essence of bipolar queries in the sense advocated in this paper. This concept was also applied by Bordogna and Pasi [12] for the textual information retrieval task.

Lacroix and Lavency [8] consider only the case of crisp conditions C and P. Then a bipolar query may be, in fact, processed using the "first select using C then order using P" strategy, i.e., the answer to the bipolar query (C, P) is obtained by, first, finding tuples satisfying C and, second, choosing from among them those satisfying the condition P, if any. If C is crisp and P is fuzzy then the second step consists in non-increasingly *ordering* the tuples satisfying C according to their degree of satisfaction of P. This understanding is predominant in the literature dealing with fuzzy extensions of the original concept of Lacroix and Lavency. Both, direct extensions proposed by Bosc and Pivert [13, 14] as well as sophisticated possibility theory-based interpretation of this concept by Dubois and Prade [6] focus, in fact, on the proper treatment of *multiple* required and preferred conditions, basically assuming the above strategy as the way of combining the negative and positive conditions.

In [15, 16] we propose a "fuzzification" of the formula (3) and study its basic properties. Here we focus on the question of combining fuzzy conditions C and P and will treat them in what follows as atomic. From this point of view it is worth mentioning some other approaches which are of relevance here. Dujmović [17] introduced the concept of the *partial absorption function* which may be used to combine the values of two variables in such a way that one variable controls the influence of the other on the result of their combina-

ISBN: 978-989-95079-6-8

tion. It makes it possible to express the requirement that for a high value of the result a high value of the first variable is mandatory while the high value of the second is desired but not mandatory. When applied to the aggregation of the values of $C(t)$ and $P(t)$ this is somehow similar to the idea of bipolar query expressed by (3), but lacks its dependence on the content of the whole database. This approach may be seen as based on a sophisticated, *dynamic* weighting of the importance of the combined values, where the weights itself depend on the combined values. A similar approach has been proposed by Dubois and Prade; cf., e.g., [4].

The operator "among", very close to the discussed here operator "and possibly", has been proposed by Tudorie [18]. She considers queries of the type "find tuples satisfying a condition P among those satisfying a condition C", which are in fact equivalent to the bipolar queries understood as in (2). The evaluation of a query with the "among" operator is very similar to the one adopted by us for the bipolar query, but is expressed in terms of the rescaling of the linguistic terms used in the condition P. Namely, first the set of tuples satisfying the condition C to a non-zero degree is selected. Then the membership functions of the fuzzy sets representing the linguistic terms appearing in P (such as "near" in (1)) are rescaled taking into account the actual range of the corresponding attributes in the set of tuples selected in the first step. For example, if originally the distance of 2 kilometers from the station has the membership value to the fuzzy set representing the term "near" equals 0.5, and it turns out that it is the shortest distance among the houses selected in the first step (i.e., among non-expensive houses, in case of the query (1)), then this membership degree may be changed to 1 (the actual algorithm of rescaling may, of course, take different forms). Finally, the overall matching degree is computed as a conjunction of the matching degrees against the condition C and the "modified" condition P, i.e., the one for which the rescaled membership function of the linguistic terms is used. Please note that if there is no interference between both conditions (in the sense discussed earlier) then there is no need for rescaling the membership functions and the query turns into a conjunction of both conditions, like in the case of the bipolar query.

Bipolar queries may be also seen a special case of *queries with preferences* proposed recently, for the crisp case, by Chomicki [19]. In the framework of this approach a new relational algebra operator called *winnow* is introduced. This unary operator selects from a set of tuples T those which are *non-dominated* with respect to a given *preference relation* R, $R \subseteq T \times T$. If two tuples $t, s \in T$ are in relation R, i.e., $R(t, s)$, then it is said that the tuple t *dominates* the tuple s with respect to the relation R. Then the *winnow* operator ω_R is defined as follows

$$\omega_R(T) = \{t \in T : \neg\exists_{s \in T}\ R(s, t)\} \tag{4}$$

Thus, for a given set of tuples it yields a subset of the *non-dominated* tuples with respect to R.

The concept of the *winnow* operator may be illustrated with the following example. Let us consider a database of a real-estate agency with a table `HOUSES` describing the details of particular real-estate properties offered by an agency. The schema of the relation `HOUSES` contains, among other, the attributes `city` and `price`. Let us assume that we are interested in the list of the *cheapest* houses in each city. Then the preference relation should be defined as follows

$$R(t, s) \Leftrightarrow (t.\texttt{city} = s.\texttt{city}) \wedge (t.\texttt{price} < s.\texttt{price})$$

where $t.A$ denotes the value of attribute A (e.g., `price`) at a tuple t. Then the *winnow* operator $\omega_R(\texttt{HOUSES})$ will select the houses that are sought. Indeed, according to the definition of the *winnow* operator, we will get as an answer a set of houses, which are non-dominated with respect to R, i.e., for which there is no other house in the same city which has a lower price.

In [16] we proposed a fuzzy counterpart of the *winnow* operator taking into account the fuzziness of the preference relation R and of the related concept of non-dominance as well as the fact that the set of tuples T is also a fuzzy set. It may be expressed as follows:

$$\mu_{\omega_R(T)}(t) = \text{truth}(T(t) \wedge \forall_s\ (T(s) \rightarrow \neg R(s, t))) \tag{5}$$

where $\mu_{\omega_R(T)}(t)$ denotes the value of the membership degree of the tuple t to the fuzzy set of tuples defined by $\omega_R(T)$.

A bipolar query (C, P) may be expressed using the concept of the fuzzy *winnow* operator as follows [19, 16]. Let R be a fuzzy preference relation of the following form (symbols R and P denote both the fuzzy predicates and the membership functions of corresponding fuzzy sets, depending on the context):

$$R(t, s) \Leftrightarrow P(t) \wedge \neg P(s) \tag{6}$$

Then the bipolar query may be expressed as the combination of the selection and fuzzy *winnow* operators $\omega_R(\sigma_C(T))$, i.e.,

$$\mu_{\omega_R(\sigma_C(T))} = \text{truth}(C(t) \wedge \forall_s\ (C(s) \rightarrow (\neg P(s) \vee P(t)))) \tag{7}$$

(where $\sigma_C(T)$ is a usual "fuzzy" extension of the standard relational algebra selection operator, i.e., $\mu_{\sigma_C(T)}(t) = C(t)$).

The definition of the fuzzy *winnow* operator (5) as well as the "fuzzification" of the formula defining a bipolar query (3) leave open the question of a choice of the logical operators which should be used to model particular logical connectives occurring in both formulas. In [16] we show that for a specific choice of them the fuzzy set of tuples obtained using (7) is identical with the fuzzy set defined by (3). In [20] we analyze the properties of the "fuzzified" version of (3) for the broader class of the logical operators. In the next section we further advance this study.

3 Various interpretations of bipolar queries and their properties

In our previous work [15, 16, 20] we studied a specific fuzzy version of the Lacroix and Lavency formula (3) representing an interpretation of the concept of bipolar queries. We have also shown its basic relation with a fuzzy version of the *winnow* operator we proposed. Here we extend this study comparing three formulas that may be used to represent the bipolar query and their properties under different interpretations of the logical connectives occurring in them.

We derive the logical formulas expressing the matching degree of a bipolar query in three different ways (we repeat here some formulas introduced in the previous section for the convenience of the reader):

ISBN: 978-989-95079-6-8

- making a direct "fuzzification" of the formula (3) proposed for the crisp case by Lacroix and Lavency [8]:

$$C(t) \textit{ and possibly } P(t) \equiv C(t) \wedge (\exists s\ (C(s) \wedge P(s)) \Rightarrow P(t)) \quad (8)$$

- making a direct "fuzzification" of the crisp *winnow* operator (4) and applying it with a specific preference relation (6) to obtain a bipolar query representation:

$$C(t) \textit{ and possibly } P(t) \equiv C(t) \wedge \neg\exists s\ ((C(s) \wedge P(s) \wedge \neg P(t))) \quad (9)$$

- using our fuzzy version of the *winnow* operator (5) and applying it as above:

$$C(t) \textit{ and possibly } P(t) \equiv C(t) \wedge \forall s\ (C(s) \Rightarrow (\neg P(s) \vee P(t))) \quad (10)$$

It may be easily seen that in the framework of the classical logic all three above formulas are equivalent. Here we study their properties in case of fuzzy (multivalued) interpretation, in particular taking into account various operators which may be used to model logical connectives. We follow usual approach of modeling conjunction and disjunction by the t-norm and t-conorm operators, respectively [21].

In order to carry out the analysis we consider so-called De Morgan Triples $(\wedge, \vee, \neg)$ that comprise a t-norm operator $\wedge$, a t-conorm operator $\vee$ and a negation operator $\neg$, where $\neg(x \vee y) = \neg x \wedge \neg y$ holds. Three following De Morgan Triples play the most important role in fuzzy logic (cf., e.g., [21] for a justification) $(\wedge_{min}, \vee_{max}, \neg)$, $(\wedge_{\Pi}, \vee_{\Pi}, \neg)$, $(\wedge_{W}, \vee_{W}, \neg)$, where particular t-norms and t-conorms are defined as follows:

$t-$ *norms*		
$x\ \wedge_{min}\ y$	$= \min(x, y)$	*minimum*
$x\ \wedge_{\Pi}\ y$	$= x \cdot y$	*product*
$x\ \wedge_{W}\ y$	$= \max(0, x + y - 1)$	*Łukasiewicz*

$t-$ *conorms*		
$x\ \vee_{max}\ y$	$= \max(x, y)$	*maximum*
$x\ \vee_{\Pi}\ y$	$= x + y - x \cdot y$	*probabilistic sum*
$x\ \vee_{W}\ y$	$= \min(1, x + y)$	*Łukasiewicz*

We will refer to these De Morgan Triples in what follows as, respectively, MinMax, Π and W triples. The negation operator $\neg$ in case of all the above De Morgan Triples is defined as: $\neg x = 1 - x$. Both t-norms and t-conorms are by definition associative and thus may be treated as m-ary operators, i.e., expressions like $x \wedge y \wedge \ldots$ and $x \vee y \vee \ldots$ are well defined.

Basically, the general and existential quantifiers are identified in fuzzy logic, for the case of a finite universe, with the maximum and minimum operators, respectively. They may be generalized via the use of other t-norms and t-conorms what leads to the concept of *t-quantifiers* and *s-quantifiers*; cf., e.g., [22]. The truth of a statement involving such a quantifier is computed as follows ($\{a_1, \ldots, a_m\}$ is a finite universe under consideration):

$$\text{truth}(\forall x A(x)) = \mu_A(a_1) \wedge \mu_A(a_2) \wedge \ldots \wedge \mu_A(a_m) \quad (11)$$

$$\text{truth}(\exists x A(x)) = \mu_A(a_1) \vee \mu_A(a_2) \vee \ldots \vee \mu_A(a_m) \quad (12)$$

We use generalized quantifiers while interpreting formulas (8)-(10) and particular t- and s-quantifiers will be denoted in what follows by the $\forall$ and $\exists$ symbol with a subscript indicating underlying t-norm or s-norm, e.g., $\exists_{max}$ denotes a fuzzy existential quantifier which obtains when the t-conorm "maximum" is used.

We consider two implication operators related to a given De Morgan Triple $(\wedge, \vee, \neg)$, so-called S-implications:

$$x \rightarrow_{S-\vee} y = \neg x \vee y \quad (13)$$

and R-implications:

$$x \rightarrow_{R-\wedge} y = \sup\{z : x \wedge z \leq y\} \quad (14)$$

Thus, for particular De Morgan Triples one obtains the following R-implication operators:

$$x\ \rightarrow_{R-min}\ y = \begin{cases} 1 & \text{for } x \leq y \\ y & \text{for } x > y \end{cases}$$

$$x\ \rightarrow_{R-\Pi}\ y = \begin{cases} 1 & \text{for } x = 0 \\ \min\{1, \frac{y}{x}\} & \text{for } x \neq 0 \end{cases}$$

$$x\ \rightarrow_{R-W}\ y = \min(1 - x + y, 1)$$

and the following S-implication operators:

$$x\ \rightarrow_{S-max}\ y = \max(1 - x, y)$$

$$x\ \rightarrow_{S-\Pi}\ y = 1 - x + x \cdot y$$

$$x\ \rightarrow_{S-W}\ y = \min(1 - x + y, 1)$$

Now let us consider the question of the choice of one of the formulas (8)-(10) to represent bipolar queries and an appropriate modeling of the logical connectives occurring therein, i.e., the choice of one of the De Morgan Triples.

In [20] we have shown certain basic properties of the fuzzified version of the original formula (8). Some of them are valid for any choice of logical operators, some are limited to some special cases. We study some of them here again, checking if they are valid also for the formulas (9) and (10). However, first we start with a property identifying the equivalence between formulas (8)-(10) for a certain choice of the logical operators.

Property 1 For a distributive De Morgan triple, i.e., when $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$, and a related S-implication all formulas (8)-(10) are equivalent.

We will show the equivalence between (8) and (9) and the remaining equivalences may be shown in a similar way (the equivalence of (8) and (10) was shown by us in [16] for the specific case of the MinMax De Morgan Triple). Formula (9) may be rewritten as follows:

$$C(t) \wedge \forall s\ (\neg C(s) \vee \neg P(s) \vee P(t))$$

what, due to the assumed distributivity of the triple is equivalent to:

$$C(t) \wedge ((\forall s\ (\neg C(s) \vee \neg P(s))) \vee P(t))$$

and via the following series of transformations leads to (8):

$$C(t) \wedge ((\forall s\ \neg(C(s) \wedge P(s))) \vee P(t)) \equiv$$
$$C(t) \wedge ((\neg\exists s\ (C(s) \wedge P(s))) \vee P(t)) \equiv$$
$$C(t) \wedge (\exists s\ (C(s) \wedge P(s)) \Rightarrow P(t))$$

ISBN: 978-989-95079-6-8

The last equivalence holds for "⇒" being an S-implication.

Among three De Morgan Triples that we consider in this paper only the MinMax triple is distributive. Thus for it and related S-implication we get equivalence of all three formulas (8)-(9). Property 1 shows only a sufficient condition for this equivalence, but for the other two triples some counterexamples for such an equivalence may be easily shown.

In [20] we have shown a property which is worth reminding as it best characterizes the understanding of the bipolar queries adopted here. Namely, if there is no conflict between the required (C) and preferred (P) conditions at all, i.e., there is a tuple fully (to a degree equal 1) satisfying both of them then the formula (8) turns into a regular conjunction of both conditions. This may be formally expressed as follows.

Property 2 If there exists a tuple t such that $C(t) = P(t) = 1$ then a bipolar query (C, P) turns into the conjunction $C \wedge P$.

In [20] we have shown that for any combination of a t-norm, t-conorm and S-implication or R-implication such that $(1 \Rightarrow x) = x$, Property 2 holds in case of the formula (8). Note that the conjunction mentioned in the Property is modeled then via assumed t-norm.

Here, in view of the Property 1, it is clear that in case of the MinMax De Morgan Triple and the S-implication, Property 2 holds also for the formulas (9) and (10). Moreover, it may be easily proved that this property also holds for the MinMax De Morgan Triple and the R-implication in case of the formula (10). Namely, we will show that:

$$\text{truth}(\forall s\ (C(s) \Rightarrow_{R-min} (\neg P(s) \vee_{max} P(t))) = \min_s \begin{cases} 1 & \text{if } C(s) \leq \neg P(s) \vee_{max} P(t) \\ \neg P(s) \vee_{max} P(t) & \text{otherwise} \end{cases} \tag{15}$$

is equal $P(t)$ for any tuple t, under the assumptions of the Property 2. If $P(t) = 1$ then for all tuples s the value of $C(s)$ is lower or equal $\max(1 - P(s), P(t))$ and thus (15) is equal 1, i.e., is equal to $P(t)$. Now let us assume that $P(t) < 1$ and let us denote with u a tuple for which $C(u) = P(u) = 1$ (the existence of such a tuple is assumed in the Property 2). Then, for any tuple t, the minimum over s in (15) is realized for $s = u$ and thus is equal $\max(1 - 1, P(t))$, i.e., $P(t)$, what completes our proof of the Property 2 also for the formula (10) and the R-implication operator.

Some examples may be easily found showing that Property 2 fails for the W and Π De Morgan triples.

Another property of the formula (8), shown in [20], is valid for two other formulas (9) and (10), and may be expressed as follows.

Property 3 If for a tuple t the value of $P(t)$ is equal to 1, then a bipolar query (C, P) turns into $C(x)$.

This property holds for all formulas (8)-(10), for any combination of a t-norm, t-conorm and S-implication or R-implication.

This property is a direct consequence of the general properties of the t-norm ($(x \wedge 0) = 0$), t-conorm ($(x \vee 1) = 1$ and $(0 \vee 0) = 0$) and implication operators ($(x \Rightarrow 1) = 1$).

Now we show some additional properties of the formulas (8)-(10).

Property 4 Let us assume that truth($\exists s\ (C(s) \wedge P(s))) > 0$. Then, for a De Morgan Triple with a t-norm without zero divisors, i.e., where $\forall x, y > 0\ (x \wedge y) \neq 0$, and the related R-implication, the matching degree computed using (8) for a tuple t fully satisfying the required condition and not satisfying the preferred condition at all (i.e., $C(t) = 1$ and $P(t) = 0$) is equal 0.

This is a property of the R-implication. Notice that this means that such a tuple t will get lower matching degree than a tuple s which satisfies both conditions to a degree ϵ, whatever small ϵ is. Thus it is surely a property we would like to avoid and which makes (8) under both MinMax and Π De Morgan Triples (whose t-norms do not have zero divisors) with related R-implications less appealing as models of the bipolar query.

A similar property, formulated below, is exhibited by (10).

Property 5 Let us assume that there is at least one tuple u such that $C(u) > 0$ and $P(u) = 1$. Then, for a De Morgan Triple with a t-norm without zero divisors and the related R-implication the matching degree computed using (10) for a tuple t fully satisfying the required condition and not satisfying the preferred condition at all (i.e., $C(t) = 1$ and $P(t) = 0$) is equal 0.

Thus also (10) under both MinMax and Π De Morgan Triples with related R-implications is not very appealing as a model of the bipolar query. This property seems to favor the W De Morgan Triple, at least in case of the R-implication and formulas (8) and (10).

Another negative property of (8) for a specific combination of logical operators may be expressed as follows.

Property 6 For the MinMax De Morgan Triple used with related S-implication, the aggregation scheme defined by (8) may lead to the same matching degree for two tuples t and u while t strongly Pareto dominates u, i.e., $C(t) > C(u)$ and $P(t) > P(u)$.

This may be demonstrated with the following example. Let us denote $\exists_s\ (C(s) \wedge P(s))$ with $\exists CP$. Let $\exists CP = 0.7$ and $C(t) = 1$, $P(t) = 0.3$, $C(u) = 0.3$ and $P(u) = 0$. Then the matching degree computed for both tuples is equal 0.3, while t strongly Pareto dominates u.

In fact this property may be supplemented by observing that all tuples t such that $P(t) \leq (1 - \exists CP)$ and $C(t) \geq (1 - \exists CP)$ obtain the same matching degree, equal $1 - \exists CP$. This fact has been observed by Dubois and Prade [5] for a formula similar to (8). However, it should be noted that even for the tuples not satisfying the above conditions, the Pareto domination may be not reflected by (8) used with the logical operators specified by Property 6. For example, for a tuple u, such that $C(u) = P(u) = 0.6$, still assuming $\exists CP = 0.7$, the matching degree is equal 0.6. The same matching degree obtains for t such that $C(t) = 1$ and $P(t) = 0.6$ as well as for t such that $C(t) = 0.6$ and $P(t) = 1.0$, while in both these cases the tuple t Pareto dominates u.

Note, that due to the Property 1, the Property 6 is also valid for two other formulas (9)-(10).

Certainly the list of properties discussed in the paper is not exhaustive and they should be seen as a first attempt at a more comprehensive analysis of the bipolar queries and their various representations proposed here. Such an analysis should provide some hints which representation should be used un-

ISBN: 978-989-95079-6-8

der which conditions and using which set of logical operators. However, already the properties discussed here provide some hints which may be summarized as follows.

We first list some general properties shared by all three formulas (8)-(10) under any combination of the logical operators. These properties may be expressed concisely as the properties of the "and possibly" operator, which is the essence of our understanding of bipolar queries (this operator is denoted below as $\wedge_{possibly}$, but it should be remembered that this operator is not truth-functional):

- monotonicity (but not strict) in both arguments,
- boundary properties: $1 \wedge_{possibly} 1 = 1$ and $x \wedge_{possibly} 1 = x$ (Property 3).

Now let us look at particular formulas (8)-(10) and summarize their properties in combinations with all considered logical operators.

Formula (8) exhibits Property 2 when used with any combination of logical operators. This is surely its advantage as this property is characteristic for our understanding of bipolar queries. Property 4 seems to suggest that the R-implication should be avoided in case of formula (8) (unless it is used in the framework of the W De Morgan Triple, but then both types of implication operators are identical, thus a general hint of avoiding R-implication may be still seen as valid). Property 6 suggests that the MinMax De Morgan Triple is generally not appropriate for the formula (8).

Formula (9) exhibits Property 2 only for the MinMax De Morgan Triple, but Property 6 makes this triple inappropriate to some extent.

Formula (10) also satisfies Property 2 only for the MinMax De Morgan Triple, which is on the other hand somehow inappropriate due to the Properties 5 and 6.

Concluding, if Property 2 is required, what seems to be a reasonable postulate, then the best choice of the representation of the bipolar queries and logical operators to model the logical connectives therein– according to studied properties– seems to be formula (8) with the Π De Morgan Triple and the S-implication operator. Such a choice saves the obtained representation from some negative properties discussed in this paper. Further studies are needed in order to identify a more comprehensive list of postulated properties.

References

[1] P. Bosc and O. Pivert. SQLf: A relational database language for fuzzy querying. *IEEE Transactions on Fuzzy Systems*, 3(1):1–17, 1995.

[2] J. Kacprzyk and S. Zadrożny. Computing with words in intelligent database querying: standalone and internet-based applications. *Information Sciences*, 134(1-4):71–109, 2001.

[3] S. Zadrożny, G. De Tre, R. De Caluwe, and J. Kacprzyk. An overview of fuzzy approaches to flexible database querying. In Galindo [23], pages 34–53.

[4] D. Dubois and H. Prade. Using fuzzy sets in flexible querying: why and how? In T. Andreasen, H. Christiansen, and H.L. Larsen, editors, *Flexible Query Answering Systems*, pages 45–60. Kluwer Academic Publishers, 1997.

[5] D. Dubois and H. Prade. Bipolarity in flexible querying. In T. Andreasen, A. Motro, H. Christiansen, and H.L. Larsen, editors, *Flexible Query Answering Systems, FQAS 2002*, volume 2522 of *Lecture Notes in Computer Science*, pages 174–182. Springer, Berlin, Heidelberg, 2002.

[6] D. Dubois and H. Prade. Handling bipolar queries in fuzzy information processing. In Galindo [23], pages 97–114.

[7] D. Dubois and H. Prade. An overview of the asymmetric bipolar representation of positive and negative information in possibility theory. *Fuzzy Sets and Systems*, 160(10):1355–1366, 2009.

[8] M. Lacroix and P. Lavency. Preferences: Putting more knowledge into queries. In *Proceedings of the 13 International Conference on Very Large Databases*, pages 217–225, Brighton, UK, 1987.

[9] D. Dubois and H. Prade. Default reasoning and possibility theory. *Artificial Intelligence*, 35(2):243–257, 1988.

[10] R.R. Yager. Higher structures in multi-criteria decision making. *International Journal of Man-Machine Studies*, 36:553–570, 1992.

[11] R.R. Yager. Fuzzy logic in the formulation of decision functions from linguistic specifications. *Kybernetes*, 25(4):119–130, 1996.

[12] G. Bordogna and G. Pasi. Linguistic aggregation operators of selection criteria in fuzzy information retrieval. *International Journal of Intelligent Systems*, 10(2):233–248, 1995.

[13] P. Bosc and O. Pivert. Discriminated answers and databases: fuzzy sets as a unifying expression means. In *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 745–752, San Diego, USA, 1992.

[14] P. Bosc and O. Pivert. An approach for a hierarchical aggregation of fuzzy predicates. In *Proceedings of the Second IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'93)*, pages 1231–1236, San Francisco, USA, 1993.

[15] S. Zadrożny. Bipolar queries revisited. In V. Torra, Y. Narukawa, and S. Miyamoto, editors, *Modelling Decisions for Artificial Intelligence (MDAI 2005)*, volume 3558 of *LNAI*, pages 387–398. Springer-Verlag, Berlin, Heidelberg, 2005.

[16] S. Zadrożny and J. Kacprzyk. Bipolar queries and queries with preferences. In *Proceedings of the 17th International Conference on Database and Expert Systems Applications (DEXA'06)*, pages 415–419, Krakow, Poland, 2006. IEEE Comp. Soc.

[17] J.J. Dujmović. Partial absorption function. *Journal of teh University of Belgrade, EE Dept.*, 659:156–163, 1979.

[18] C. Tudorie. Qualifying objects in classical relational database querying. In Galindo [23], pages 218–245.

[19] J. Chomicki. Querying with intrinsic preferences. *Lecture Notes in Computer Science*, 2287:34–51, 2002.

[20] S. Zadrożny and J. Kacprzyk. Bipolar queries using various interpretations of logical connectives. In *Foundations of Fuzzy Logic and Soft Computing*, Lecture Notes in Computer Science, pages 181–190. Springer, 2007.

[21] J. Fodor and M. Roubens. *Fuzzy Preference Modelling and Multicriteria Decision Support*. Series D: System Theory, Knowledge Engineering and Problem Solving. Kluwer Academic Publishers, 1994.

[22] R. Mesiar and H. Thiele. On T-Quantifiers and S-Quantifiers. In V. Novak and I. Perfilieva, editors, *Discovering the World with Fuzzy Logic*, pages 310–326. Physica-Verlag, Heidelberg New York, 2000.

[23] J. Galindo, editor. *Handbook of Research on Fuzzy Information Processing in Databases*. Information Science Reference, New York, USA, 2008.

ISBN: 978-989-95079-6-8

A Fuzzy Variant of the Rand Index for Comparing Clustering Structures

Eyke Hüllermeier[1] and Maria Rifqi[2]

1. FB Mathematik/Informatik, Philipps-Universität Marburg
D-35032 Marburg, Germany
2. University Pierre et Marie Curie–Paris 6, CNRS UMR 7606, LIP6,
Paris, F-75016, France
Email: eyke@mathematik.uni-marburg.de, maria.rifqi@lip6.fr

***Abstract*— *In this paper, we introduce a fuzzy extension of the Rand index, a well-known measure for comparing two clustering structures. In contrast to an existing proposal, which is restricted to the comparison of a fuzzy partition with a non-fuzzy reference partition, our extension is able to compare two proper fuzzy partitions with each other. Elaborating on the formal properties of our fuzzy Rand index, we show that it exhibits desirable metrical properties.*

***Keywords*—** Clustering, Distance, Fuzzy Partition, Metric, Rand Index, Similarity

1 Introduction

The problem to compare two partitions of a set of objects occurs quite naturally in various domains, notably in data analysis and clustering. For example, one way to evaluate the result of a clustering algorithm is to compare the clustering structure produced by the algorithm with a correct partition of the data (which of course presumes that this information is available). In cluster analysis, so called *external evaluation measures* have been developed for this purpose [1, 2]. However, measures of that kind are not only of interest as evaluation criteria, i.e., for comparing a hypothetical partition with a true one. Instead, distance measures for partitions are interesting in their own right and can be used for different purposes.

In [3], for example, the authors consider the problem of clustering data in a very high-dimensional space. To increase efficiency, they propose to map the data into a low-dimensional space first and to cluster the transformed data thus obtained afterward. In this context, a distance measure for clustering structures (partitions) is useful to measure the loss of information incurred by the data transformation: If the transformation is (almost) lossless, the clustering structures in the two spaces should be highly similar, i.e., their distance should be small. On the other hand, a significant difference between the two partitions would indicate that the transformation does have a strong effect in the sense of distorting the structure of the data set.

Even though a large number of evaluation criteria and similarity indexes for clustering structures have been proposed in the literature, their extension to the case of fuzzy partitions has received much less attention so far. This is especially true for external evaluation criteria and measures comparing two clustering structures, whereas *internal criteria* for evaluating a single partition[1] have been studied more thoroughly (see, e.g., [4] and [5] for early proposals).

In a recent paper by Campello [6], the author has proposed an extension of the *Rand index* [7], a well-known measure of similarity between two partitions of a data set. Even though Campello's proposal is quite interesting, it also exhibits a number of disadvantages. Most notably, it is properly defined only for the comparison of a fuzzy partition with a non-fuzzy reference partition. It is true that this restriction can be tolerated if the index is used as an external evaluation criterion since, as correctly argued by the author, a reference partition provided by an external source is typically non-fuzzy. Yet, our example above has clearly shown that there is also a need for measures comparing two fuzzy partitions.

In this paper, we propose an alternative extension of the Rand index (which is, in principle, also applicable to related similarity measures for clustering structures). As opposed to Campello's proposal, our variant is able to compare two proper fuzzy partitions with each other. Moreover, we study our fuzzy Rand index from a formal point of view and show that it satisfies the desirable properties of a metric (when being used as a distance function).

The remainder of the paper is organized as follows. In the next section, we briefly review the proposal of Campello and discuss some of its properties in a critical way. In Section 3, we introduce our new measure and elaborate on its formal properties. The paper concludes with a short summary and an outlook on future work in Section 4.

2 Review of Campello's Proposal

Before reviewing Campello's proposal and discussing some of its properties, we briefly recall the original definition of the Rand index.

2.1 The Rand Index

Let $\mathbf{P} = \{P_1, \ldots, P_k\} \subset 2^X$ and $\mathbf{Q} = \{Q_1, \ldots, Q_\ell\} \subset 2^X$ be two (crisp) partitions of a finite set $X = \{x_1, x_2 \ldots x_n\}$ with n elements, which means that $P_i \neq \emptyset$, $P_i \cap P_j = \emptyset$ for all $1 \leq i \neq j \leq k$, and $P_1 \cup P_2 \cup \ldots \cup P_k = X$ (and analogously for $\mathbf{Q}$). Let

$$C = \{ (x_i, x_j) \in X \times X \mid 1 \leq i < j \leq n \}$$

denote the set of all tuples of elements in X.[2] We say that two elements $(x, x') \in C$ are *paired* in $\mathbf{P}$ if they belong to the same

[1] Typically, such criteria compare the intra-cluster variability, i.e., the variability among objects within the same cluster (which should be small) with the inter-cluster variability, i.e., the variability among objects from different clusters (which should be high).

[2] Since we consider unordered tuples, we should more correctly write $\{x_i, x_j\}$ instead of (x_i, x_j).

ISBN: 978-989-95079-6-8

cluster, i.e., if there is a cluster $P_i \in \mathbf{P}$ such that $x \in P_i$ and $x' \in P_i$. Moreover, we distinguish the following subsets of C:

- $C_1 \equiv$ the set of tuples $(x,x') \in C$ that are paired in $\mathbf{P}$ and paired in $\mathbf{Q}$;
- $C_2 \equiv$ the set of tuples $(x,x') \in C$ that are paired in $\mathbf{P}$ but not paired in $\mathbf{Q}$;
- $C_3 \equiv$ the set of tuples $(x,x') \in C$ that are not paired in $\mathbf{P}$ but paired in $\mathbf{Q}$;
- $C_4 \equiv$ the set of tuples $(x,x') \in C$ that are neither paired in $\mathbf{P}$ nor in $\mathbf{Q}$.

Obviously, $\{C_1,C_2,C_3,C_4\}$ is a partition of C, and $a+b+c+d = |C| = n(n-1)/2$, where

$$a = |C_1|,\ b = |C_2|,\ c = |C_3|,\ d = |C_4|. \tag{1}$$

The tuples $(x,x') \in C_1 \cup C_4$ are the *concordant* pairs, i.e., the pairs for which there is agreement between $\mathbf{P}$ and $\mathbf{Q}$, while the tuples $(x,x') \in C_2 \cup C_3$ are the *discordant* pairs for which the two partitions disagree. The Rand index is defined by the number of concordant pairs divided by the total number of pairs:

$$\mathrm{R}(\mathbf{P},\mathbf{Q}) = \frac{a+d}{a+b+c+d} \tag{2}$$

Thus defined, the Rand index is a similarity measure which assumes values between 0 and 1. It can easily be turned into a distance function by defining

$$\mathrm{D}_R(\mathbf{P},\mathbf{Q}) = 1 - \mathrm{R}(\mathbf{P},\mathbf{Q}) = \frac{b+c}{a+b+c+d}\ .$$

It is worth mentioning that D_R satisfies the classical properties of a distance (reflexivity, separation, symmetry, and triangular inequality).

2.2 Campello's Fuzzy Rand Index

The aim of Campello's paper is to extend the Rand index to the case of fuzzy partitions. To this end, he first reformulates it within a set-theoretic framework. An extension to the fuzzy case can then be accomplished in a straightforward way by using generalized set-theoretical operators. Recall that $k = |\mathbf{P}|$ and $\ell = |\mathbf{Q}|$, and consider the following sets:

- $V \equiv$ the set of pairs $(x,x') \in C$ that belong to the same cluster in $\mathbf{P}$; it can be expressed as $V = \bigcup_{i=1\ldots k} V_i$, where V_i is the set of pairs that both belong to the i-th cluster $P_i \in \mathbf{P}$.
- $W \equiv$ the set of pairs $(x,x') \in C$ that belong to different clusters in $\mathbf{P}$; it can be expressed as $W = \bigcup_{1\le i\ne j\le k} W_{ij}$, where W_{ij} is the set of pairs such that $x \in P_i$ and $x' \in P_j$.
- $Y \equiv$ the set of pairs $(x,x') \in C$ that belong to the same cluster in $\mathbf{Q}$; it can be expressed as $Y = \bigcup_{i=1\ldots \ell} Y_i$, where Y_i is the set of pairs that both belong to the i-th cluster $Q_i \in \mathbf{Q}$.
- $Z \equiv$ the set of pairs $(x,x') \in C$ that belong to different clusters in $\mathbf{Q}$; it can be expressed as $Z = \bigcup_{1\le i\ne j\le \ell} Z_{ij}$, where Z_{ij} is the set of pairs such that $x \in Q_i$ and $x' \in Q_j$.

The Rand index can directly be written in terms of the cardinalities of these sets, since the four quantities (1) are obviously given by

$$a = |V \cap W|, \qquad b = |V \cap Z|,$$
$$c = |W \cap Y|, \qquad d = |W \cap Z|.$$

In the fuzzy case, the above sets become fuzzy sets. Let $P_i(x) \in [0,1]$ denote the degree of membership of element $x \in X$ in the cluster $P_i \in \mathbf{P}$. The sets V, W, Y, and Z can then be defined through fuzzy-logical expressions involving a t-norm $\top$ and t-conorm $\bot$:

$$\begin{aligned} V(x,x') &= \bot_{i=1}^{k} \top(P_i(x),P_i(x')) \\ W(x,x') &= \bot_{1\le i\ne j\le k} \top(P_i(x),P_j(x')) \\ Y(x,x') &= \bot_{i=1}^{\ell} \top(Q_i(x),Q_i(x')) \\ Z(x,x') &= \bot_{1\le i\ne j\le \ell} \top(Q_i(x),Q_j(x')) \end{aligned} \tag{3}$$

Moreover, defining the intersection of sets by the t-norm combination of membership degrees and resorting to the commonly used sigma-count principle [8] for defining set cardinality, one obtains

$$\begin{array}{rcccl} a & = & |V \cap Y| & = & \sum_{(x,x')\in C} \top(V(x,x'),Y(x,x')) \\ b & = & |V \cap Z| & = & \sum_{(x,x')\in C} \top(V(x,x'),Z(x,x')) \\ c & = & |W \cap Y| & = & \sum_{(x,x')\in C} \top(W(x,x'),Y(x,x')) \\ d & = & |W \cap Z| & = & \sum_{(x,x')\in C} \top(W(x,x'),Z(x,x')) \end{array} \tag{4}$$

As before, the Rand index can then be defined as in (2), namely as the fraction

$$\frac{a+d}{a+b+c+d}\ .$$

2.3 Properties of Campello's Fuzzy Rand Index

Having defined a similarity or, equivalently, a distance function, it is natural to ask for desirable metrical properties of that function. In the case of the above fuzzy Rand index, however, this question has to be considered with caution, since Campello is actually only interested in comparing a fuzzy partition $\mathbf{P}$ with a non-fuzzy partition $\mathbf{Q}$. And indeed, formal properties of the measure are not investigated in his paper.

On the other hand, it is noted by Campello himself that, formally, the measure can in principle be applied to compare two fuzzy partitions. When doing so, however, it turns out quickly that it fails to be a proper metric. In fact, it does not even satisfy reflexivity, the perhaps most basic axiom: Even for two identical partitions $\mathbf{P}$ and $\mathbf{Q}$, the quantities b and c in (4) will generally not vanish, a necessary condition for having $\mathrm{R}(\mathbf{P},\mathbf{Q}) = 1$.

Consider, for example, the simple fuzzy partition $\mathbf{P}$ illustrated in Fig. 1, which consists of two clusters P_1 and P_2. Instead of a hard boundary, there is a "soft" transition between P_1 and P_2; the elements x_1, x_2, x_3, and x_4 partially belong to both clusters and have membership degrees, respectively, of $3/4$, $1/2$, $1/2$, $1/4$ in P_1 and $1/4$, $1/2$, $1/2$, $3/4$ in P_2. Comparing $\mathbf{P}$ to itself in terms of the fuzzy Rand index, we obtain $\mathrm{R}(\mathbf{P},\mathbf{P}) < 1$.

Upon closer examination, it seems that the core principle of Campello's extension is not suitable for comparing partitions in a fuzzy sense. In fact, despite being defined in terms

ISBN: 978-989-95079-6-8

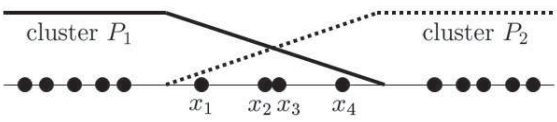


Figure 1: Illustration of a simple fuzzy partition of a subset of the reals (indicated by circles). The partition consists of two clusters, P_1 (left) and P_2 (right). While some elements definitely belong to only one of the clusters, some "critical" points in the middle have partial membership in both clusters.

of fuzzy logical formulas, the very idea of the approach has arguably a more "probabilistic flavor". This becomes especially obvious when using the product as a t-norm and the (bounded) sum as t-conorm. Then, if $P_i(x)$ is interpreted as the probability that x belongs to the i-th cluster, $V(x,x')$ is nothing else than the probability that x and x' are put in the same cluster, given that the two corresponding clusters are chosen *independently of each other* according to the distributions $(P_1(x),P_2(x)\dots P_k(x))$ and $(P_1(x'),P_2(x')\dots P_k(x'))$, respectively. Likewise, $W(x,x')$ is the probability that x and x' are put into different clusters.

Even if one accepts the probabilistic interpretation of a single membership degree, the additional assumption of independence is clearly not tenable. In fact, this property is obviously violated when comparing a partition with itself, since for each element $x \in X$, a cluster can then only be chosen once and not two times independently of each other. But even if $\mathbf{P}$ and $\mathbf{Q}$ are not identical, independence of cluster membership is in conflict with the topological relationships between the elements and clusters. In the example in Fig. 1, for instance, it is not reasonable to put x_1 and x_4 into cluster P_2 and x_2 and x_3 into cluster P_1. When putting elements independently of each other into clusters, however, this is a possible scenario. And indeed, this scenario contributes to Campello's fuzzy Rand index according to (3).

Seen from this point of view, one may even question the usefulness of the approach for its original purpose, namely the comparison of a fuzzy with a non-fuzzy partition. What the fuzzy partition in our example truly suggests is that we are uncertain about the boundary between the two clusters. More concretely, the fuzzy partition suggests four possible non-fuzzy partitions:

- $\mathbf{P}_1$ which puts the boundary left to x_1;
- $\mathbf{P}_2$ with boundary between x_1 and x_2;
- $\mathbf{P}_3$ with boundary between x_3 and x_4;
- $\mathbf{P}_4$ which puts the boundary right to x_4.

Thus, it seems reasonable to define an extension of the Rand index as an aggregation (e.g., weighted average) of the results of the non-fuzzy comparisons, namely

$$\mathrm{R}(\mathbf{P}_1,\mathbf{Q}), \mathrm{R}(\mathbf{P}_2,\mathbf{Q}), \mathrm{R}(\mathbf{P}_3,\mathbf{Q}), \mathrm{R}(\mathbf{P}_4,\mathbf{Q}).$$

In Campello's approach, there are not 4 but 16 scenarios which have an influence on the result, since each of the four cluster memberships is determined independently of each other. In general, the result will therefore be different. In fact, differences already occur for single pairs of elements. For example, since x_2 and x_3 are always in the same cluster in $\mathbf{P}_1,\dots,\mathbf{P}_4$, it is natural to say that they are paired with degree 1. According to Campello's approach, however, the degree to which x_2 and x_3 are in the same cluster in $\mathbf{P}$ is given by

$$V(x_2,x_3) = \bot(\top(1/2,1/2),\top(1/2,1/2)),$$

which corresponds to the truth degree of the proposition that "x_1 is put into P_1 AND x_2 is put into P_1 OR x_1 is put into P_2 AND x_2 is put into P_2". In general, this degree will be < 1 (except for special $(\top,\bot)$-combinations such as $\top = \min$ and $\bot =$ bounded sum).

3 A New Fuzzy Rand Index

In this section, we propose a new fuzzy variant of the Rand index which is able to compare any pair of fuzzy partitions and, moreover, has desirable metric properties. In the following, we focus on the view of the Rand index as a distance function. Thanks to the affine transformation $\mathrm{D}_R = 1 - \mathrm{R}$, all results can directly be transferred to the original conception as a measure of similarity.

3.1 Definition

Given a fuzzy partition $\mathbf{P} = \{P_1,P_2\dots P_k\}$ of X, each element $x \in X$ can be characterized by its membership vector

$$\mathbf{P}(x) = (P_1(x),P_2(x)\dots P_k(x)) \in [0,1]^k, \tag{5}$$

where $P_i(x)$ is the degree of membership of x in the i-th cluster P_i. We define a fuzzy equivalence relation on X in terms of a similarity measure on the associated membership vectors (5). Generally, this relation is of the form

$$E_{\mathbf{P}}(x,x') = 1 - \|\mathbf{P}(x) - \mathbf{P}(x')\|, \tag{6}$$

where $\|\cdot\|$ is a proper distance on $[0,1]^k$. The basic requirement on this distance is that it yields values in $[0,1]$. The relation (6) generalizes the equivalence relation induced by a conventional partition (where each cluster forms an equivalence class). In passing, we note that this definition is invariant toward a permutation (renumbering) of the clusters in $\mathbf{P}$, which is clearly a desirable property.

Now, given two fuzzy partitions $\mathbf{P}$ and $\mathbf{Q}$, the idea is to generalize the concept of concordance as follows. We consider a pair (x,x') as being concordant in so far as $\mathbf{P}$ and $\mathbf{Q}$ agree on their degree of equivalence. This suggest to define the *degree of concordance* as

$$1 - |E_{\mathbf{P}}(x,x') - E_{\mathbf{Q}}(x,x')| \in [0,1]. \tag{7}$$

Analogously, the *degree of discordance* is

$$|E_{\mathbf{P}}(x,x') - E_{\mathbf{Q}}(x,x')| \ .$$

Our distance measure on fuzzy partitions is then defined by the normalized sum of degrees of discordance:

$$d(\mathbf{P},\mathbf{Q}) = \frac{\sum_{(x,x')\in C} |E_{\mathbf{P}}(x,x') - E_{\mathbf{Q}}(x,x')|}{n(n-1)/2} \tag{8}$$

Likewise,

$$1 - d(\mathbf{P},\mathbf{Q}) \tag{9}$$

corresponds to the normalized degree of concordance and, therefore, is a direct generalization of the original Rand index.

ISBN: 978-989-95079-6-8

3.2 Formal Properties

In this section, we first show that our proposal is indeed a proper generalization of the Rand index. Afterward, we study the metrical properties of the measure.

Proposition: *In the case where* **P** *and* **Q** *are non-fuzzy partitions, the measure (9) reduces to the original Rand index.*

Proof: In the non-fuzzy case, the membership vectors (5) are 0/1-vectors. More specifically, each vector has a single entry $P_i(x) = 1$, while all other entries are 0. Consequently, the fuzzy equivalence (6) reduces to the conventional equivalence, that is, $E_{\mathbf{P}}(x,x') = 1$ if x and x' are in the same cluster and $E_{\mathbf{P}}(x,x') = 0$ otherwise. Likewise, (7) yields 1 if (x,x') is a concordant pair and 0 otherwise. Consequently, the measure (9) is the (normalized) sum of concordant pairs and, therefore, equals the original Rand index. □

Recall that a non-negative $U^2 \to \mathbb{R}$ mapping $d(\cdot)$ is called a metric on U if is satisfies the following properties for all $u,v,w \in U$:

- Reflexivity: $d(u,u) = 0$
- Separation: $d(u,v) = 0$ implies $u = v$
- Symmetry: $d(u,v) = d(v,u)$
- Triangle inequality: $d(u,w) \leq d(u,v) + d(v,w)$

The properties of reflexivity and symmetry are quite obviously valid for our measure (8). To show the triangle inequality, consider three fuzzy partitions **P**, **Q**, **R** and fix a single tuple $(x,x') \in C$. Let

$$a = E_{\mathbf{P}}(x,x'),\ b = E_{\mathbf{Q}}(x,x'),\ c = E_{\mathbf{R}}(x,x').$$

Since a, b, and c are real numbers (from the unit interval), and the simple difference on the reals satisfies the triangle inequality, we have $|a-c| \leq |a-b| + |b-c|$. Now, since this inequality holds for each pair $(x,x') \in C$, it remains valid when summing over all these pairs. In other words, it is also satisfied by (8), which means that

$$d(\mathbf{P},\mathbf{R}) \leq d(\mathbf{P},\mathbf{Q}) + d(\mathbf{Q},\mathbf{R}).$$

The separation property is not immediately valid for (8). Roughly speaking, this is due to the fact that, by mapping elements to their membership vectors (5), some information about the partition itself is lost. In particular, it is possible that two partitions, even though they are not identical, cannot be distinguished in terms of the distances between these vectors.

Nevertheless, we can guarantee the separation property by restricting to a reasonable subclass of fuzzy partitions. We call a fuzzy partition $\mathbf{P} = \{P_1, P_2 \ldots P_k\}$ *normal*, if it satisfies the following:

N1 For each $x \in X$: $P_1(x) + \ldots + P_k(x) = 1$.

N2 For each $P_i \in \mathbf{P}$, there exists an $x \in X$ such that $P_i(x) = 1$.

In other words, we consider Ruspini partitions [9] and assume that each cluster has a prototypical element. Moreover, we assume the following equivalence relation on X:

$$E_{\mathbf{P}}(x,x') = 1 - \frac{1}{2}\sum_{i=1}^{k} |P_i(x) - P_i(x')|. \tag{10}$$

Note that $0 \leq E_{\mathbf{P}}(x,x') \leq 1$ for all $(x,x') \in X^2$ under assumption N1.

Now, consider two normal fuzzy partitions **P** and **Q**, and suppose that $d(\mathbf{P},\mathbf{Q}) = 0$. According to our definition of $d(\cdot)$, this obviously means that

$$E_{\mathbf{P}}(x,x') = E_{\mathbf{Q}}(x,x') \tag{11}$$

for all $(x,x') \in C$. We call a set $\{p_1, p_2 \ldots p_k\} \subset X$ a prototype set for **P**, if $P_i(p_i) = 1$ for all $i = 1,\ldots,k$ (note that a prototype set is not necessarily unique). We distinguish two cases.

(a) There are no identical prototype sets for **P** and **Q** (note that this is necessarily the case if **P** and **Q** have a different number of clusters). Then, we can find elements $x,x' \in X$ such that x and x' are prototypes for **P** but not for **Q**. Note that N1 and N2 jointly imply that a prototype is represented by a 0/1 membership vector, and that $|\mathbf{P}(x) - \mathbf{P}(x')| = 1$ for two different prototypes x and x'. Moreover, these properties imply that the extreme distance of 1 can *only* be assumed for prototypes, whereas $|\mathbf{P}(x) - \mathbf{P}(x')| < 1$ if either x or x' is not a prototype. Thus, it follows that $E_{\mathbf{P}}(x,x') = 0$ and $E_{\mathbf{Q}}(x,x') > 0$, which means that condition (11) is violated. Hence, we have constructed a contradiction with the assumption that $d(\mathbf{P},\mathbf{Q}) = 0$.

(b) There are identical prototype sets $\{p_1,\ldots,p_k\} = \{q_1,\ldots,q_\ell\}$, respectively, for **P** and **Q** (which means that $k = \ell$, i.e., **P** and **Q** do have the same number of clusters). We can then establish a one-to-one correspondence between prototypes such that, without loss of generality, $p_i = q_i$ for $i = 1,\ldots,k$. From properties N1 and N2, it follows that the membership degree of any element x in the cluster P_i is a function of $E_{\mathbf{P}}(x,p_i)$. In fact, noting that $\mathbf{P}(p_i)$ is a 0/1 vector with a single 1 on position i, we get

$$\begin{aligned} E_{\mathbf{P}}(x,p_i) &= 1 - \frac{1}{2}\sum_{j=1}^{k} |P_j(x) - P_j(p_i)| \\ &= 1 - \frac{1}{2}\left((1-P_i(x)) - \sum_{j\neq i} P_j(x)\right) \\ &= 1 - \frac{1}{2}\left((1-P_i(x)) - (1-P_i(x))\right) \\ &= P_i(x). \end{aligned}$$

From (11), it thus follows that $P_i(x) = Q_i(x)$ for all $x \in X$, i.e., the i-th cluster in **P** and the i-th cluster in **Q** are identical. Since this holds for all $i \in \{1,2\ldots k\}$, we have shown that $\mathbf{P} = \mathbf{Q}$.

The above results can be summarized as follows.

Theorem: *The distance function (8) on fuzzy partitions is a pseudometric, i.e., it is reflexive, symmetric, and subadditive. Moreover, on the restricted class of normal fuzzy partitions or, more specifically, under the assumptions N1, N2, and (10), it also satisfies the separation property and, therefore, is a metric.*

4 Summary and Outlook

We have introduced a generalization of the Rand index for comparing two fuzzy clustering structures. Elaborating on the formal properties of our measure, we have shown that it is a

ISBN: 978-989-95079-6-8

pseudo-metric and, on a subclass of fuzzy partitions obeying certain normality assumptions, even a metric.

In future work, we plan to extent our approach to other similarity measures for (non-fuzzy) clustering structures which are related to the Rand index in the sense of being defined in terms of the same basic quantities, namely the numbers a, b, c, and d of concordant and discordant object pairs. An example of such a measure is the Jaccard coefficient, which is defined as $a/(a+b+c)$.

References

[1] M. Halkidi, Y. Batistakis, and M. Vazirgiannis. Clustering validity checking methods: Part I. *ACM SIGMOD Record*, 31(2):40–45, 2002.

[2] M. Halkidi, Y. Batistakis, and M. Vazirgiannis. Clustering validity checking methods: Part II. *ACM SIGMOD Record*, 31(3):19–27, 2002.

[3] J. Beringer and E. Hüllermeier. Fuzzy clustering of parallel data streams. In J. Valente de Oliveira and W. Pedrycz, editors, *Advances in Fuzzy Clustering and Its Application*, pages 333–352. John Wiley and Sons, 2007.

[4] MP. Windham. Cluster validity for the fuzzy c-means algorithm. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 4(4):357–363, 1982.

[5] XL. Xie and GA. Beni. Validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 3(8):841–846, 1991.

[6] R. J. G. B. Campello. A fuzzy extension of the Rand index and other related indexes for clustering and classification assessment. *Pattern Recognition Letters*, 28(7):833–841, 2007.

[7] William M. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336):846–850, 1971.

[8] A. DeLuca and S. Termini. A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory. *Information and Control*, 20:301–312, 1972.

[9] E.H. Ruspini. A new approach to clustering. *Information Control*, 15:22–32, 1969.

ISBN: 978-989-95079-6-8

Fuzzy Modeling of Labeled Point Cloud Superposition for the Comparison of Protein Binding Sites

Thomas Fober and Eyke Hüllermeier

FB Mathematik/Informatik, Philipps-Universität Marburg
D-35032 Marburg, Germany
Email: {thomas,eyke}@mathematik.uni-marburg.de

Abstract— *Geometric objects are often represented approximately in terms of a finite set of points in three-dimensional Euclidean space. In this paper, we extend this representation to what we call labeled point clouds. A labeled point cloud is a finite set of points, where each point is not only associated with a position in three-dimensional space, but also with a discrete class label that represents a specific property. This type of model is especially suitable for modeling biomolecules such as proteins and protein binding sites, where a label may represent an atom type or a physico-chemical property. Proceeding from this representation, we address the question of how to compare two labeled points clouds in terms of similarity. Using fuzzy modeling techniques, we develop a suitable similarity measure as well as an efficient evolutionary algorithm to compute it. Finally, an application study is presented in which the approach is used to classify protein binding sites.*

Keywords— structural bioinformatics, proteins, similarity, classification

1 Introduction

Geometric objects are often represented in terms of a set of points in three-dimensional Euclidean space. This type of representation is finite and hence approximate (even though the number of points can become very large, as for example in laser range scanning), focusing on the most important characteristics of the object while ignoring less important details. A well-known example of a representation of this kind is the *Molfile* format [1], where molecules are described in terms of the spatial coordinates of all atoms. However, since not only the position but also the type of an atom is of interest, this representation is not a simple point cloud. Likewise, other biomolecular structures, such as proteins and protein binding sites, are not only characterized by their geometry but also by additional features, such as physico-chemical properties. In this paper, we therefore introduce the concept of a *labeled point cloud*. A labeled point cloud is a finite set of points, where each point is not only associated with a position in three-dimensional space, but also with a discrete class label that represents a specific property.

Since theory formation in the biological sciences is largely founded on similarity-based and analogical reasoning principles, the comparison of two (or more) objects with each other is a fundamental problem in bioinformatics. To compare two point clouds, the authors in [2] make use of a measure based on the *Gromov-Hausdorff distance* of sets. This approach is limited to unlabeled point clouds, however. Another possibility is to transform a labeled point cloud into a (labeled) graph first, capturing, in one way or the other, geometrical information in terms of edges, and to apply graph matching techniques afterward. This strategy was recently proposed in [3], where the use of *graph kernels* as similarity measures [4, 5, 6] has been especially advocated. At first sight, this idea looks appealing, especially since methods for comparing graphs abound in the literature. Nevertheless, it also comes with a number of disadvantages. For example, many techniques for matching and comparing graphs capture aspects of similarity which are reasonable for graphs but not necessarily for geometric objects. Besides, graph matching techniques are typically quite complex from a computational point of view.

Perhaps most importantly, however, a graph representation captures the geometrical information only in an *implicit* way, namely through the presence, absence, and possibly the labeling of edges. Moreover, the transformation is often not even lossless. Matching objects while obeying geometrical constraints can then become troublesome, since the geometrical information is not explicitly available. Instead, it must be reconstructed from the graph representation whenever needed.

As an alternative to an indirect approach of that kind, we therefore propose the method of *labeled point cloud superposition* (LPCS), which operates on labeled point clouds directly. Thus, it preserves as much geometrical information as possible and facilitates the exploitation thereof. Related to the concept of an LPCS, we introduce a similarity measure which makes use of modeling techniques from fuzzy set theory. This measure proceeds from the idea of equivalence (inclusion) of point clouds in a set-theoretic sense, but is tolerant toward exceptions (on the level of label information) and geometric deformations.

The remainder of the paper if organized as follows. Subsequent to a brief introduction to protein binding sites and their representation in Section 2, we introduce the concept of LPCS in Section 3. The problem of computing an LPCS is then addressed in Section 4, where an evolution strategy is proposed for this purpose. Section 5 is devoted to the experimental validation of the approach, and Section 6 concludes the paper.

2 Modeling Protein Binding Sites

In this paper, our special interest concerns the modeling of protein binding sites. More specifically, our work builds upon CavBase [7], a database for the automated detection, extraction, and storing of protein cavities (hypothetical binding sites) from experimentally determined protein structures (available through the PDB). In CavBase, a set of points is used as a first approximation to describe a binding pocket. The database currently contains 113,718 hypothetical binding sites that have been extracted from 23,780 publicly available protein structures using the LIGSITE algorithm [8].

ISBN: 978-989-95079-6-8

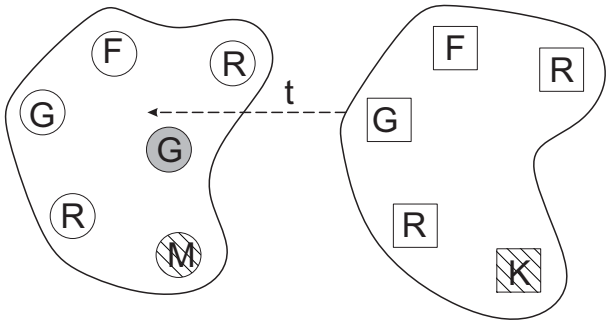


Figure 1: Two point clouds A (left, points as circle) and B (right, points as squares): The intra-point distances are the same in both point clouds, except for the additional gray point in A. Labels are depicted as letters within the circles and boxes, respectively.

The geometrical arrangement of the pocket and its physicochemical properties are first represented by predefined *pseudocenters* – spatial points that represent the center of a particular property. The type and the spatial position of the centers depend on the amino acids that border the binding pocket and expose their functional groups. They are derived from the protein structure using a set of predefined rules [7]. As possible types for pseudocenters, hydrogen-bond donor, acceptor, mixed donor/acceptor, hydrophobic aliphatic, metal ion, pi (accounts for the ability to form π–π interactions) and aromatic properties are considered.

Pseudocenters can be regarded as a compressed representation of areas on the cavity surface where certain protein-ligand interactions are experienced. Consequently, a set of pseudocenters is an approximate representation of a spatial distribution of physicochemical properties. Obviously, just like in the case of Molfile, this representation is already in the form of a labeled point cloud: pseudocenters are given with their coordinates and labels, so that no further transformation is needed.

3 Labeled Point Cloud Superposition

Intuitively, two labeled point clouds are similar if they can be spatially superimposed. That is, by fixing the first and "moving" the second one (as a whole, i.e., without changing the internal arrangement of points) in a proper way, an approximate superposition of the two structures is obtained. More specifically, we will say that two point clouds are well superimposed if, for each point in one of the structures, there exists a point in the other cloud which is spatially close and has the same label. As an illustration, the example in Fig. 1 shows two point clouds A and B, for simplicity only in two dimensions. By moving B to the left (or A to the right), a superposition can be found so that, except for the hatched and gray nodes, all points in A spatially coincide with a corresponding point in B having the same label, and vice versa. So, A and B can be considered as being similar, at least to some extent.

More formally, let

$$A = \{(x_1, \ell(x_1)), \ldots, (x_m, \ell(x_m))\}$$

be a point cloud consisting of m points $x_i = (x_{i1}, x_{i2}, x_{i3}) \in \mathbb{R}^3$ with associated label $\ell(x_i) \in \mathcal{L}$, where $\mathcal{L}$ is a discrete set of labels (in the context of modeling protein binding sites, as discussed in the previous section, $\mathcal{L}$ is given by the seven types of pseudocenters). Moreover, let

$$B = \{(y_1, \ell(y_1)), \ldots, (y_n, \ell(y_n))\}$$

be a second point cloud to be compared with A. In the following, we define a function $\mathrm{SIM}(\cdot,\cdot)$ that returns a degree of similarity between two such structures A and B.

Roughly speaking, we consider similarity as a generalized (fuzzy) equivalence, which we in turn reduce to two inclusion relations, namely the inclusion of A in B and, vice versa, of B in A. Thus, we are first of all interested in whether each point $y \in B$ is also present in A (and each point $x \in A$ also present in B). For a fixed $y \in B$, we define the membership degree of this point in A by

$$\mu_A(y) = \exp\left(-\gamma \cdot d(y, A)\right) \quad , \tag{1}$$

where

$$d(y, A) = \min_{\substack{x \in A \\ \ell(x) = \ell(y)}} \|y - x\|_1$$

is the distance between a point $y \in B$ and the closest point $x \in A$ having the same label ($d(y,A) = \infty$ and hence $\mu_A(y) = 0$ if no such point exists); for $x \in A$, $\mu_B(x)$ and $d(x,B)$ are defined analogously.

In its proper sense, the inclusion of a set B in a set A means that *each* point $y \in B$ is also contained in A or, stated differently, if a point y is in B, then it is also present in A. If membership is a matter of degree, i.e., if A and B are fuzzy sets, this condition is often formalized in terms of a fuzzy implication [9]:

$$\min_{y \in B} \left(\mu_B(y) \to \mu_A(y)\right) \quad .$$

Here, the minimum operator plays the role of a generalization of the universal quantifier. In our case, $\mu_B(y) \equiv 1$, so that the above expression can be simplified as follows:

$$inc(B, A) = \min_{y \in B} \mu_A(y) \quad . \tag{2}$$

However, a universal quantification (modeled by the min operator) is too strict in our biological context, where data is typically inexact and noisy. To relax this definition of fuzzy inclusion, we replace the minimum by a fuzzy quantifier Q, which is specified in the form of a non-decreasing $[0,1] \to [0,1]$ mapping [10, 11]. This leads to

$$inc(B, A) = \min_{i=1\ldots|B|} \max\{Q(i/|B|), m_i\} \quad ,$$

where m_i is the i-th largest membership degree in the fuzzy set $\{\mu_A(y) \,|\, y \in B\}$. (Note that we recover (2) for Q defined by $Q(1) = 1$ and $Q(t) = 0$ for $0 \leq t < 1$.) Here, we simply take Q as the identical mapping $t \mapsto t$. Roughly speaking, $inc(B,A)$ thus defined can be interpreted as the generalized truth degree of the proposition that A is *almost* contained in B. The degree of inclusion of A in B, $inc(A,B)$, is defined analogously.

As mentioned above, the idea of our approach is to define the similarity between two labeled point clouds in terms of the best superposition of these two clouds. Therefore, let $\mathrm{TF}(\cdot, t)$ be a function that moves a point cloud via rotation and translation, as specified by the six-dimensional vector $t = (\theta_1, \theta_2, \theta_3, \delta_1, \delta_2, \delta_3) \in [0, 2\pi]^3 \times \mathbb{R}^3$. Thus,

$$B^* = \mathrm{TF}(B, t) = \{(y_1^*, \ell(y_1^*)), \ldots, (y_n^*, \ell(y_n^*))\}$$

ISBN: 978-989-95079-6-8

is the point cloud obtained by translating the point cloud B by $\delta = (\delta_1, \delta_2, \delta_3)$ (which means adding δ to each point $y \in B$) and rotating the result thus obtained by the angles θ_1, θ_2, and θ_3. Note that this operation leaves the label information unchanged (i.e., $\ell(y_i) = \ell(y_i^*)$). The position-invariant degree of inclusion of B in A is then given by

$$\mathrm{INC}(B,A) = \max_{t \in [0,2\pi]^3 \times \mathbb{R}^3} inc(\mathrm{TF}(B,t),A) \ , \qquad (3)$$

and $\mathrm{INC}(A,B)$ is defined analogously.

Based on these degrees, the similarity between A and B, in the sense of a generalized equivalence, can be defined as

$$\mathrm{SIM}(A,B) = \min\{\,\mathrm{INC}(A,B), \mathrm{INC}(B,A)\,\} \ . \qquad (4)$$

It is worth mentioning, however, that (4) is not always appropriate, especially if A and B greatly differ in size. In some applications, it makes sense to have a high similarity degree even if A is only a substructure of B, for example if A is a subpocket of B containing the most important catalytic residues (while the rest of the binding site B is functionally less important). Obviously, this is not guaranteed by (4). An interesting generalization, therefore, is to let

$$\begin{aligned} \mathrm{SIM}(A,B) = & \alpha \cdot \min\{\mathrm{INC}(A,B), \mathrm{INC}(B,A)\} + \\ & + (1-\alpha) \cdot \max\{\mathrm{INC}(A,B), \mathrm{INC}(B,A)\} \ . \end{aligned} \qquad (5)$$

Formally, this similarity measure can be motivated from a fuzzy logical point of view as follows. Considering the min (max) operator as a generalized conjunction (disjunction), the first (second) combination of the two inclusion degrees is the truth degree of the proposition that A is contained in B AND (OR) B is contained in A. A conjunctive combination of the two degrees of inclusion is obviously more demanding than a disjunctive one, as the former requires equality between A and B while the latter only requires inclusion of A in B or B in A. The measure (5), which formally corresponds to an OWA (ordered weighted average) combination of the two degrees of inclusion [12], achieves a trade-off between these two extreme aggregation modes, which is controlled by the parameter $\alpha \in [0,1]$: The closer α is to 0, the closer the aggregation is to the maximum, i.e., the less demanding it becomes. The optimal α is application-specific and depends on the purpose of the similarity measure.

4 Solving the LPCS Problem

The computation of the similarity (5) involves the solution of a real-valued optimization problem, namely the problem of finding an optimal vector t in (3) and, thus, an optimal point cloud superposition. The objective function to be maximized here is highly non-linear and multimodal. As an illustration, Fig. 2 shows the objective function obtained for the superposition of a randomly generated two-dimensional point cloud A (in which all points have the same label) with itself. This function maps each two-dimensional translation vector $t = (x,y)$ to the corresponding similarity degree between $\mathrm{TF}(A)$ and A (where we used $\alpha = 0.5$ in (5) and did not consider rotation). As can be seen, there is a sharp peak at $t = (0,0)$, which corresponds to the optimal superposition. Surrounding this solution, however, there are also many local optima.

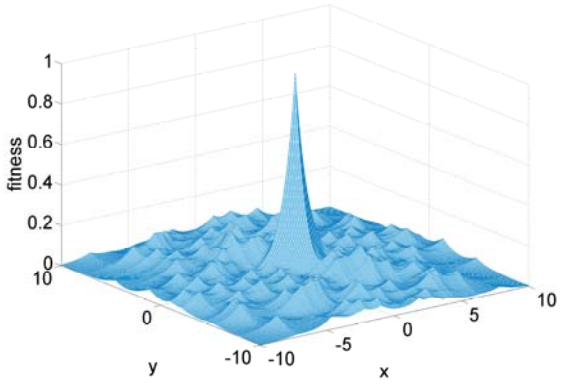


Figure 2: Example of an LPCS objective function.

The problem of local optima also becomes clear from the small example in Fig. 1. Moving the point cloud A from left to right, into the direction of B, has the following effect: First, a good superposition of two sub-clouds will be found, namely the right part of cloud A and the left part of cloud B. This results in a local maximum. Moving A further to the right leads to a larger local maximum (sub-clouds are growing), until the global maximum will eventually be reached.

4.1 Evolution Strategies

To solve the LPCS problem, we resort to *evolution strategies* (ES), a population-based, stochastic optimization method inspired by biological evolution and specifically developed for real-valued optimization problems [13]. An evolution strategy is based on a population, a set of μ (sub-optimal) candidate solutions that are initially spread randomly over the search space. In each generation, new solutions are generated by applying the genetic operators *recombination* and *mutation*. Recombination randomly selects ρ individuals from the current population and combines them to a new solution. Mutation takes this solution and shifts it randomly in the search space. An ES produces $\lambda = \lceil \mu \cdot \nu \rceil$ offsprings per iteration, so that this procedure has to be repeated λ times. A selection operator implements the "survival of the fittest" principle by picking the best individuals for the new population. There are two kinds of selection: The *plus*-selection chooses the best μ individuals among the offsprings plus the parents, while the *comma*-selection ignores the parent generation (this requires $\nu > 1$). A main advantage of the ES is its self-adaptation mechanism that controls the step sizes used in the mutation operator. One property of this mechanism (the advantage during optimization is obvious) is that step sizes decrease dramatically if the optimization reached a maximum. This property can be used as a qualitative termination criterion (stop when the largest step size falls below a given threshold).

Population-based optimization methods are especially advantageous for highly multimodal problems. Using a large population leads to an increased probability to generate a candidate solution in a region where the direction of descent points to the global maximum. Choosing the membership function (1) as a strictly monotone decreasing function which converges to zero ensures to have this direction in each point $t \in [0,2\pi]^3 \times \mathbb{R}^3$ and thus greatly simplifies the maximization problem. However, our experiments indicated that the solu-

ISBN: 978-989-95079-6-8

tion we found was most often only a local maximum. Therefore, we propose to use *fast restarts* of the ES. This means that the ES is started n times using comma-selection and weak termination criteria to achieve a large and quick but inexact exploration of our search space. We thus obtain n results in total. In a last step, we use the ES with plus-selection and strong termination criterion. Additionally, we include the best solution so far in the start population. The last run of the ES usually yields a globally optimal degree of similarity.

4.2 Complexity

Even though evolution strategies are generally known to be quite efficient solvers, the concrete complexity does of course depend on the application at hand. The application-specific part is the fitness function, i.e., the objective function to be optimized. This function has to be evaluated frequently and, therefore, is an important factor for the runtime. In our case, this function is given by the similarity measure (5), and its evaluation is strongly dominated by the nearest neighbor search which has to be conducted for each single point in both structures (recall that, according to (1), membership degrees are determined by the distance to closest points with the same label).

There exist a lot of data structures for supporting nearest neighbor search; see e.g. [14]. The most efficient among them need time $O(n\log^2 n)$ for construction and $O(\log^3 n)$ for answering a query. Unfortunately, we are not aware of an approach that allows for updating a data structure in an efficient and dynamic way. This would be desirable for our problem, in which the point clouds permanently change (the point cloud associated with an individual changes in each iteration). Instead, conventional approaches necessitate a construction from scratch in every iteration.

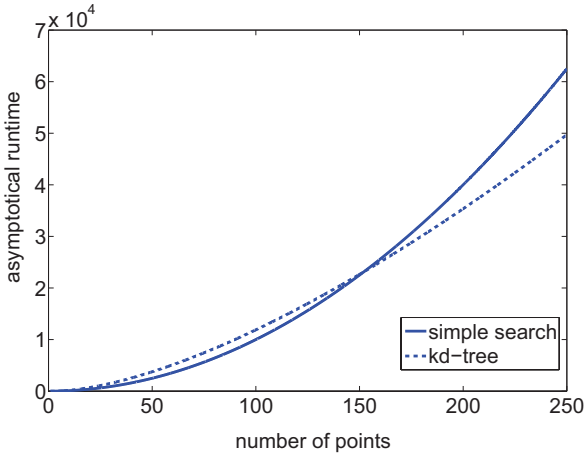


Figure 3: Runtime of a simple procedure and a more complex data structure as a function of the number of points.

Fig. 3 compares the runtimes, as a function of the number of points, for two approaches: (1) The use of a kd-tree data structure, which is reconstructed in each iteration and then used for query processing. (2) The use of a simple linear data structure, in which the points are stored in a fixed order. It needs linear instead of logarithmic time to answer a query but, on the other hand, does not cause additional costs for reconstruction. As can be seen, the use of a more complex approach pays off only for sufficiently large point clouds: The kd-tree reaches a break-even point at approximately 150 points.

In our application, we are mainly concerned with protein binding sites, which are characterized by around 180 points on average (even though much larger structures do of course exist). The use of a complex data structure did therefore not pay off. Nevertheless, we increased efficiency by hashing the points x_i of a point cloud, using the label $\ell(x_i) \in \mathcal{L}$ as a key. Since nearest neighbors are only searched among points having the same label, this obviously reduces runtime by a factor of approximately $|\mathcal{L}|$.

5 Experimental Results

5.1 Methods

In our experiments, we compared our novel method (LPCS) with existing graph-based approaches, namely the random walk (RW) kernel [6], the shortest path (SP) kernel [5], and the method of multiple graph alignment (MGA) recently introduced in [15]. Given two labeled points clouds as input, all these methods produce a degree of similarity as an output. Yet, for the graph-based approaches, it is of course necessary to transform a point cloud into a graph representation in a preprocessing step. This was done as as proposed in [15]:

1. each point is transformed into a node with corresponding node label
2. for each pair of nodes:
   (a) the Euclidean distance between both nodes is calculated
   (b) if the distance is below a certain threshold (here 11 Å to ensure connected graphs), an edge with weight equal to this distance is added

Our ES was restarted $n = 5$ times. The parameterization was optimized with the *sequential parameter optimization toolbox* [16] and was chosen as follows:

- inexact ES: $\mu = 30, \nu = 4, \rho = 2$, comma-selection, termination criteria: largest step size < 0.05, discrete recombination for strategy- and object-component.
- exact ES: $\mu = 30, \nu = 4, \rho = 6$, plus-selection, termination criteria: largest step size < 0.00001, intermediate recombination for object and discrete recombination for strategy-component.

A comprehensive explanation of the different ES parameters and operators can be found in [13].

For both variants we initialized the object-component in $[-150, 150]^3$ for translation and $[0, 2\pi]^3$ for rotation: The step sizes were initialized in $[5, 15]^3$ and $[1, \pi]^3$, respectively. The SP-kernel is parameter-free, the RW-kernel expects a parameter λ that is set to the largest degree of a node in the data set to ensure a geometric series during calculation, which results in a simpler evaluation [4]. Since the geometric information of real-world data is noisy, we also need a tolerance parameter ε to decide whether two edges have equal length (difference $\leq \varepsilon$) or not; in our experiments, we used $\varepsilon = 0.2$. For MGA, we chose the parameterization proposed in [15].

ISBN: 978-989-95079-6-8

The assessment of a similarity measure for biomolecular structures, such as protein binding sites, is clearly a non-trivial problem. In particular, since the concept of similarity by itself is rather vague and subjective, it is difficult to evaluate corresponding measures in an objective way. To circumvent this problem, we propose to evaluate similarity measures in an indirect way, namely by means of their performance in the context of nearest neighbor (NN) classification. The underlying idea is that, the better a similarity measure is, the better should be the predictive performance of an NN classifier using this measure for determining similar cases.

5.2 Data

One important problem in pharmaceutical chemistry is the identification of protein binding sites that bind a certain ligand. We selected two classes of binding sites that bind, respectively, to NADH or ATP. This gives rise to a binary classification problem: Given a protein binding site, predict whether it binds NADH or ATP.

More concretely, we compiled a set of 355 protein binding pockets representing two classes of proteins that share, respectively, ATP and NADH as a cofactor. To this end, we used CavBase to retrieve all known ATP and NADH binding pockets that were co-crystallized with the respective ligand. Subsequently, we reduced the set to one cavity per protein, thus representing the enzymes by a single binding pocket. As protein ligands adopt different conformations due to their structural flexibility, it is likely that the ligands in our data set are bound in completely different ways, hence the corresponding binding pocket does not necessarily share much structural similarity. We thus had to ensure the selection of binding pockets with ligands bound in similar conformation. To achieve this, we used the Kabsch algorithm [17] to calculate the root mean square deviation (RMSD) between pairs of ligand structures. Subsequently, we combined all proteins whose ligands yielded a RMSD value below a threshold of 0.2, thereby ensuring a certain degree of similarity. This value was chosen as a trade-off between data set size and similarity. Eventually, we thus obtained a two-class data set comprising 214 NADH-binding proteins and 141 ATP-binding proteins.

5.3 Results

The results of a leave-one-out cross validation, using the simple 1-NN classifier for prediction, are summarized in Table 1. As can be seen, the kernel-based methods (SP and RW) perform very poorly and are hardly better than random guessing. In terms of accuracy, MGA is much better, though still significantly worse than LPCS. In fact, LPCS performs clearly best on this problem.

Table 1: Accuracy and runtimes (in seconds with standard deviation, referring to a single comparison) of LPCS ($\alpha = 0.5$, with restarts like described above), MGA, RW, and SP on the NADH/APT data set.

Method	Accuracy	Runtime
MGA	0.7662	121.74 ± 418.02
SP	0.6056	$\mathbf{9.75 \pm 97.77}$
RW	0.5972	65.51 ± 89.07
LPCS	**0.9352**	20.04 ± 24.65

Table 2 furthermore shows how the performance of LPCS depends on the choice of the trade-off parameter α in (5). As can be seen, this parameter does indeed have an influence, even though the differences are not extreme. For this data set, α-values around 0.5 yield better results than extreme values close to 0 or 1; the optimal choice would be $\alpha = 0.7$. In practice, α can be considered as a tuning parameter to be adapted to the problem at hand (e.g., by means of a cross-validation on the training data).

Table 2: Accuracy of LPCS for different values of α in (5).

α	accuracy	α	accuracy
0	0.9042	0.6	0.9352
0.1	0.9183	0.7	0.9380
0.2	0.9126	0.8	0.9239
0.3	0.9154	0.9	0.9267
0.4	0.9267	1	0.9183
0.5	0.9352		

Regarding runtime, the experiment shows that LPCS is quite efficient, abandon restarts of the ES it would be the fastest of all alternatives, however with an increasing risk to get stuck in a local optimum. Using restarts, only SP is still a bit faster on average, however, it completely fails in terms of predictive accuracy. Even though we did not investigate this issue in a systematic way so far, our experience has shown that LPCS scales much better than typical graph-based approaches. This is hardly surprising, since the dimensionality of the LPCS optimization problem is constant (six parameters have to be optimized) and does not depend on the number of data points. It is true that the size of the point clouds does have an influence on the evaluation of the objective function, which involves a nearest neighbor search for each point. The increase in runtime is at most quadratic, however. As opposed to this, the complexity of graph-based methods such as MGA, in which graph matching is reduced to a combinatorial optimization problem, grows exponentially with the number of nodes.

6 Conclusions

In this paper, we have introduced labeled point cloud superposition (LPCS) as a novel tool for structural bioinformatics, namely as a method for comparing biomolecules on a structural level. Besides, using fuzzy modeling techniques, we have defined a related similarity measure. The concept of a labeled point cloud appears to be a quite natural representation for biological structures, especially since it is closely leaned on existing database formats. In comparison to other approaches, such as the prevalent graph-based methods, the modeling is hence simplified and does not involve any complex transformations. More importantly, a labeled point cloud preserves the full geometric information and makes it easily accessible to computational procedures.

A labeled point cloud superposition is a spatial "alignment" of two point clouds which is optimal in the sense of a given scoring (similarity) function. As for related problems in bioinformatics, such as sequence alignment, the computation of the similarity between two objects hence involves the solution of an optimization problem. To this end, we have proposed the

ISBN: 978-989-95079-6-8

use of an evolution strategy, an approach from the family of evolutionary algorithms, which appears to be especially suitable for this problem.

First experimental results with classification data are quite promising and suggest that our approach is able to compare protein binding sites in a reasonable way. In terms of classification accuracy, LPCS turned out to be significantly better than existing (graph-based) methods used for comparison. Moreover, even though it is computationally more complex than these methods for small data sets, it scales much better and becomes more efficient for larger data sets. This is due to the fact that, in contrast to graph-based methods, the search space does not depend on the size of the point clouds and remains low-dimensional.

References

[1] Johann Gasteiger and Thomas Engel. *Chemoinformatics*. Wiley-Vch, Weinheim, 2003.

[2] F. Mémoli and G. Sapiro. Comparing point clouds. In *Eurographics / ACM SIGGRAPH symposium on Geometry processing*, pages 32–40, Nice, France, 2004.

[3] Francis R. Bach. Graph kernels between point clouds. In *International Conference on Machine Learning*, pages 25–32, Helsinki, Finland, 2008.

[4] K. M. Borgwardt. *Graph Kernels*. PhD thesis, Ludwig-Maximilians-Universität München, Germany, 2007.

[5] K. M. Borgwardt and H. P. Kriegel. Shortest-path kernels on graphs. In *International Conference on Data Mining*, pages 74–81, Houston, Texas, 2005.

[6] Thomas Gärtner. A survey of kernels for structured data. *SIGKKD Explorations*, 5(1):49 – 58, 2003.

[7] S. Schmitt, D. Kuhn, and G. Klebe. A new method to detect related function among proteins independent of sequence and fold homology. *Journal of Molecular Biology*, 323(2):387–406, 2002.

[8] M. Hendlich, F. Rippmann, and G. Barnickel. LIGSITE: Automatic and efficient detection of potential small molecule-binding sites in proteins. *Journal of Molecular Graphics and Modelling*, 15:359–363, 1997.

[9] D. Sinha and E.R. Dougherty. Fuzzification of set inclusion: theory and applications. *Fuzzy Sets and Systems*, 55(1):15–42, 1993.

[10] L.A. Zadeh. A computational approach to fuzzy quantifiers in natural languages. *Comput. Math. Appl.*, 9:149–184, 1983.

[11] J. Fodor and R.R. Yager. Fuzzy set-theoretic operators and quantifiers. In D. Dubois and H. Prade, editors, *Fundamentals of Fuzzy Sets*, pages 125–194. Kluwer Academic Publishers, Boston/London/Dordrecht, 2002.

[12] R.R. Yager. On ordered weighted averaging aggregation operators in multicriteria decision making. 18(1):183–190, 1988.

[13] Hans-Georg Beyer and Hans-Paul Schwefel. Evolution strategies: A comprehensive introduction. *Journal Natural Computing*, 1(1):3–52, 2002.

[14] M. de Berg, M. van Kreveld, M. Overmars, and O. Schwarzkopf. *Computational Geometry*. Springer, New York, 2000.

[15] N. Weskamp, E. Hüllermeier, D. Kuhn, and G. Klebe. Multiple graph alignment for the structural analysis of protein active sites. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 4(2):310–320, 2007.

[16] Thomas Bartz-Beielstein. *Experimental research in evolutionary computation: The new experimentalism*. Springer, 2006.

[17] Wolfgang Kabsch. A solution of the best rotation to relate two sets of vectors. *Acta Crystallographica*, 32:922–923, 1976.

ISBN: 978-989-95079-6-8

IFSA-EUSFLAT 2009

An Interval Approach for Fuzzy Linear Regression with Imprecise Data

Amory Bisserier[1] Reda Boukezzoula[1] Sylvie Galichet[1]

1 Laboratoire d'Informatique, Systèmes, Traitement de l'Information et de la Connaissance (LISTIC)
Université de Savoie BP 80439 74941 Annecy le Vieux, Cedex France
Email: {amory.bisserier, reda.boukezzoula, sylvie.galichet}@univ-savoie.fr

Abstract *In this paper, a revisited approach for fuzzy regression linear model representation and identification is introduced. By adopting the commonly used principle of α-cuts, the fuzzy regression implementation is reduced to the handling of conventional intervals, for inputs, parameters and outputs. Using the Midpoint-Radius representation of intervals, the uncertainty attached to linear models becomes more interpretable. Actually, it is possible to determine the output uncertainty origin (model parameters and/or inputs). In this context, a possibilistic regression method is proposed to identify models of minimal global uncertainty, that is with respect to all possible inputs.*

Keywords Interval Regression, Fuzzy Regression, Uncertainty Representation, Fuzzy Inputs-Fuzzy Outputs.

1 Introduction

Fuzzy regression, a fuzzy type of conventional regression analysis, has been proposed to evaluate the functional relationship between input and output variables in a fuzzy environment. Indeed, unlike statistical regression modeling based on probability theory, fuzzy regression is based on possibility theory and fuzzy set theory [14].

In the fuzzy literature, the regression problem with fuzzy data has been previously treated from different points of view. According to [5], fuzzy regression techniques can be classified into two distinct areas. The first proposed by Tanaka ([10], [12]) which minimizes the total spread of the output is named possibilistic regression. In this case, the problem is viewed as finding fuzzy coefficients of a regression model according to a mathematical programming problem. The second approach developed by Diamond [4], which minimizes the total square error of the output is called the fuzzy least square method. In this paper a possibilistic approach is adopted.

In the chosen possibilistic context, different kinds of input/output data can be considered. Actually, the complete specification of regression problems highly depends on the nature of input-output data [5]. Some researches are thus devoted to Crisp-Input Crisp-Output (CICO) data [9] while others [6], [7], [8], [13] consider the regression problem using Fuzzy-Inputs and Fuzzy-Outputs (FIFO) data. Most commonly, a mixed approach with Crisp-Inputs and Fuzzy-Outputs (CIFO) is used [1], [2], [10], [11], [12].

In this framework, the fuzzy regression with CICO and CIFO data aroused a major interest. However, the FIFO model regression remains a little studied field. Indeed, only a few papers concerning this problem have been published ([6], [7], [8], [13]). For instance, in [6] and [13], different distances are proposed in order to extend the least square method to FIFO data. In [8], a multiobjective optimization technique for FIFO systems is proposed, whereas in [7] the optimization is based on the minimization of the deviation between observations and predictions.

All above mentioned methods lead to interesting results but present a lack of representativity and an awkward illustration of the uncertainties in the regression model. In order to overcome this problem, an extension of the regression methodology initially developed for CIFO systems in [1], [2] to FIFO ones is proposed. The key idea of the proposed approach resides in using a Midpoint-Radius regression model representation. By doing so, the regression problem can be viewed as a conventional linear optimization problem where a new criterion is proposed.

From a practical point of view and as commonly used in fuzzy regression literature ([5], [7], [8], [12]), the a-cut principle is adopted. In this case, for a specified a-cut, the fuzzy interval becomes a conventional interval, which states that a fuzzy interval representation is a generalization of a conventional one. In order to take into account the maximum of uncertainty in the system (the pessimistic case), an a-cut equal to 0 (0-cut) is considered. In this case, the fuzzy regression problem can be reduced to an interval regression one. That is the approach adopted in this paper where a new regression methodology is proposed.

The structure of this paper is as follows. In section 2, relevant concepts and notations are introduced. Section 3 is devoted to fuzzy regressive models representation. An identification method of such models based on a linear problem optimization is proposed in section 4. An application is shown in section 5. Finally, conclusion and perspectives are presented in section 6.

2 Relevant Concepts and Notations

For the sake of rigor and clarity, let us define the basic notions and notations used in this paper.
A fuzzy trapezoidal interval A is a particular case of a fuzzy number where a possibility distribution, represented by a membership function μ_A, is associated with the conventional intervals of the support and the kernel (see Figure 1).

ISBN: 978-989-95079-6-8

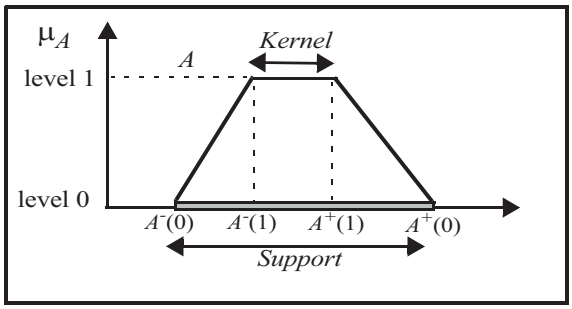


Figure 1: Trapezoidal fuzzy interval representation

In this case, for a given α-cut (α∈[0, 1]) on the fuzzy interval A, a conventional interval is obtained:

$$A(\alpha) = [A^{-}(\alpha), A^{+}(\alpha)] \tag{1}$$

Using the α-cut representation, a symmetrical fuzzy interval is viewed as a weighted family of nested intervals. In the sequel, in order to consider the maximum uncertainty, only α-cuts at level 0 will be considered. Consequently, all computations are performed with conventional intervals.

In this framework, an interval A is defined by the set of elements lying between its lower and upper limits (endpoints) as:

$$A = [A^{-},\ A^{+}] = \{x \mid A^{-} \le x \le A^{+}, x \in \Re\} \tag{2}$$

The interval A is called thick if $A^{-} < A^{+}$ and thin (or point) if $A^{-} = A^{+}$.

In order to manipulate and exhibit the uncertainty with intervals, an efficient representation is needed. Indeed, the latter should be able to represent arbitrary intervals and their sets in a single and uniform graphical way. In this framework, it is obvious that to uniquely define an interval, two parameters are necessary, hence the space of intervals is two-dimensional. There are many possible representations of an interval. The most used one is the endpoints representation given in equation (2). However, the main drawback of this representation resides in the fact that the uncertainty (the Radius) is awkward and not directly exhibited in the interval. In order to overcome this difficulty, another representation is adopted. Indeed, in order to facilitate the uncertainty representation, Midpoint (M) and Radius (R) coordinates are used (see Figure 2). In this case, an interval is represented as a point in the space (M, R).

For an interval A, its Midpoint and its Radius are defined by:

$$M_A = (A^{-} + A^{+})/2 \ \text{ and: } \ R_A = (A^{+} - A^{-})/2 \tag{3}$$

Moreover, the set of all the intervals contained in the interval A can be represented by a domain $C(A)$. As shown in Figure 2, the domain $C(A)$ is represented by the upper and lower bound lines of A, forming a triangle.

In the (M, R) representation, it is clear that the most imprecise element of $C(A)$ is the interval A itself which corresponds to the highest value of Radius, i.e. the top of the triangle. On the opposite, the triangle base represents all the crisp values between the lower and the upper bounds (the set of all the thin intervals).

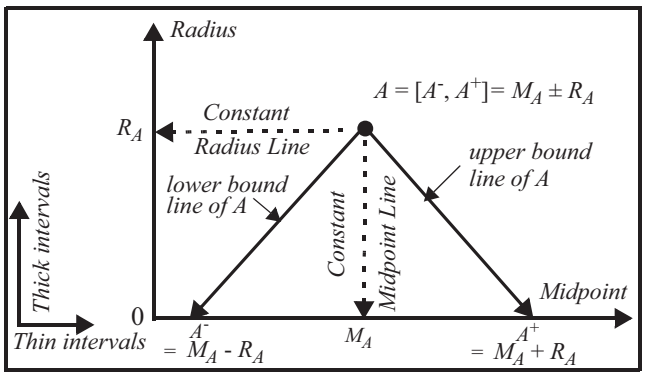


Figure 2: Midpoint-Radius representation of A and $C(A)$

For two intervals $A = [A^{-}, A^{+}]$ and $B = [B^{-}, B^{+}]$, the addition and multiplication operations are defined as follows:

$$\begin{cases} A \oplus B = [A^{-} + B^{-}, A^{+} + B^{+}] \\ A \otimes B = [min(Z), max(Z)] \end{cases} \tag{4}$$

where:

$$Z = \{A^{-}B^{-}, A^{-}B^{+}, A^{+}B^{-}, A^{+}B^{+}\} \tag{5}$$

In this case, it can be stated that the Midpoint and Radius of these operations are given by:

$$\begin{cases} M_{A \oplus B} = M_A \oplus M_B \\ R_{A \oplus B} = R_A \oplus R_B \end{cases} \tag{6}$$

and:

$$\begin{cases} M_{A \otimes B} = M_A M_B + \varphi_{A \otimes B} \cdot sign(M_A M_B) \\ R_{A \otimes B} = R_A R_B + |M_A| R_B + R_A |M_B| - \varphi_{A \otimes B} \end{cases} \tag{7}$$

where:

$$\varphi_{A \otimes B} = min\{|M_A| R_B, R_A |M_B|, R_A R_B\} \tag{8}$$

For two intervals [3] A and B, an inclusion relation of A in B (see Figure 3) is defined as follows:

$$A \subseteq B \Leftrightarrow \begin{cases} B^{-} \le A^{-} \\ A^{+} \le B^{+} \end{cases} \Leftrightarrow \begin{cases} M_B - R_B \le M_A - R_A \\ M_A + R_A \le M_B + R_B \end{cases} \tag{9}$$

$$\Leftrightarrow \begin{cases} M_B - M_A \le R_B - R_A \\ M_A - M_B \le R_B - R_A \end{cases} \tag{10}$$

From equation (10) it follows:

$$A \subseteq B \Leftrightarrow |M_B - M_A| \le R_B - R_A \tag{11}$$

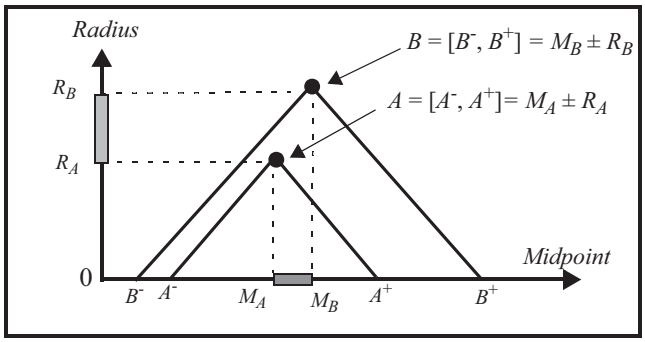


Figure 3: Inclusion of two intervals ($A \subseteq B$)

3 Fuzzy Regressive Model Representation

ISBN: 978-989-95079-6-8

The fuzzy linear regression model is the most frequently used form in regression analysis for expressing the relationship between one or more explanatory variables and response. For the sake of simplicity and as commonly assumed in the fuzzy regression literature, the case of simple linear regression model involving a single independent variable is considered. The case of multiple inputs is a straightforward generalization of this methodology but more complex in a computational point of view.

An interval regression model is mathematically expressed in the following form:

$$Y(X) = A_0 \oplus A_1 \otimes X \tag{12}$$

where Y is the interval output (dependant variable), A_0 and A_1 the interval parameters, i.e. unknowns to be estimated from observed data. According to (3), the parameters A_0 and A_1 are denoted as follows:

$$\begin{aligned} A_0 &= [A_0^-, A_0^+] = [M_{A_0} - R_{A_0}, M_{A_0} + R_{A_0}] \\ A_1 &= [A_1^-, A_1^+] = [M_{A_1} - R_{A_1}, M_{A_1} + R_{A_1}] \end{aligned} \tag{13}$$

The input X can take any value in a definition domain to be chosen.

As mentioned in the introduction, the regression problem has been previously treated from different points of view and by considering different kinds of input/output data. Indeed, according to the nature of X (crisp or imprecise) two regression problems can been distinguished. The first one deals with Crisp-Input Imprecise-Output data ([12], [10], [1], [2]). In this case, as the model is linear and the inputs are considered as crisp values, the model output (12) will also be an interval. So, for a crisp input $X = X_*$, equation (12) becomes:

$$Y(X_*) = [Y_*^-, Y_*^+] = A_0 \oplus A_1 \cdot X_* \tag{14}$$

In this case, the definition domain is the interval $D = [inf(D), sup(D)]$, which means that the input X can take any crisp value in D.

For the sake of illustration simplicity, in the sequel, the particular case when the parameter A_1 and the input are positive is considered. The other cases are a straightforward generalization of this particular case. The evolution of the model output (14) is illustrated in Figure 4. Actually, as the input X is a crisp value of D, the model output uncertainty is only due to the uncertainty of the two parameters A_0 and A_1.

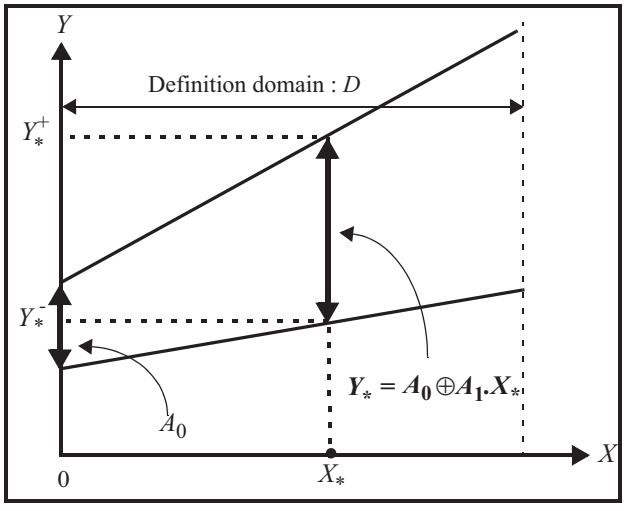


Figure 4: Interval model representation with a crisp input

The second regression problem is devoted to systems with Imprecise-Inputs and Imprecise-Outputs. For an interval input $X = X_* = [X_*^-, X_*^+]$, the interval model output is given by the following equation with Z defined according to (5):

$$\begin{aligned} Y(X_*) = [Y_*^-, Y_*^+] &= [A_0^- + min(Z), A_0^+ + max(Z)] \\ &= [A_0^- + A_1^- \cdot X_*^-, A_0^+ + A_1^+ \cdot X_*^+] \end{aligned} \tag{15}$$

In this case, the input X can be any interval included in the interval D previously defined for crisp inputs. Consequently, the model definition domain is now $\Delta = C(D)$ (see Figure 2). The model output evolution is presented in Figure 5.

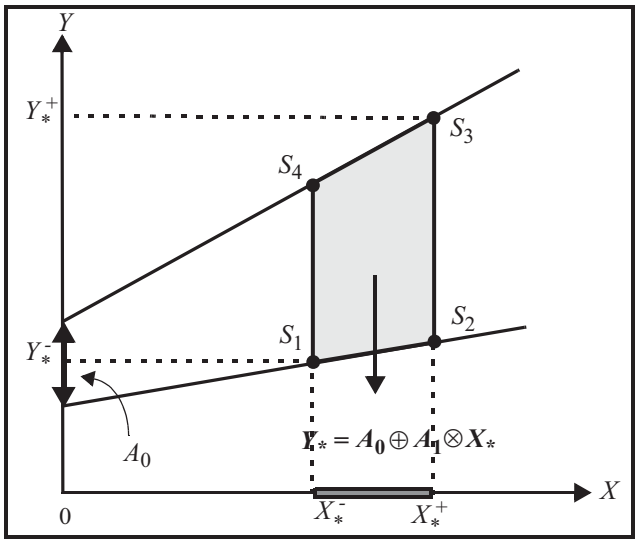


Figure 5: Conventional interval model representation

Several main drawbacks can be underlined:

- The interval input X_* is represented by its endpoints coordinates on a real line with only one dimension. It leads to a lack of information about the input uncertainty (Radius). In this case, neither the definition domain Δ nor the output uncertainty on it can be exhibited in a conventional 2D space.
- According to Figure 5, the model output is defined by the 4-sided area delimited by the four vertices (S_1, S_2, S_3, S_4). If the vertices S_1 and S_3 exhibit the model output bounds, the others (S_2 and S_4) have no interpretability in the model output. In this situation, two intervals are defined to represent the model output, which limits the interpretability of this kind of representation ([7], [8]).

In order to deal with these drawbacks, the Midpoint-Radius representation is used to define all handled intervals. In this context, the output Y is expressed by:

$$\begin{aligned} Y(X_*) &= [Y_*^-, Y_*^+] \\ &= [M_{Y(X_*)} - R_{Y(X_*)}, M_{Y(X_*)} + R_{Y(X_*)}] \end{aligned} \tag{16}$$

where:

$$\begin{cases} M_{Y(X_*)} = M_{A_0} + M_{A_1} \otimes X_* \\ R_{Y(X_*)} = R_{A_0} + R_{A_1} \otimes X_* \end{cases} \tag{17}$$

For positive parameter A_1 and input X_*, we have:

$$M_{A_1 \otimes X_*} = M_{A_1} M_{X_*} + \varphi_{A_1 \otimes X_*} \cdot sign(M_{A_1} M_{X_*}) \tag{18}$$

As:

ISBN: 978-989-95079-6-8

$$\varphi_{A_1 \otimes X_*} = min\{M_{A_1}R_{X_*}, R_{A_1}M_{X_*}, R_{A_1}R_{X_*}\} = R_{A_1}R_{X_*}, \quad (19)$$

the equation (18) becomes:

$$M_{A_1 \otimes X_*} = M_{A_1}M_{X_*} + R_{A_1}R_{X_*} \quad (20)$$

By adopting the same principle, the following Radius expression is determined:

$$R_{A_1 \otimes X_*} = R_{A_1}R_{X_*} + M_{A_1}R_{X_*} + R_{A_1}M_{X_*} - R_{A_1}R_{X_*} = M_{A_1}R_{X_*} + R_{A_1}M_{X_*} \quad (21)$$

By substitution of equations (20) and (21) in (17), it follows:

$$\begin{cases} M_{Y(X_*)} = M_{A_0} + M_{A_1}M_{X_*} + R_{A_1}R_{X_*} \\ R_{Y(X_*)} = R_{A_0} + M_{A_1}R_{X_*} + R_{A_1}M_{X_*} \end{cases} \quad (22)$$

The Midpoint and Radius output evolutions on the domain Δ are illustrated in Figure 6 and Figure 7.

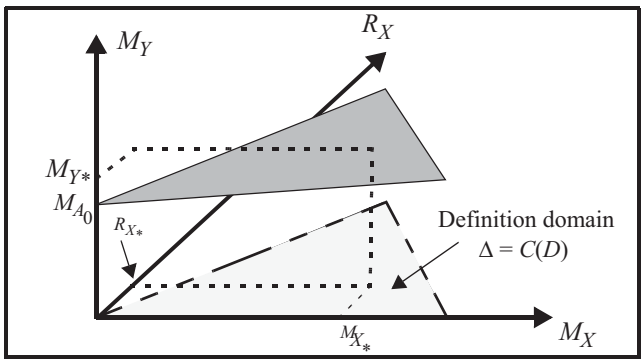


Figure 6: The midpoint output evolution

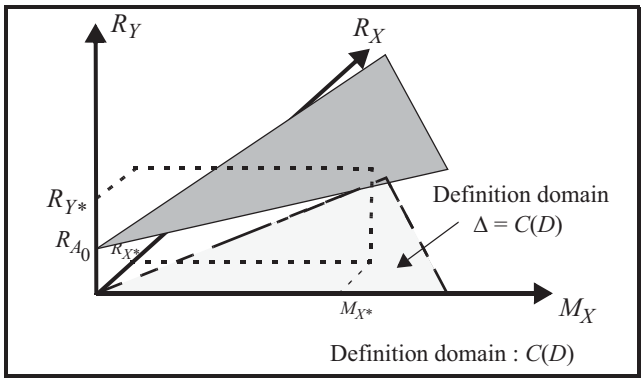


Figure 7: The radius output evolution.

From equations (16) and (22), it is obvious that:

$$\begin{cases} Y_*^- = M_{A_0} - R_{A_0} + (M_{A_1} - R_{A_1})(M_{X_*} - R_{X_*}) \\ Y_*^+ = M_{A_0} + R_{A_0} + (M_{A_1} + R_{A_1})(M_{X_*} + R_{X_*}) \end{cases} \quad (23)$$

A global representation of the model output evolution on the domain Δ is illustrated in Figure 8.

This Midpoint-Radius representation presents several advantages:

- In this representation, only one interval output is defined for a considered input. Indeed, the output expression represents as well the contribution of the input Midpoint and Radius.
- According to this representation, the origin of uncertainty in the output can be precisely determined. In other words, the contributions of the parameters A_0, A_1 and the input X are well defined in the model output (22).

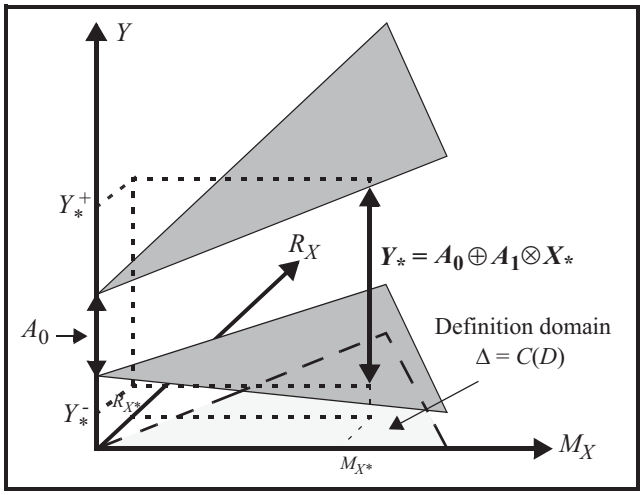


Figure 8: The model output evolution

4 Interval Regressive Model Identification

In the context of regression model identification, two main problems come up, namely, the model structure specification and the estimation of the given model. The first problem is focused on the choice of a suitable model structure for a data set. This problem is traditionally addressed *a priori.* In this paper, as commonly used, the model structure is assumed to be linear. In this case, the regression problem is reduced to an estimation problem of the model parameters.

Let us consider a set of N observed data samples. Let the j^{th} sample be represented by the couple $(X_j, Y_j), j = 1,..., N$ where X_j are interval inputs and the Y_j are the corresponding interval output.

All the observed input intervals are assumed to be included in the interval $D = [D^-, D^+]$ given by:

$$D^- = min(X_j^-) \text{ and } D^+ = max(X_j^+) \quad (24)$$

Like any regression technique, the fuzzy regression objective is to determine a predicted functional relationship between inputs X and outputs Y:

$$\hat{Y}(X) = A_0 \oplus A_1 \otimes X \quad (25)$$

In this case, the definition domain of the identified model is the set of all the intervals included in D, i.e. $\Delta = C(D)$.

As discussed in the paper introduction, a possibilistic regression approach is adopted where the objective is to determine the model parameters such that the observed data are included in the predicted ones (possibility model). In this framework, another kind of model (necessity one) could be identified [11].

In the proposed identification methodology, two points have to be considered for its implementation:

- the interval inclusion constraints to be introduced in the optimization problem,
- the identification criterion to be optimized.

4.1 The optimization constraints

In this case, the inclusion constraints are given by:

ISBN: 978-989-95079-6-8

$$Y_j \subseteq \hat{Y}_j \Leftrightarrow \left|M_{\hat{Y}_j} - M_{Y_j}\right| \leq R_{\hat{Y}_j} - R_{Y_j} \quad (26)$$

with:

$$M_{\hat{Y}_j} = M_{A_0} + M_{A_1} \otimes X_j \text{ and: } R_{\hat{Y}_j} = R_{A_0} + R_{A_1} \otimes X_j \quad (27)$$

By substitution of (27) in (26), the constraints inclusion can be written as:

$$\begin{aligned} &\left|M_{A_0} + M_{A_1} M_{X_j} + R_{A_1} R_{X_j} - M_{Y_j}\right| \leq \\ &R_{A_0} + M_{A_1} R_{X_j} + R_{A_1} M_{X_j} - R_{Y_j} \end{aligned} \quad (28)$$

Moreover, the optimization constraints must ensure the identification of intervals, whose bounds are ordered. Moreover, the sign of the parameter A_1 must be constrained as well. In the case of a model whose parameter A_1 is identified as positive, it follows:

$$R_{A_0} \geq 0, \text{ and } : M_{A_1} \geq R_{A_1} \geq 0 \quad (29)$$

It is important to note that all the considered constraints in the identification problem are linear according to the identification variables (Midpoint and Radius of A_0 and A_1).

4.2 The used criterion

The objective is to identify the less imprecise model on its definition domain, as it has been proposed in our previous works [1]. The computation of the global uncertainty on the whole domain D proposed in [1] is here extended to the whole domain Δ, whose elements are intervals. It follows that the criterion J to be optimized, i.e. the global uncertainty on Δ, is expressed as::

$$\begin{aligned} J &= \iint_{\Delta} (\hat{Y}^{+} - \hat{Y}^{-}) dM_X dR_X = 2 \iint_{\Delta} R_{\hat{Y}} dM_X dR_X \\ &= 2 \iint_{\Delta} R_{A_0} + M_{A_1} R_X + R_{A_1} M_X) dM_X dR_X \end{aligned} \quad (30)$$

The substitution of equation (23) in (30) yields to:

$$J = 2R_{A_0} \cdot R_D^2 + 2R_{A_1} \cdot R_D^2 \cdot M_D + \frac{2}{3} M_{A_1} \cdot R_D^3 \quad (31)$$

To sum up, the identification method is performed by minimizing the criterion (31) under the constraints (28) and (29).

At this stage, several points can be discussed about the proposed approach:

- As the interval D is defined with the observed inputs, its Radius and Midpoint are known numerical values. So, the criterion J is linear according to the optimization variables. Moreover, J is independent of the observed inputs, i.e. the optimization is made on the whole domain Δ of the model, not only on the observed data. This point contributes to improve the model robustness.
- It is obvious that the criterion must be optimized under constraints in order to ensure inclusion property. If the constraints are released, the identified model corresponds to the minimal possible uncertainty, achieved with null parameters, and has no sense.
- As the minimization of the model uncertainty is performed on Δ, the input uncertainty is considered in the model output expression in the identification process.

5 Illustrative example

In this part, the proposed identification method is applied on the illustrative example given in [8], [7]. Given the data set [8], [7] illustrated in Table 1, corresponding of α-cuts of level 0 of proposed fuzzy data set, the objective is to identify the model parameters A_0 and A_1.

Table 1: Observed data set and predicted one

inputs X_j	observed outputs Y_j	predicted outputs $\hat{Y}_j$
[1.5 ; 2.5]	[3.5 ; 4.5]	[3.5 ; 6.5]
[3 ; 4]	[5 ; 6]	[4.16 ; 7.25]
[4.5 ; 6.5]	[6.5 ; 8.5]	[4.82 ; 8.5]
[6.5 ; 7.5]	[6 ; 7]	[5.69 ; 9]
[8 ; 9]	[8 ; 9]	[6.35 ; 9.75]
[9.5 ; 11.5]	[7 ; 9]	[7 ; 11]
[10.5 ; 11.5]	[10 ; 11]	[7.45 ; 11]
[12 ; 13]	[9 ; 10]	[8.1 ; 11.75]

The considered definition domain is D = [1.5, 13] and represented in Figure 9. This domain corresponds to a set of positive inputs, and, as discussed previously, the parameter A_1 is assumed to be positive too.

The identified parameters corresponding to the less imprecise model according to the criterion J defined in (31) are:

$$A_0 = [2.844 ; 5.25] \text{ and } A_1 = [0.438 ; 0.5] \quad (32)$$

In this case, the optimal value of uncertainty obtained is $J = 153.95$.

The identified model output representation is given in Figure 9, whereas the predicted outputs are presented in Table 1.

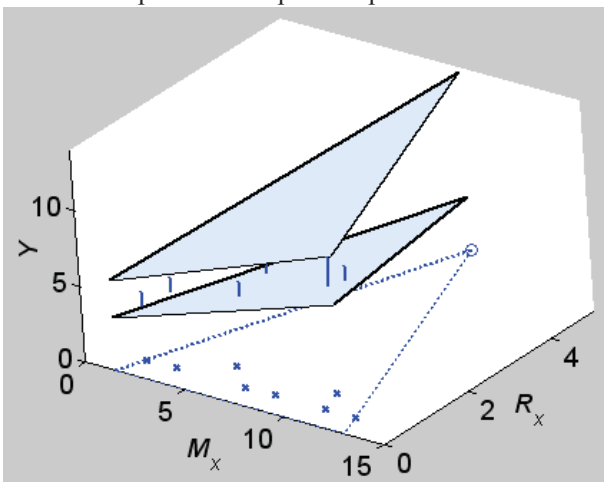


Figure 9: Identified model representation.

As presented in Table 1, the main advantage of the proposed approach compared to those presented in [7] and [8] is that only one interval output corresponds to a considered input. In Figure 9, it can be stated that the desired inclusion is respected.

Another advantage of the representation using the (Midpoint, Radius) space for the interval inputs concerns the exploitation

ISBN: 978-989-95079-6-8

of the model characteristics. According to (22), the output Radius evolution on Δ can be represented (see Figure 10). On the axe (M_X, $R_X = 0$), i.e. the triangle basis, only the parameter uncertainty is exhibited, whereas at the top of the triangle, the output uncertainty is maximum, corresponding to the most uncertain input in Δ, i.e. D. In this example (Figure 10), the output radius is nearly constant on the basis of the triangle, corresponding to the case of a nearly crisp parameter A_1. The output radius increasing on Δ is essentially due to the input uncertainty.

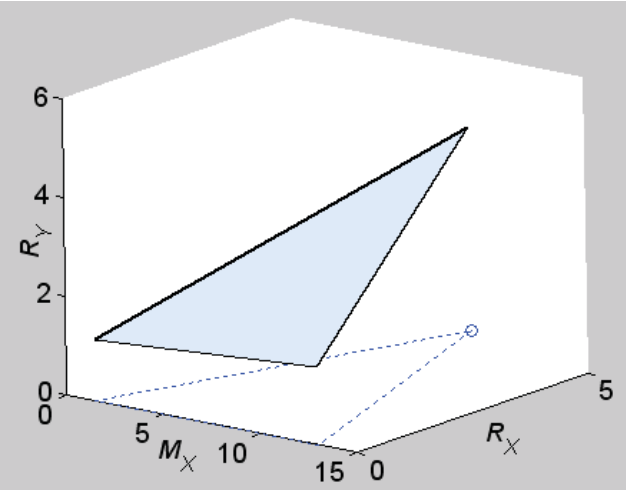


Figure 10: Identified model output Radius evolution.

The same analysis can be performed on the output Midpoint evolution on Δ, corresponding to the global tendency of the output on the domain.

In [8], the identified model, whose output respects the inclusion statement, corresponds to a criterion value $J = 281.03$, according to (31). Its parameters are the following:

$$A_0 = [1.25 ; 6.75] \text{and } A_1 = [0.4 ; 0.56] \quad (33)$$

In fact, the vertical dimension α is introduced into the proposed multiobjective optimization problem. It is known that the higher the α value is set for optimization, the more uncertain the identified model is and inclusion is respected for any lower α levels. So, this compromise between higher level of inclusion and lower uncertainty leads to the identification of a model whose uncertainty is not minimal on the domain.

On the other hand, the proposed method in [7], which minimizes deviations between observations and predictions without constraining inclusion, leads to the identification of a model which corresponds to a criterion value $J = 102.6$, and whose parameters are:

$$A_0 = [3 ; 3.82] \text{and } A_1 = [0.5 ; 0.54] \quad (34)$$

It is obvious that guaranteeing the inclusion increase the model uncertainty.

6 Conclusion

In this paper, a revisited representation of interval regressive model is proposed and used in an identification problem.

By adopting the Midpoint-Radius formalism for intervals, uncertainty representation is improved. Applied to regressive interval models, this formalism allows a better handling of the input uncertainty in the definition domain establishment and in the model output definition. In an identification problem, based on the optimization of a linear expression of the model output global uncertainty on its definition domain, the desired inclusion is ensured.

Further work will focus on the extension of this approach to fuzzy linear systems, by considering the vertical dimension α. In the proposed space, a symmetrical trapezoidal fuzzy interval can be interpreted by a zooming operation as illustrated in the following Figure 11:

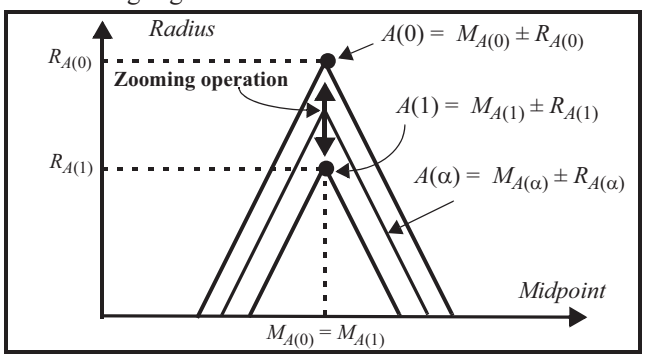


Figure 11: Midpoint-Radius representation of symmetrical trapezoidal fuzzy interval α-cuts.

A particular attention should be paid to the loss of linearity of the membership function of the model output, due to the fuzzy product between the input X and the parameter A_1.

References

[1] A. Bisserier, S. Galichet, R. Boukezzoula, Fuzzy Piecewise Linear Regression, *5th IEEE World Congress on Computational Intelligence (WCCI 2008 / FUZZ-IEEE 2008)*, Hong-Kong, China, 2008, 2089-2094.

[2] A. Bisserier, R. Boukezzoula, S. Galichet, Linear fuzzy regression using trapezoidal fuzzy intervals, *Information Processing and Management of Uncertainty (IPMU2008)*, Malaga, Spain, 2008, 181-188

[3] R.Boukezzoula, S. Galichet and L. Foulloy, MIN and MAX Operators for Fuzzy Intervals and their Potential Use in Aggregation Operators, *IEEE Trans. on Fuzzy Systems* 15(6): 1135-1144, 2007

[4] P. Diamond, Fuzzy Least Squares, *Information Sciences* 46: 141-157, 1988.

[5] P. Diamond , H. Tanaka, Fuzzy regression analysis, *Fuzzy sets in decision analysis, operations research and statistics*, Kluwer Academic Publishers, Norwell, MA, 1999.

[6] P. D'Urso, Linear regression analysis for fuzzy/crisp input and fuzzy/crisp output data, *Computational Statistics & Data Analysis* 42 (1-2), 2003, 47-72.

[7] M. Hojati, C.R. Bector, K. Smimou, A simple method for computation of fuzzy linear regression, *European Journal of Operational Research* 166: 172-184, 2005.

[8] M. Sakawa, H. Yano, Multiobjective fuzzy linear regression analysis for fuzzy input-output data, *Fuzzy Sets and Systems* 47: 173-181, 1992.

[9] D.A. Savic, W. Pedrycz, Evaluation of fuzzy linear regression models, *Fuzzy Sets and Systems* 39, 1991, 51-63.

[10] H.Tanaka, I. Hayashi, J. Watada, Possibilistic linear regression analysis for fuzzy data, *European Journal of Operational Research* 40: 389-396, 1989.

[11] H. Tanaka, H. Lee, Interval regression analysis by quadratic programming approach, *IEEE Trans. on Fuzzy Systems* 6(4): 473-481, 1998.

[12] H.Tanaka, S. Uejima, K. Asai, Linear regression analysis with fuzzy model, *IEEE, Systems, Trans. Systems Man Cybernet*. SMC-2: 903-907, 1982.

[13] H.-C. Wu, Fuzzy estimates of regression parameters in linear regression models for imprecise input and output data, *Computational Statistics & Data Analysis* 42 (1-2), 2003, 203-217.

[14] L.A. Zadeh, Fuzzy Sets, *Information and Control* 8: 338-353, 1965.

ISBN: 978-989-95079-6-8

A Predator-Prey Model with Fuzzy Initial Populations

Muhammad Zaini Ahmad Bernard De Baets

Department of Applied Mathematics, Biometrics and Process Control, Ghent University
Coupure links 653, B-9000 Ghent, Belgium
Email: mzaini_ahmad@yahoo.com, Bernard.DeBaets@ugent.be

Abstract— *The aim of this paper is to study a predator-prey population model which takes into account the uncertainty that arises when determining the initial populations of predator and prey . This model is solved numerically by means of a 4^{th}-order Runge-Kutta method. Simulations are made and a graphical representation is also provided to show the evolution of both populations over time. In addition to that, a new phase-plane in this fuzzy setting, referred to as fuzzy phase-plane, is introduced. The stability of the equilibrium points is also described. Finally, this paper points out a new research direction on fuzzy dynamical systems, especially in non-linear cases.*

Keywords— Equilibrium points, Fuzzy predator-prey population model, Fuzzy phase-plane, Fuzzy stability, Uncertainty.

1 Introduction

Nowadays, differential equations have proved a successful modelling paradigm. In a wide range of disciplines, the behaviour of an idealised version of a problem under study has been adequately described by one or more ordinary differential equations. Real-world problems, however, are pervaded with uncertainty. According to Diniz *et al.* [1], the uncertainty can arise in the experimental part, the data collection, the measurement process, as well as when determining the initial conditions. These are patently obvious when dealing with "living" material, such as soil, water, microbial populations, etc. In order to handle these problems, the use of fuzzy sets may be seen as an effective tool for a better understanding of the studied phenomena. It is therefore not surprising that there is a vast literature dealing with fuzzy differential equations (FDEs) for modelling this kind of perception. In the study of differential equations in a fuzzy environment, the term fuzzy differential equations is used for referring to differential equations with fuzzy coefficients, differential equations with fuzzy initial values or fuzzy boundary values, or even differential equations dealing with functions on the space of fuzzy intervals (see [2, 3, 4, 5, 6, 7, 8]).

In population dynamics, Bassanezi *et al.* [9] have started using FDEs to study the stability of fuzzy dynamical systems. The variables and/or the parameters in the systems under consideration were supposed to be uncertain and modelled by fuzzy sets. The authors have suggested a new definition of fixed point: if f is a continuous function, then u is a fixed point if $f([u]^\alpha) = [u]^\alpha$ for $\alpha \in [0,1]$, where $[u]^\alpha = [u_1^\alpha, u_2^\alpha] = [\min_{x\in[u_1^\alpha,u_2^\alpha]} f(x), \max_{x\in[u_1^\alpha,u_2^\alpha]} f(x)]$ (see [9] for more details). According to Ortega *et al.* [10], the proposed idea is not easy to implement in practice because several details should be treated carefully. Guo *et al.* [11] have adopted Hüllermeier's approach [7] to establish fuzzy impulsive functional differential equations and some results were applied in the logistic model and the Gompertz model. In [12], the author has established existence and uniqueness results for fuzzy functional differential equations. These results were applied in population models whose solutions define fuzzy intervals. Recently, Peixoto *et al.* [13] have studied the fuzzy predator-prey population model. The authors have elaborated the classical deterministic model by means of a fuzzy rule-based system. The fuzzy rule-based system considered in their work consists of two inputs and two outputs. The fuzzy rule base is given by 30 rules of the type: "If the number of preys is large AND the potential of predation is very small, THEN the variation of preys increase a little AND the variation of the potential of predation increase a lot". The words such as "large", "very small", "increase a little" and "increase a lot" are modelled by fuzzy sets. This might be a possible way to treat the model in a fuzzy setting. In this paper, we propose a new way of studying the fuzzy predator-prey population model. For this purpose, we only consider the initial populations of predator and prey to be fuzzy intervals.

2 Preliminary Concepts

In this section, let us first introduce the notations and some theoretical background we will be using throughout this paper. Denote $\mathcal{F}(\mathbb{R}) = \{U \mid U : \mathbb{R} \to [0,1]\}$, where U satisfies the following conditions:

1. U is normal: there exists $x_0 \in \mathbb{R}$ such that $U(x_0) = 1$;
2. U is convex: for $x, y \in \mathbb{R}$ and $0 \leq \lambda \leq 1$, one has
$$U(\lambda x + (1-\lambda)y) \geq \min\left(U(x), U(y)\right);$$
3. U is upper semi-continuous: for any $x_0 \in \mathbb{R}$, one has
$$U(x_0) \geq \lim_{x \to x_0^{\pm}} U(x);$$
4. $[U]^0 = \overline{\{x \in \mathbb{R} \mid U(x) > 0\}}$ is a compact subset of $\mathbb{R}$.

If U satisfies (1)–(4), then it is called a fuzzy interval. We can define its α-cuts as follows:

$$[U]^\alpha = \{x \in \mathbb{R} \mid U(x) \geq \alpha\}, \;\; 0 < \alpha \leq 1$$

For a fuzzy interval U, its α-cuts are closed intervals in $\mathbb{R}$. Let us denote them by $[U]^\alpha = [u_1^\alpha, u_2^\alpha]$.

In what follows, we introduce some results we adopt from [14] concerning the notion of interactivity of fuzzy intervals.

ISBN: 978-989-95079-6-8

Definition 1 *Two fuzzy intervals A and B are said to be interactive fuzzy intervals if there exist $q, r \in \mathbb{R}$, $q \neq 0$, such that their joint possibility distribution is given by*

$$\begin{aligned} C(x_1, x_2) &= A(x_1) \cdot \chi_{\{qx_1+r=x_2\}}(x_1, x_2) \\ &= B(x_2) \cdot \chi_{\{qx_1+r=x_2\}}(x_1, x_2) \end{aligned} \tag{1}$$

where $\chi_{\{qx_1+r=x_2\}}(x_1, x_2)$ stands for the characteristic function of the line

$$\left\{(x_1, x_2) \in \mathbb{R}^2 \mid qx_1 + r = x_2\right\}.$$

In this case we have

$$[C]^\alpha = \left\{(x, qx + r) \in \mathbb{R}^2 \mid x = (1-t)a_1^\alpha + ta_2^\alpha, t \in [0,1]\right\}$$

where $[A]^\alpha = [a_1^\alpha, a_2^\alpha]$ and $[B]^\alpha = q[A]^\alpha + r$ for any $\alpha \in [0,1]$, and, finally

$$B(x) = A\left(\frac{x-r}{q}\right)$$

for all $x \in \mathbb{R}$.

Definition 2 *Two fuzzy intervals A and B are said to be completely positively interactive if q is positive in (1).*

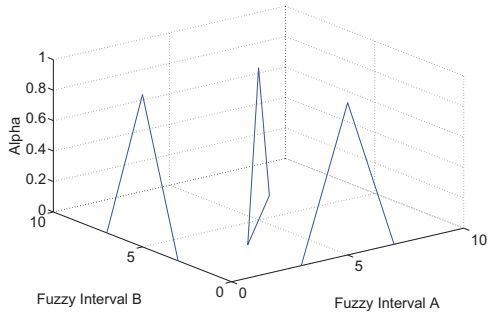


Figure 1: Completely positively interactive fuzzy intervals.

Definition 3 *Let C be a joint possibility distribution with marginal possibility distributions A and B, and let $f : \mathbb{R}^2 \to \mathbb{R}$ be a continuous function. If A and B are interactive fuzzy intervals, then the extension of f via C is defined as*

$$f(A, B)(y) = \sup_{y=f(x_1,x_2)} C(x_1, x_2), \quad \forall y \in \mathbb{R}. \tag{2}$$

Remark 1 *If A and B are non-interactive fuzzy intervals, that is, their joint possibility distribution is defined by*

$$C(x_1, x_2) = \min\left(A(x_1), B(x_2)\right),$$

then (2) turns into the extension principle introduced by Zadeh [15].

Definition 4 *If A and B are two fuzzy intervals, then the distance D between A and B is defined as*

$$D(A, B) = \sup_{\alpha \in [0,1]} d_H([A]^\alpha, [B]^\alpha)$$

where

$$d_H([A]^\alpha, [B]^\alpha) =$$

$$\max\left\{\sup_{a \in [A]^\alpha} \inf_{b \in [B]^\alpha} d(a,b), \sup_{b \in [B]^\alpha} \inf_{a \in [A]^\alpha} d(a,b)\right\}$$

3 Two-Dimensional Autonomous Systems

In this section, we first consider a general two-dimensional autonomous system with crisp initial values:

$$\begin{cases} x'(t) = f(x,y), & x(t_0) = x_0 \\ y'(t) = g(x,y), & y(t_0) = y_0 \end{cases} \tag{3}$$

where $f, g : \mathbb{R}^2 \to \mathbb{R}$ are continuous functions and $x_0, y_0 \in \mathbb{R}$. This model is called an autonomous system because the functions appearing on the right-hand side of (3) do not depend on the independent variable t. Suppose that the initial values in (3) are uncertain and modelled by fuzzy intervals, then we have the following fuzzy autonomous system [16]:

$$\begin{cases} x'(t) = f(x,y), & x(t_0) = X_0 \\ y'(t) = g(x,y), & y(t_0) = Y_0 \end{cases} \tag{4}$$

where $f, g : \mathbb{R}^2 \to \mathbb{R}$ are continuous functions and $X_0, Y_0 \in \mathcal{F}(\mathbb{R})$ with $[X_0]^\alpha = [x_{01}^\alpha, x_{02}^\alpha]$ and $[Y_0]^\alpha = [qx_{01}^\alpha + r, qx_{02}^\alpha + r]$ for $\alpha \in [0,1]$, $q \in \mathbb{R}^+$ and $r \in \mathbb{R}$. Here we assume that X_0 and Y_0 are completely positively interactive fuzzy intervals.

A solution to (4) can be approximated by any numerical method existing in the literature. For any fixed value of t, the solution is represented as a joint possibility distribution in the xy-plane. As t varies, the solution traces out the joint possibility distributions in the plane. By plotting a collection of such joint possibility distributions, we gain graphical insight into the behaviour of solutions. In this paper, the set of joint possibility distributions and the xy-plane are called fuzzy trajectories and the fuzzy phase-plane, respectively.

4 Numerical Methods

In many cases, model (4) presented in Section 3 cannot be solved analytically and therefore we need a numerical scheme to approximate the exact solution. We consider the following 4^{th}-order Runge-Kutta method:

$$x_{i+1} = x_i + \frac{h}{6}\left[K_1 + 2K_2 + 2K_3 + K_4\right] \tag{5}$$

$$y_{i+1} = y_i + \frac{h}{6}\left[L_1 + 2L_2 + 2L_3 + L_4\right] \tag{6}$$

where

$$\begin{aligned} K_1 &= f\left(x_i, y_i\right) \\ L_1 &= g\left(x_i, y_i\right) \\ K_2 &= f\left(x_i + \frac{h}{2}K_1, y_i + \frac{h}{2}L_1\right) \\ L_2 &= g\left(x_i + \frac{h}{2}K_1, y_i + \frac{h}{2}L_1\right) \\ K_3 &= f\left(x_i + \frac{h}{2}K_2, y_i + \frac{h}{2}L_2\right) \\ L_3 &= g\left(x_i + \frac{h}{2}K_2, y_i + \frac{h}{2}L_2\right) \\ K_4 &= f\left(x_i + hK_3, y_i + hL_3\right) \\ L_4 &= g\left(x_i + hK_3, y_i + hL_3\right) \end{aligned}$$

ISBN: 978-989-95079-6-8

Since the arguments x_i and y_i on the right-hand side of (5) and (6) are interactive, we can define new functions as follows:

$$F_h(x_i, y_i) = x_i + \frac{h}{6}\left[K_1 + 2K_2 + 2K_3 + K_4\right] \quad (7)$$

$$G_h(x_i, y_i) = y_i + \frac{h}{6}\left[L_1 + 2L_2 + 2L_3 + L_4\right] \quad (8)$$

By using the concept of interactivity of fuzzy intervals proposed by Majlender [14], we have the following generalised 4^{th}-order Runge-Kutta method for solving (4), where X_0 and Y_0 are interactive fuzzy intervals.

$$X_{i+1} = F_h(X_i, Y_i) \quad (9)$$

$$Y_{i+1} = G_h(X_i, Y_i). \quad (10)$$

The idea is that if $X_i(u) \geq \alpha$ for some $u \in \mathbb{R}$, then there exists a unique $v \in \mathbb{R}$ that Y_i can take. Furthermore, if u is moved to the left (right), then the corresponding value (that Y_i can take) will also move to the left (right). Vice versa, if u is moved to the left (right), then the corresponding value (that Y_i can take) will move to the right (left) [14]. From this explanation, we can compute the numerical approximations as follows:

$$x^{\alpha}_{1,i+1} = F_h(x^{\alpha}_{1,i}, y^{\alpha}_{1,i}) \quad (11)$$

$$x^{\alpha}_{2,i+1} = F_h(x^{\alpha}_{2,i}, y^{\alpha}_{2,i}) \quad (12)$$

$$y^{\alpha}_{1,i+1} = G_h(x^{\alpha}_{1,i}, y^{\alpha}_{1,i}) \quad (13)$$

$$y^{\alpha}_{2,i+1} = G_h(x^{\alpha}_{2,i}, y^{\alpha}_{2,i}) \quad (14)$$

Our purpose here is to generate accurate approximations at each α-cut. We begin by making a partition of the form $t_0 < t_1 < t_2 < ... < t_{N-1} < t_N = T$ on the interval $[t_0, T]$. This partition is uniformly spaced, that is the partition points are $t_i = t_0 + ih$, $i = 0, 1, 2, ..., N$ and the partition spacing $h = \frac{T - t_0}{N} > 0$ is sufficiently small and we called it the step-size or step-length.

5 Stability

In this section, we consider the following fuzzy autonomous system in vector form [16]:

$$\mathbf{z}'(t) = \mathbf{f}(\mathbf{z}), \quad \mathbf{z}(t_0) = \mathbf{Z}_0 \quad (15)$$

where $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ and $\mathbf{Z}_0 \in \mathcal{F}(\mathbb{R}^2)$ is a joint possibility distribution of which the marginal possibility distributions are interactive fuzzy intervals.

Next, we give several definitions and sufficient conditions for stability of the model studied.

Definition 5 *A point $\mathbf{z}_e$ is called an equilibrium point of (15) at time t_0, if for all $t \geq t_0$, it holds that $\mathbf{f}(\mathbf{z}_e) = 0$.*

This definition asserts that an equilibrium point $\mathbf{z}_e$ is a constant solution of (15), since $\mathbf{f}(\mathbf{z}_e) = 0$ and thus $\mathbf{z}'(t) = 0$ at such a point.

Lemma 1 *If $\mathbf{Z}$ is joint possibility distribution and $\mathbf{z}_e$ is an equilibrium point, then*

$$D(\mathbf{Z}, \mathbf{z}_e) = \sup_{\mathbf{s} \in [\mathbf{Z}]^0} d(\mathbf{s}, \mathbf{z}_e)$$

Definition 6 *An equilibrium point $\mathbf{z}_e$ of (15) is defined to be fuzzy stable if for all $\epsilon > 0$, there exists $\delta > 0$ such that if $D(\mathbf{Z}(t_0), \mathbf{z}_e) < \delta$ then $D(\mathbf{Z}(t), \mathbf{z}_e) < \epsilon$ for all $t \geq t_0$.*

This definition states that all solutions of (15) that start "sufficiently close" to an equilibrium point $\mathbf{z}_e$ stay "close" to $\mathbf{z}_e$. Note that this definition of fuzzy stability does not require the solution to approach the equilibrium point.

Proposition 1 *If an equilibrium point $\mathbf{z}_e$ is stable, then it is also fuzzy stable.*

Definition 7 *An equilibrium point $\mathbf{z}_e$ of (15) is defined to be asymptotically fuzzy stable if it is fuzzy stable and there exists $\delta > 0$ such that if $D(\mathbf{Z}(t_0), \mathbf{z}_e) < \delta$ then $\lim_{t \to \infty} D(\mathbf{Z}(t), \mathbf{z}_e) = 0$ for all $t \geq t_0$.*

This definition asserts that solutions of (15) that start "sufficiently close" to an equilibrium point $\mathbf{z}_e$ not only stay "close" to $\mathbf{z}_e$, but must eventually approach the equilibrium point $\mathbf{z}_e$ as t approaches infinity. Note that this is a stronger requirement than fuzzy stability, since the equilibrium point $\mathbf{z}_e$ must be stable before one can talk about whether or not it is asymptotically fuzzy stable.

Proposition 2 *If an equilibrium point $\mathbf{z}_e$ is asymptotically stable, then it is also asymptotically fuzzy stable.*

Definition 8 *An equilibrium point $\mathbf{z}_e$ of (15) is defined to be fuzzy unstable if it is not fuzzy stable. Hence, there exists an $\epsilon > 0$ such that for any choice of $\delta > 0$, there exists $\mathbf{Z}(t)$ satisfying $D(\mathbf{Z}(t_0), \mathbf{z}_e) < \delta$ such that $D(\mathbf{Z}(t), \mathbf{z}_e) \geq \epsilon$ for some t.*

This definition states that all solutions of (15) that start "sufficiently close" to an equilibrium point $\mathbf{z}_e$ move away from $\mathbf{z}_e$ when t increases.

Proposition 3 *If an equilibrium point $\mathbf{z}_e$ is unstable, then it is also fuzzy unstable.*

6 Fuzzy Predator-Prey Population Model

Consider the following fuzzy predator-prey population model

$$\begin{cases} x'(t) = x - 0.1xy \\ y'(t) = -0.5y + 0.02xy \end{cases}$$

where the initial populations of prey and predator are interactive fuzzy intervals with $q = 1$ and $r = -5$, and they are defined as follows, respectively:

$$X_0 = \begin{cases} 0, & \text{if } x < 14, \\ x - 14, & \text{if } 14 \leq x < 15, \\ -x + 16, & \text{if } 15 \leq x < 16, \\ 0, & \text{if } x > 16 \end{cases}$$

and

$$Y_0 = \begin{cases} 0, & \text{if } y < 9, \\ y - 9, & \text{if } 9 \leq y < 10, \\ -y + 11, & \text{if } 10 \leq y < 11, \\ 0, & \text{if } y > 11 \end{cases}$$

ISBN: 978-989-95079-6-8

First we find the equilibrium points of the model by setting the derivatives equal to zero. One can note that the model has two equilibrium points: $(0, 0)$ and $(25, 10)$. The equilibrium point $(0, 0)$ is uninteresting because there are no populations to observe in the model. However, the second equilibrium point is of interest. From the numerical approximation, we can see that the point $(25, 10)$ is fuzzy stable since the supports of the joint possibility distributions move around the equilibrium point when t increases (see Fig. 4). This is true since the prey population increases when the predator population is minimum, and the prey population decreases when the predator population is maximal (see Figs. 2 and 3).

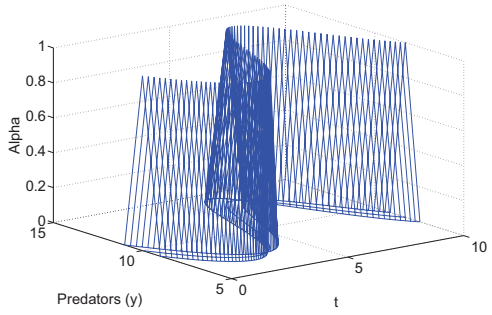


Figure 2: The evolution of the predator population over time.

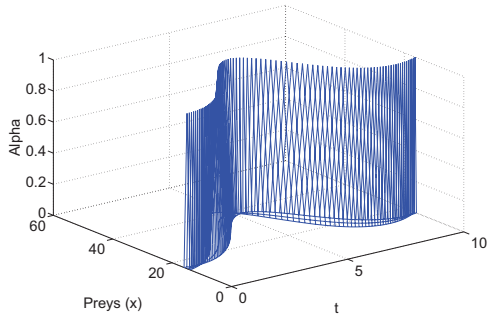


Figure 3: The evolution of the prey population over time.

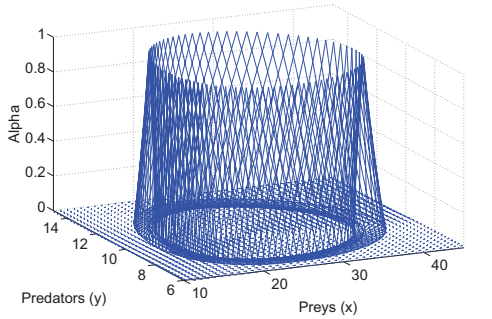


Figure 4: Fuzzy phase-plane with direction field ($N = 100$, $t = 10$, $h = 0.1$).

7 Conclusions

In this paper, we have proposed a new way of studying predator-prey population models, which takes into account the uncertainty that arises when determining the initial populations of predator and prey. By using the concept of interactivity of fuzzy intervals proposed by Majlender [14], we have developed a generalised numerical method, based on a 4^{th}-order Runge-Kutta method, to approximate the number of predators and preys over time.

Acknowledgment

This research was co-funded by the Ministry of Higher Education of Malaysia (MOHE) and partially supported by Universiti Malaysia Perlis (UniMAP) under the program "Skim Latihan Akademik IPTA (SLAI)".

References

[1] G.L. Diniz, J.F.R. Fernandes, J.F.C.A. Meyer, and L.C. Barros. A fuzzy Cauchy problem modelling the decay of the biochemical oxygen demand in water. *Proc. Joint 9th IFSA World Congress and 20th NAFIPS International Conferences*, 1:512–516, 25-28 July 2001.

[2] S. Abbasbandy and T. Allahviranloo. Numerical solution of fuzzy differential equation by Runge-Kutta method. *Nonlinear Studies*, 11(1):117–129, 2004.

[3] M. Ma, M. Friedman, and A. Kandel. Numerical solution of fuzzy differential equations. *Fuzzy Sets and Systems*, 105:133–138, 1999.

[4] B. Bede and S.G. Gal. Generalizations of the differentiability of fuzzy-number-valued functions with applications to fuzzy differential equations. *Fuzzy Sets and Systems*, 151:581–599, 2005.

[5] J.J. Buckley and T. Feuring. Fuzzy differential equations. *Fuzzy Sets and Systems*, 110:43–54, 2000.

[6] Y. Chalco-Cano and H. Román-Flores. On new solutions of fuzzy differential equations. *Chaos, Solitons and Fractals*, 38 (1):112–119, 2008.

[7] E. Hüllermeier. An approach to modelling and simulation of uncertain dynamical systems. *Int. J. Uncertainty Fuzziness Knowledge-Based Systems*, 5:117–137, 1997.

[8] M.T. Mizukoshi, L.C Barros, Y. Chalco-Cano, H. Román-Flores, and R.C. Bassanezi. Fuzzy differential equations and the extension principle. *Information Sciences*, 177:3627–3635, 2007.

[9] R. C Bassanezi, L.C. Barros, and P.A. Toneli. Attractor and asymptotic stability for dynamical systems. *Fuzzy Sets and Systems*, 113:473–483, 2000.

[10] N. R. S. Ortega, P. C. Sallum, and E. Massad. Fuzzy dynamical systems in epidemic modelling. *Kybernetes*, 29 (2):201–218, 2000.

[11] M. Guo, X. Xu, and R. Li. Impulsive functional differential inclusions and fuzzy population models. *Fuzzy Sets and Systems*, 138:601–615, 2003.

[12] V. Lupulescu. On a class of fuzzy functional differential equations. *Fuzzy Sets and Systems*, 160:1547–1562, 2009.

[13] M.S. Peixoto, L.C. Barros, and R.C. Bassanezi. Predator-prey fuzzy model. *Ecological Modelling*, 214:39–44, 2008.

[14] C. Carlsson, R. Fullér, and P. Majlender. Additions of completely correlated fuzzy numbers. *Proceeding of IEEE International Conference on Fuzzy Systems*, 1:353–539, 25-29 July 2004.

[15] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[16] S. Seikkala. On the fuzzy initial value problem. *Fuzzy Sets and Systems*, 24:319–330, 1987.

ISBN: 978-989-95079-6-8

Local semantic context analysis for automatic ontology matching

Giuseppe Fenza, Vincenzo Loia and Sabrina Senatore

Dipartimento di Matematica e Informatica,
Università degli Studi di Salerno,
via ponte don Melillo, 84084 - Fisciano (SA), Italy
E-mails: {gfenza, loia, ssenatore}@unisa.it.

Abstract—*This paper proposes an algorithm for concept matching, applied in the ontology mapping domain. Basic idea is to seek the effective semantics embedded in the concept name by analyzing the context in which it appears. Through simple interactions with the known lexicon WordNet, the right meaning associated to a concept is unequivocally elicited by exploring their local semantic context, viz. the surrounding concepts.*

This approach reveals interesting results for the word sense disambiguation, when polysemy problems requires a semantic interpretation.

Keywords— Semantic Web, concept matching, ontology mapping, WordNet, concept similarity.

1 Introduction

The Web comprises huge and disparate collections of information, whose size is estimated to overcome 11.5 billion web pages [1], though the search engines do not seem to cover the whole size (for instance, Google indexes approximately 9 billion pages). The diffusion of the Semantic Web has promised new models to support integrated access to web resources and services as well as intelligent applications for information processing on the Web.

Ontologies represents a conceivable solution for data representation as well as the knowledge sharing, aimed at the integration of the Web content in a unique and coherent view. Nevertheless, due to the decentralized nature of the Web, a plethora of ontologies has been defined and disseminated on the net; often they describe overlapped application domains; sometimes are specialized for specific domain.

It is evident the exigency to find some semantic correspondence among concepts which refer to different ontologies in order to get a "semantic reconciliation", aimed at establishing interoperability between semantic Web applications and a more homogeneous integration of information [22].

Ontology mapping has been proposed as an effective way of handling the semantic heterogeneity problem. It plays a central role in many application domains, such as e-commerce, semantic web services matchmaking [22], information integration, query mediation [23], etc..

It is the process that discovers a set of semantic correspondence between some entities of different ontologies. Many research studies have yield approaches and tools for (semi) automatic ontology matching [2], [3]; structural and linguistic matching has been taken into account; often combined approaches for the matching valuation provide efficient results too.

This paper presents a simple approach to concept matching, based on the linguistic similarity. The main idea is to discriminate the effective meaning of a word by analyzing the context in which it appears, in order to overcome the polysemy problems and the lexical ambiguity.

The approach achieves a primitive ontology mapping, by discovering semantic correspondences between concepts of two ontologies (implemented as graph-like structures). A graphical interface presents the discovered mapping.

The paper is organized as follows. Section 2 sketches a state of art in the ontology matching domain. Section 3 introduces our approach, describing the basic algorithm and then in Section 4, implementation details and some relevant case studies are presented in order to validate its applicability. Finally, conclusions close the paper.

2 Related works

In the last years, ontologies are increasingly being used to support the integration of information. Yet, their diffusion in many Web areas emphasizes impressive heterogeneity among information sources and in particular in the formal model exploited to encode the domain conceptualizations.

In literature, different types of heterogeneity have been identified, mainly split in the following classes.

- *Syntactic heterogeneity*: represents the heterogeneity due to differences at the language level; for instance, when two ontologies are defined by using different knowledge representation formalisms even though the meaning is the same [4].
- *Terminological heterogeneity*: occurs when different names are given to similar ontology entities [6].
- *Semantic heterogeneity*: is often called conceptual heterogeneity. It occurs when the same domain of interest is modelled in different way, for instance, exploiting different types of axioms for defining concepts or just giving diverse expressive values to them [5].

Many efforts have been done from the research communities to find a solution to the matching problem by developing a variety of tools and mostly providing a well-founded formalization [7, 8, 9]. Matching ontologies (or schemas) has been recognized as a critical operation in many

ISBN: 978-989-95079-6-8

application domains. It takes as input two ontologies, each one consisting of a set of entities and produces as output the "connections" (which may represent equivalences or just relations such as consequence, subsumption, or disjointness) holding between these entities.
Systems for ontology matching such as Cupid [11], OntoBuilder [10], Similarity Flooding [2], Clio [12], Glue [13], S-Match [3], OLA [14] and Prompt [15] aim at providing an alignment, namely a set of correspondences between semantically related entities of different ontologies. Finding the correspondences which are effective from the user's viewpoint is often not easy.
In [17] a tree-based classification of elementary matching approaches is given. At granularity level, matching techniques can be distinguished into the following classification criteria:
– Element-level vs. structure-level: Element-level matching techniques discover correspondences by analyzing entities or their relative instances, without considering their relations with other entities or their instances. Structure-level techniques instead compute correspondences by studying the structure in which the concepts or their instances appear [16].
– Syntactic vs. external vs. semantic: The syntactic techniques consider the input with regard to its sole structure, according to some well stated algorithm. On the other hand, external techniques exploit external resources (such as human support, thesaurus, etc.) and common knowledge in order to interpret the input [17]. Finally, semantic techniques use some formal semantics (i.e. model-theoretic approaches) to interpret the input and justify their results. Semantic matching [3] is based on the idea that labels at concept nodes, which are written in natural language, are translated into propositional formulas which codify the intended meaning of the labels themselves. This way the matching problem becomes a propositional unsatisfiability problem, which can then be efficiently implemented through propositional satisfiability (SAT) solvers [18].
Although many advances in the ontology matching research topic have been addressed, nevertheless, there are still open issues to investigate in order to get a more efficient and effective integration of ontology matching tools in the web applications domain.

3 Ontology Matching

The ontology matching problem can be seen as a problem of concept matching. Two concepts match, if they are similar. In this section some formal notions are given to delineate a common background, useful to describe our ontology matching approach. Then the ontology matching method is detailed.

3.1 Scope of a concept into an ontology

Concept matching requires the evaluation of the similarity between two concepts. This approach aims at discovering linguistic similarities between the involved entities. In general, linguistic similarities are based on morphology and semantics, which are associated to the words that describe the relative entities. Thus, in this approach the similarity between two entities of different ontologies is evaluated not only by investigating the semantics of the entities names, but also taking into account the local context, through which the effective meaning is described.
Often the same word placed in different textual contexts assumes completely different meanings. In addition, lexicons are not able to disambiguate situations in which homonyms occur.
In order to deal with lexical ambiguity, this approach introduces the notion of "scope" of a concept which represents the context where the concept is placed.

Definition 1: *Let O be ontology and c is a concept in O. The scope of c, with radius r,* ***scope(c, r)*** *is a set of all the concepts outgoing from c included in a path of length r, with centre c. More formally:*

$$scope(c,r) = \{c' |\, c' \in O\ , dist(c,c') < r\}$$

where $dist(c,c')$ *is the number of edges that are in the path from concept c to concept c'. Let us note that* $dist(c,c') = 0\ iff\ c = c'$.

Figure 1 sketches the idea. The scope defines a round area composed of all concepts that are connected directly or indirectly to the central node. This area represents the context. Increasing the radius means to enlarge the scope (i.e. this area) and, consequently, the set of neighbour concepts that intervene in the description of the context.
Fixed the radius of the scope of two concepts α and β, belonging to different ontologies, our goal is to find out a semantic relationship exist between them.

Definition 2. Let *α be a concept in the ontology O. The concept name of α,* ***a= label(α)*** *is the linguistic label associated to the concept α.*

Definition 3. Given two concepts *α , β belonging respectively to* ontologies *O and O', let* $\mathcal{L}$ *be a lexicon and a= label(α) and b= label(β). Then, let lex (a, b) ∈ [0, 1] be a lexical similarity associated to the pair of concept names (α, β), with α∈O , β∈O'. The set L is composed of all pairs, defined as follows:*

$$L = \{(\alpha', \beta') | \forall\ \alpha \in scope(\alpha', r)\ and\ \forall\ \beta \in scope(\beta', r)\ and\ \exists lex(a, b) \neq 0)\}$$

WordNet is the lexicon exploited in this approach. It is one of the most common lexical databases, used for research studies in computational linguistics, text analysis, etc. The function *lex(a, b)* is the similarity metric.
Many similarity functions have been implemented in the WordNet package. In particular, a measure specifically developed for WordNet is the similarity proposed by Wu and Palmer [19]. It is simple to compute and results an efficacy metric for ontology matching.

ISBN: 978-989-95079-6-8

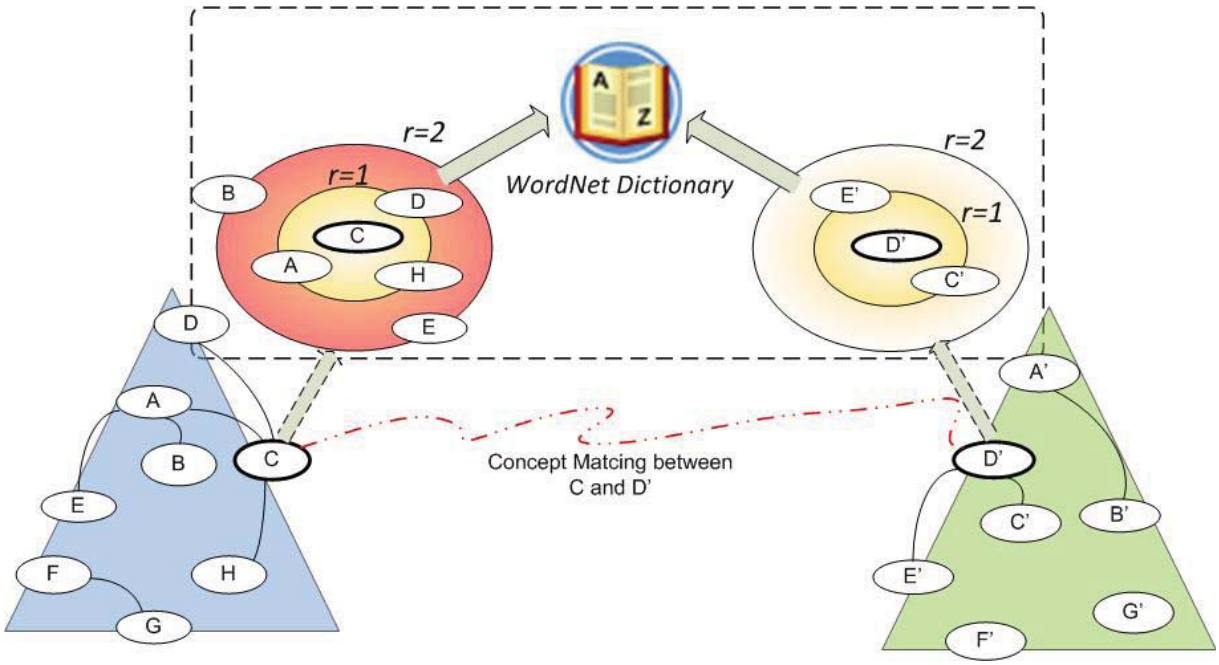


Figure 1: Intuitive idea of concept matching, exploiting the *scope* and the effective *sense* of the concept names

The Wu-Palmer similarity is based on the observation that although two concepts near the root of a hierarchy are close to each other in terms of edges, they can be very different semantically; conversely, two classes under one concept that are distant as number of edges, could be more similar conceptually. The idea is to find a common closest hyperonym in the path from them to the root.
The principle of similarity computation between two nodes is based on the distance which separates these two nodes from the root node and the distance which separates their closest common ancestor from the root too.
Let us assume that the arcs in ontology represent uniform distances (i.e. all the semantic edges have the same weight).

Definition 4. *Let c_1 and c_2 be two ontology elements in the ontology O. The distance between two nodes $\delta(c_1, c_2)$ is represented by the minimum number of edges that connect them.*

Definition 5. *Given an ontology O formed by a set of nodes and a root node R. Let c_1 and c_2 be two ontology concepts of which we will calculate the similarity. Then, g is the common ancestor of c and c'. The Wu–Palmer similarity is defined by the following expression[19]:*

$$sim(c_1, c_2) = \frac{2 \times \delta(g, R)}{\delta(c_1, R) + \delta(c_2, R)}$$

3.2 Concept Sense Discrimination

Through the given definitions, it is possible to individuate the meaning of a name associated to an ontology concept. Given a word, WordNet provides a list of all the synsets and word senses, related to that word. Just to give an example, submitting the word "table" to WordNet, the result is the following list.

```
The noun table has 6 sense (first 3 form tagged
texts)

1. (57) table, tabular array -- (a set of data
   arranged in rows and columns; "see table 1")
2. (25) table -- (a piece of furniture having a
   smooth flat top that is usually supported by
   one or more vertical legs; "it was a sturdy
   table")
3. (5) table -- (a piece of furniture with
   tableware for a meal laid out on it; "I
   reserved a table at my favorite restaurant")
4. mesa, table -- (flat tableland with steep
   edges; "the tribe was relatively safe on the
   mesa but they had to descend into the valley
   for water")
5. table -- (a company of people assembled at a
   table for a meal or game; "he entertained the
   whole table with his witty remarks")
6. board, table -- (food or meals in general;
   "she sets a fine table"; "room and board")
```

Let us note all the senses are ranked with respect to the frequency of the term (shown in the parenthesis for the first three senses) in the reference context (according to the spoken common sense).
The following pseudo-code details the algorithm for discriminate the actual sense of a word associated to a given concept. The algorithm takes as input an ontology O, a reference concept α in that ontology and a word *w*. The word *w* represents the name associated to the concept α (i.e. w=label(α)).
Let us note the semantic difference between the concept (or class) in an ontology and the label associated to that concept. The algorithm replies to question like "I would like to know the effective sense of the word w, placed in the context (or scope) of the concept α".

ISBN: 978-989-95079-6-8

Algorithm1

```
Input: Ontology O, concept α ∈ O, radius r and the
word w

Output: sense number of w

1: build an array N[|synset(w)|];
2: for each t1 in synset(w)
3:    initialize N[senseNumber(t1)] = 0;
4:    for i = 1 to r
5:       set M = ∅;
6:       for each c in scope(α,i) - scope(α,i-1)
7:          set S = ∅;  //set of similarity values
8:          for each t2 in synset(label(c))
9:             sim = similarity_WP(t1, t2)
10:            S = S U {sim};
11:         end for each
12:
13:         set max = maximum in S;
14:         max = max / |scope(α, i)|;
15:         M = M U {max};
16:      end for each
17:      set Sum = summation of all the values in M
18:      Sum = Sum /i;
19:      N[senseNumber(t1)] += Sum;
20:   end for
21: end for each
22: i1 = sense number with the highest
                              frequency in N.
23: return i1
```

First step (line 1) is to declare a vector structure whose size corresponds to the number of synsets (or senses) associated to the given word w. Goal is to maintain in each cell of the vector a pertinence value that represents how much the word w is semantically related to that sense (or belongs to that synset).
The algorithm selects all the concepts in the scope of α (belonging to the reference ontology O) by varying the radius (lines 4-6), in order to get different set of terms. Then compute the Wu-Palmer similarity between two terms coming from the concept name of α and the word *w* (line 9).
Just to give an idea about how the algorithm works, let us look at Figure 2. For simplicity, let us suppose *w* = *label(α)* = "Table". The algorithm evaluates the similarity between all the terms coming from both the WordNet synset of "Table" and the scope of concept "Table" in the ontology.
In particular, Figure 2 shows the flat ring shaped areas in different colours, by varying the radius.
For each concept *c* in the ring shaped area (computed as the difference of the areas between two successive radii, see line 6), the max similarity values between the name associated to *c* and a concept name in synset of α are maintained in M (line 15).
At the end of the two loops (lines 6-11) the variable *Sum* contains the sum of all the max similarity values computed for each couple of terms coming from the fixed term t1 of synsets of α and all the terms in the synset of *w*.
The *Sum* is "weighted" with respect to the current radius (line 18). The final value of *Sum* is added to the value stored in the cell associated to the sense of some term t1. This is repeated for each t1 in the synset of *w*. At the end, in each cell of the vector N there is a value, associated to each term in the synset of *w*.

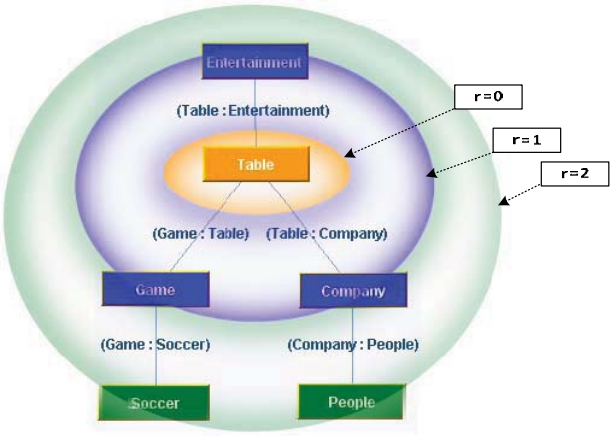


Figure 2. A sketched ontology and the scope of concept "Table" at different radius r.

The index of the vector cell, whose value is the maximum, represents the sense number to give as output.
Applying the algorithm on the ontology of Figure 2, the retuned index value is 5.
Abstractly speaking, the approach helps the user to individuate the right meaning of a word, given the context.

3.3 Ontology mapping

As said, the ontology mapping outlines correspondences or matches between concepts coming from two different ontology. In this approach, the concept matching is measured by computing a similarity between concepts at linguistic level.
The algorithm described in the previous section has been exploited to evaluate the concept matching. More formally, given two ontologies *O* e *O'* and two concepts *c* and *c'*, belonging respectively to these two ontologies, there is a match between *c* and *c'* if a similarity between them exists, computed as follows:

1. Algorithm1 is invocated twice: one time it takes as input the concept *c* and the ontology *O* and another time, by giving the concept *c'* and the ontology *O'*. Outputs of these two independent executions of Algorithm1 are two indexes, *i* for the synset of *c* and *i'* for the synset of c'. As said, they identify the sense number associated to each concept.

2. Once discovered the sense of involved concepts, the affinity between the concepts is computed by a similarity measure (for instance, the Wu-Palmer similarity defined in Section 3.1) between *c* and *c'*.

In order to obtain all the semantic correspondences among the concepts in the two ontologies, this procedure can be applied for each couple of concepts coming from two reference ontologies. Final result is an ontology mapping; a similarity value is assigned to each discovered correspondence between concepts in the two ontologies.

ISBN: 978-989-95079-6-8

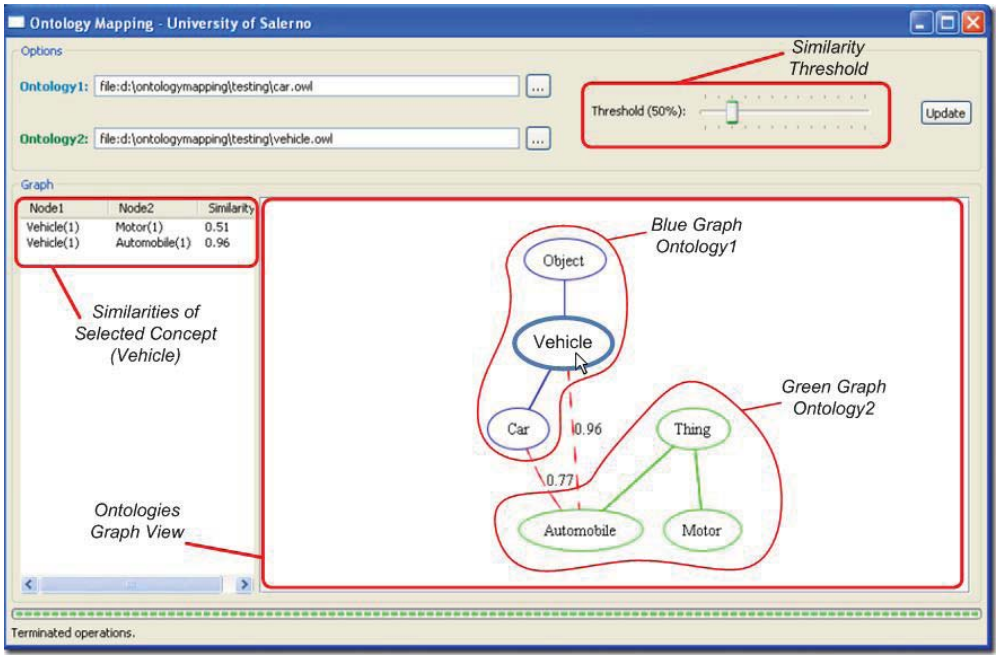


Figure 3. Ontology Mapping GUI

4 User interface and application details

The algorithm has been implemented in Java language. It exploits Java WordNet Similarity Library (JWSL) [20] for accessing the WordNet database. JSWL implements the most commons similarity and relatedness measures between words and get information about *synsets*.

A simple graphical interface has been designed for the user interaction (see Figure 3). It has been implemented by exploiting Graphviz library [21]. It is an open source graph visualization software which provides several main graph layout programs.

Let us note the ontologies have been represented (and implemented) as graphs, whose nodes represent the concepts whereas the edges connects two nodes when a relationship exists between them. In this version, a restriction on just hierarchies or taxonomies (i.e. only IS-A relationships) is considered.

The user can load the two ontologies and the final mapping is drafted in the *Ontology Graph View* panel (see Figure 3). The correspondences discovered by the mapping are sketched in Figure 3 as dotted lines between the concepts. A weight (value of similarity) is associated to each arc.

Figure 3 shows on the right hand of the interface, a sliding bar for setting the similarity threshold. Moving the cursor, the threshold is modified, returning in the graph view panel just the arcs whose similarity values are greater than this threshold.

Moreover, when the user clicks on a concept (node) in the graph view panel, all its connections with other concepts and the relative similarity values are shown on the left of the interface.

4.1 Some case studies

The algorithm has been applied on some sketched ontologies. The application acquires the ontologies in OWL format, although this prototypal version is limited to work on hierarchical structure.

In the interface of Figure 3, a simple example of ontology mapping is shown. The case study has been built ad-hoc; just the relevant portions of two ontologies are shown.

The Ontology1 (in blue color in Figure 3) is described by the concepts named “Object”, “Vehicle” and “Car” , whereas the Ontology2 (green color) is composed of the concept with names “Thing”, “Automobile” and “Motor”.

The application adds an arc between two concepts of different ontologies, when a similarity value (in the range [0, 1]) is found.

Let us note there exists an arc between the concept names “Vehicle” and “Automobile” with an high similarity (0.96), but the arc between “Vehicle” and “Motor” is not inserted in the graph view, because the similarity between these concept names assumes values lower than the selected threshold.

Similar considerations can be done on the ontology mapping shown in Figure 4 (although the GUI is not shown).

Two ontologies are drafted in different colors. Both of them describe a similar domain (the living beings).

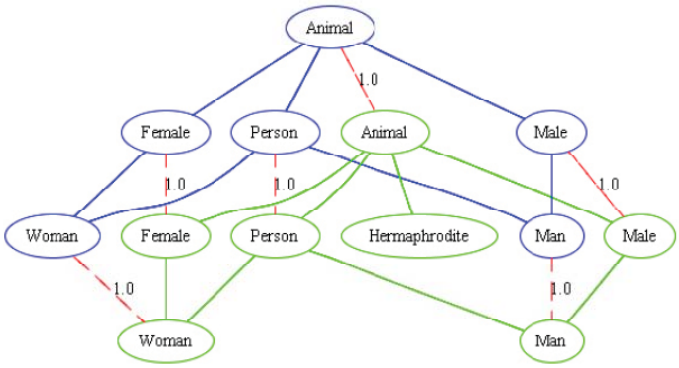


Figure 4: Example of Ontology Matching on the human beings domain.

ISBN: 978-989-95079-6-8

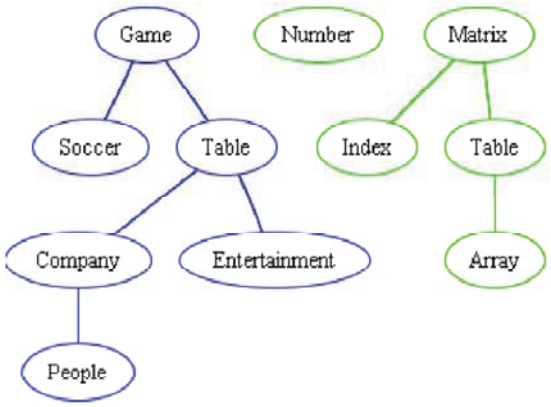


Figure 5: No concept matching occurs between these two ontologies, due to different meanings of the reference ontology domains.

Because the ontologies present concepts with similar names and meanings, the similarity values on the added arcs (in red) are equal to 1 (the maximum similarity).
Finally, Figure 5 shows a case of no concept match. The two ontologies share the same concept name "Table", but the reference context is completely different. In fact, Algorithm1 returns the sense number 5 for the ontology on the left (colored blue) and the sense number 1 for the ontology on the right, respectively (see Section 3.2 for details). Thus no arcs can be placed between them.
Let us note, in the example in Figure 5, the node with concept name "Number" is disconnected by the others, because the algorithm does not analyzes the OWL ontology relationships that are not hierarchical.

5 Conclusions

The paper describes a study for ontology mapping based on the discovering of linguistic similarity. A graphical interface allows the user to see the final mapping, presenting a similarity value for each discovered concept matching.
This algorithm represents an initial version of an ontology matching method and needs some extension for processing all the ontology relationships. Its efficacy would be visible if it was exploited in combination with structure based ontology matching. Further developments are addressed to reach these issues. Nevertheless, its applicability may not restricted to ontology matching problems. The basic algorithm could be used for Information Retrieval problems, for instance when the textual analysis requires the word sense disambiguation. The context of a word could be described by defining a "sliding window" on the text surrounding the analyzed word, in order to discover its appropriate sense. Furthermore, the algorithm could be exploited for discriminating common sense words, in specialized contexts where the word meaning depends on the application domain (or the reference ontology).

References

[1] Gulli, A. and Signorini, A. 2005. The indexable web is more than 11.5 billion pages. In *Special interest Tracks and Posters of the 14th international Conference on World Wide Web* (Chiba, Japan, May 10 - 14, 2005). WWW '05. ACM, New York, NY, 902-903.

[2] S. Melnik, H. G. Molina, E. Rahm, Similarity Flooding: A Versatile Graph Matching Algorithm and Its Application to Schema Matching, *In Proceedings of the 18th International Conference on Data Engineering*, San Jose, CA, 2002, pp.117-128.

[3] F. Giunchiglia, M. Yatskevich, E. Giuchilglia, Efficient Semantic Matching, *In Proceedings of ESWC, Heraklion*, Greece, 2005, pp.272-289.

[4] J. Euzenat and H. Stuckenschmidt. The 'family of languages' approach to semantic interoperability. In Borys Omelayenko and Michel Klein, editors, Knowledge transformation for the semantic web, pages 49–63. IOS press,Amsterdam (NL), 2003.

[5] M. Klein. Combining and relating ontologies: an analysis of problems and solutions. In Proc. IJCAI Workshop on Ontologies and Information Sharing, Seattle (WA US), 2001.

[6] N. Noy and M. Klein. Ontology evolution: Not the same as schema evolution. Knowledge and Information Systems, 6(4):428–440, 2004.

[7] P. Bernstein, A. Halevy, and R. Pottinger. A vision of management of complex models. *ACM SIGMOD Record*, 29(4):55–63, 2000

[8] P. Bouquet, M. Ehrig, J. Euzenat, E. Franconi, P. Hitzler, M. Krotzsch, L. Serafini, G. Stamou, Y. Sure, and S. Tessaris. Specification of a common framework for characterizing alignment. Deliverable D2.2.1, Knowledge web NoE, 2004

[9] A. Zimmermann, M. Krotzsch, J. Euzenat, and P. Hitzler. Formalizing ontology alignment and its operations with category theory. In Proc. 4th International Conference on Formal Ontology in Information Systems (FOIS),pages 277–288, Baltimore (MD US), 2006

[10] A. Gal, G. Modica, H. Jamil, and A. Eyal. Automatic ontology matching using application semantics. AI Magazine, 26(1):21{32, 2005.

[11] J. Madhavan, P. Bernstein, and E. Rahm. Generic schema matching with Cupid. In Proceedings of the 27th International Conference on Very Large Data Bases (VLDB), pages 48-58, Rome, Italy, 2001.

[12] R. Miller, M. Hernandez, L. Haas, L.-L. Yan, C. Ho, R. Fagin, and L. Popa. The Clio project: Managing heterogeneity. SIGMOD Record, 30(1):78-83,2001.

[13] A. Doan, J. Madhavan, P. Domingos, and A. Halevy. Learning to map between ontologies on the semantic web. In Proceedings of the 11th International Conference on World Wide Web (WWW), pages 662{673, Honolulu, USA, 2002.

[14] J. Euzenat and P. Valtchev. Similarity-based ontology alignment in OWLlite. In Proceedings of the 15th European Conference on Artificial Intelligence (ECAI), pages 333-337, Valencia, Spain, 2004.

[15] N. Noy and M. Musen. The PROMPT suite: interactive tools for ontology merging and mapping. International Journal of Human-Computer Studies, 59(6):983-1024, 2003.

[16] J. Kang and J. Naughton. On schema matching with opaque column names and data values. In Proc. 22nd International Conference on Management *of Data (SIGMOD)*, pages 205–216, San Diego (CA US), 2003.

[17] P. Shvaiko and J. Euzenat. A survey of schema-based matching approaches. Journal on Data Semantics, IV:146–171, 2005.

[18] D. Le Berre JSAT: the java satisfiability library http://www.sat4j.org/

[19] Z. Wu and M. Palmer. Verb semantics and lexical selection. In Proc. 32nd Annual Meeting of the Association for Computational Linguistics (ACL), pages 133–138, Las Cruces (NM US), 1994.

[20] Java WordNet Similarity Library (JWSL) Available: http://grid.deis.unical.it/similarity/

[21] Graphviz - Graph Visualization. Available: http://www.graphviz.org/

[22] G. Fenza, V. Loia, S. Senatore, A hybrid approach to semantic web services matchmaking, International Journal of Approximate Reasoning, Volume 48, Issue 3, August 2008, pp. 808-828.

[23] G. Fenza, V. Loia, S. Senatore: *Improving Fuzzy Service Matchmaking through Concept Matching* Discovery. FUZZ-IEEE 2007: pp. 1-6

ISBN: 978-989-95079-6-8

Using fuzzy linguistic summaries for the comparison of time series: an application to the analysis of investment fund quotations

Janusz Kacprzyk[1,2] Anna Wilbik[1*]

[1] Systems Research Institute, Polish Academy of Sciences
Newelska 6, 01-447 Warsaw, Poland
[2] WIT – Warsaw School of Information and Technology
Newelska 6, 01-447 Warsaw, Poland
Email: kacprzyk@ibspan.waw.pl, wilbik@ibspan.waw.pl

Abstract— *We propose a new, human consistent method for the evaluation of similarity of time series that uses a fuzzy quantifier base aggregation of trends (segments), within the authors' (cf. Kacprzyk, Wilbik, Zadrożny [1, 2, 3, 4, 5, 6] or Kacprzyk, Wilbik [7, 8, 9]) approach to the linguistic summarization of trends based on Zadeh's protoforms and fuzzy logic with linguistic quantifiers. The results obtain are very intuitively appealing and justified by valuable outcomes of similarity analyses between quotations of an investment fund and the two main indexes of the Warsaw Stock Exchange.*

1 Introduction

Among various application areas of advanced data mining and knowledge discovery tools and techniques, all kinds of finance related areas are of primordial and growing importance. For instance, in an interesting statistical report [10] on top data mining applications in 2008, the first two positions are, in the sense of yearly increase:

• *Investment/Stocks*, up from 3% of respondents in 2007 to 14% of respondents in 2008% (350% increase),

• *Finance*, up form 7.2% in 2007 to 16.8% in 2008 (108% increase).

This trend will presumably continue in view of serious difficulties on financial/investment market which is expected to continue well after 2009.

This paper is a continuation of our previous works (cf. Kacprzyk, Wilbik, Zadrożny [1, 2, 3, 4, 5, 6] or Kacprzyk, Wilbik [7, 8, 9]) which deal with the problem of how to effectively and efficiently support a human decision maker in making decisions concerning investments. We deal mainly with investments in investment (mutual) funds. Clearly, decision makers are here concerned with possible future gains/losses, and their decisions is related to what might happen in the future. However, our aim is not the forecasting of the future daily prices, which could have been eventually used directly for a purchasing decision. Instead, in our works, we follow a decision support paradigm, that is we try to provide the decision maker with some information that can be useful for his/her decision on whether and how many units to purchase. We do not intend to replace the human decision maker.

This problem is very complex. First of all, there may be two general approaches. The first one, which may seem to be the most natural is to provide means to derive a price forecast for an investment unit so that the decision maker could "automatically' purchase what has been forecast, and as much as he/she could afford. Unfortunately, the success in such a straightforward approach is much less than expected. Basically, statistical methods employed usually for this purpose are primitive in the sense that they just extrapolate the past and do not use domain knowledge, intuition, some inside information, etc. A natural solution may be to try to just support the human decision maker in making those investment decisions by providing him/her with some additional useful information, and not getting involved in the actual investment decision making.

Various philosophies in this respect are possible. Basically, from our perspective, the following one will be followed. In all investment decisions the future is what really counts, and the past is irrelevant. But, the past is what we know, and the future is (completely) unknown. Behavior of the human being is to a large extend driven by his/her (already known) past experience. We usually assume that what happened in the past will also happen (to some, maybe large extent) in the future. This is basically, by the way, the very underlying assumption behind the statistical methods too!

This clearly indicates that the past can be employed to help the human decision maker find a good solution. We follow here this path, i.e. we present a method to subsume the past, to be more specific the past performance of an investment (mutual) fund, by presenting results in a very human consistent way, using natural language statements.

It should be noted that this line of reasoning has often been articulated by many well known investment practitioners, and one can quote here some more relevant opinions. In any information leaflets of investment funds, one may always notice a disclaimer stating that "Past performance is no indication of future returns" which is true. However, on the other hand, for instance, in a well known posting "Past Performance Does Not Predict Future Performance" [11], they state something that may look strange in this context, namely: "... according to an Investment Company Institute study, about 75% of all mutual fund investors mistakenly use short-term past performance as their primary reason for buying a specific fund". But, in an equally well known posting "Past performance is not everything" [12], they state: "... disclaimers apart, as a practice investors continue to make investments based on a schemes past performance. To make matters worse, fund houses are only too pleased to toe the line by actively advertising the past performance of their schemes leading investors to conclude that

*Partially supported by the Ministry of Science and Higher Education Grant no. NN 516 4309 33.

ISBN: 978-989-95079-6-8

it is the single-most important parameter (if not the most important one) to be considered while investing in a mutual fund scheme".

As strange as this apparently is, we may ask ourselves why it is so. Again, in a well known posting "New Year's Eve: Past performance is no indication of future return" [13], they say "...if there is no correlation between past performance and future return, why are we so drawn to looking at charts and looking at past performance? I believe it is because it is in our nature as human beings ...because we don't know what the future holds, we look toward the past ...".

And, continuing along this line of reasoning, we can find many other examples of similar statements supporting our position. For instance, Myers [14] says: "...Does this mean you should ignore past performance data in selecting a mutual fund? No. But it does mean that you should be wary of how you use that information ...While some research has shown that consistently good performers continue to do well at a better rate than marginal performers, it also has shown a much stronger predictive value for consistently bad performers ...*Lousy performance in the past is indicative of lousy performance in the future...*". And, further: Bogle [15] states: "... there is an important role that past performance can play in helping you to make your fund selections. While you should disregard a single aggregate number showing a fund's past long-term return, you can learn a great deal by studying the *nature of its past returns*. Above all, look for consistency.". In [16], we find: "While past performance does not necessarily predict future returns, it can tell you how volatile a fund has been". In the popular "A 10-step guide to evaluating mutual funds" [17], they say in the last, tenth, advise: "Evaluate the funds performance. Every fund is benchmarked against an index like the BSE Sensex, Nifty, BSE 200 or the CNX 500 to cite a few names. Investors should compare fund performance over varying time frames vis-a-vis both the benchmark index and peers. Carefully evaluate the funds performance across market cycles particularly the downturns".

One can give many more quotations from all kinds of investment guides, analyses, etc. by leading experts. Virtually all of them emphasize the importance of looking at the past to help make future decisions. Moreover, they also generally advocate a more comprehensive look in the sense that what might be useful would rather be not particular single values but some deeper meaning, even nature of past behavior and returns.

We have followed this line of reasoning in our past papers (cf. Kacprzyk, Wilbik, Zadrożny [1, 2, 3, 4, 5, 6] or Kacprzyk, Wilbik [7, 8, 9]), i.e. to try to find a human consistent, fuzzy quantifier based scheme for a linguistic summarization of the past in terms of various aspects of how the time series representing daily quotations of the investment fund(s) behave. However, we have mainly concentrated on a sheer absolute performance, i.e. the time evolution of the quotations themselves. This may be relevant, and sometimes attractive to the users who can see a summary of their gains/loses and their temporal evolution.

However, for many more detailed analyses we need to relate the performance of an investment fund to the performance (time evolution) of some benchmark(s) against which the particular investment fund is to be compared, as stated in its prospectus. These benchmarks are usually some stock exchange indexes, or a composition thereof.

One of more interesting and relevant questions that may be posed by both a professional analyst or a customer whose money is being invested may be: how similar was the temporal evolution of quotations of the particular investment fund and its benchmark(s)? This has clearly to do with the measuring of similarity of time series.

The problem of similarity of time series or finding similar subsequences of time series is an important issue in many problems, e.g., in indexing [18, 19], clustering [20] or motif discovery [21, 22]. A similarity measure should allow for the imprecise match, i.e., should indicate time series sequences that are similar to some extent. Moreover, the algorithm should be very efficient [23], as we may consider very long time series.

Different approaches to the similarity measures of time series were considered and proposed in the literature. A simple solution is to view each sequence as a point in an n-dimensional Euclidean space, where n is the length of the segment (trend). The similarity or dissimilarity is based on the Euclidean distance, and computed as the norm in this space [24]. This approach can be used for various time series but normally some extensions should be applied providing various transformations such as scaling or shifting were proposed [21].

Another methods use, for instance, the Dynamic Time Wraping (DTW) [19] and Longest Common Subsequence (LCS) measures [24], and they allows gaps in sequences. They are successfully used in other domains such as speech recognition and text pattern matching [25].

Other approaches developed for similarity search in time series data are based on dimension reduction. This can be achieved in numerous ways, e.g. using the Discrete Fourier Transform (DFT) [23], wavelets [26], etc.

One can also mention here approaches based on the piecewise linear representation of time series [27]. In this approach time series are represented by segments (trends), not the particular values. The method proposed here falls into this category.

In this paper we propose new method of calculating the similarity of time series; as we have already indicated, this concerns the quotations of an investment fund and some benchmark. The degree of similarity of two time series is meant here as the degree to which, for instance, "most" of the long, overlapping segments are similar (for instance, such that "almost all" of their features are similar). Notice that this has very much to do with a soft definition of consensus meant as a degree to which, for instance, "*most* of the *important* individuals agree as to *almost all* of the *relevant* options". This was introduced by Kacprzyk and Fedrizzi [28, 29, 30], Kacprzyk, Fedrizzi and Nurmi [31], and recently by Kacprzyk and Zadrożny [32] and has been found useful in many areas.

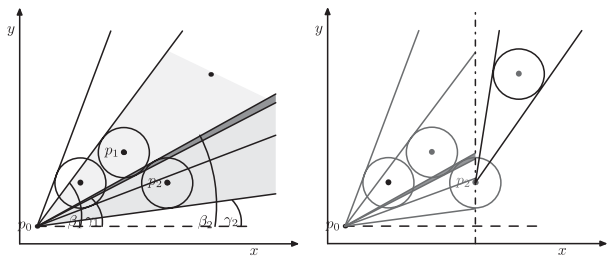
We will now discuss first the segmentation of time series, as the similarity will proceed in terms of segments (trends), not all the consecutive values. Then, we will propose some new, linguistic quantifier based measures of similarity, followed by examples on real data, and conclusions.

ISBN: 978-989-95079-6-8

2 Segmentation of time series

In our approach (cf. Kacprzyk, Wilbik, Zadrożny [1, 2, 3, 4, 5, 6] or Kacprzyk, Wilbik [7, 8, 9]) we summarize the trends (segments) extracted from a time series, and we have to extract these segments assumed to be represented by a fragment of a straight line. There are many algorithms for a piecewise linear segmentation of time series, including, e.g., on-line (sliding window) algorithms, bottom-up or top-down strategies (cf. Keogh [33, 34]). We may either divide the time series into k segments (using, e.g., bottom-up or top-down strategies), where k is specified by the user, or divide it so that the maximum error for any segment does not exceed some threshold value ε, also specified by the user. In our works [1, 2, 3, 4, 5, 6, 7, 8, 9] we used a simple on-line algorithm, a modification of one proposed by Sklansky and Gonzalez [35] in the contest of image analysis. The method is simple and fast, and fully satisfies our needs though many other methods are possible.

Our algorithm constructs the segments so that the maximum distance between the segment line and the observed time point is smaller than a user specified ε value. It works by constructing the intersection of cones starting from point p_i of the time series and including a circle of radius ε around the subsequent data points p_{i+j}, $j = 1, 2, \ldots$, until the intersection of all cones starting at p_i is empty. If for p_{i+k} the intersection is empty, then we construct a new cone starting at p_{i+k-1}. Figure 1 present the idea of the algorithm. The family of possible solutions is indicated as a gray area. Clearly other algorithms can also be used, and there is a lot of them in the literature (cg. [18, 36]).



(a) the intersection of the cones is indicated by the dark grey area

(b) a new cone starts in point p_2

Figure 1: Illustration of the segmentation algorithm used

To present details of the algorithm, let us first denote:
- `p_0` – a point starting the current cone,
- `p_1` – the last point checked in the current cone,
- `p_2` – the next point to be checked,
- `Alpha_01` – a pair of angles (γ_1, β_1), meant as an interval, that defines the current cone as shown in Figure 1(a),
- `Alpha_02` – a pair of angles of the cone starting at the point `p_0` and inscribing the circle of radius ε around the point `p_2` (cf. (γ_2, β_2) in Figure 1(a)),
- function `read_point()` reads a next point of data series,
- function `find()` finds a pair of angles of the cone starting at the point `p_0` and inscribing the circle of radius ε around the point `p_2`.

A pseudocode of the algorithm that extracts trends is depicted in Figure 2.

```
read_point(p_0);
read_point(p_1);
while(1)
{
  p_2=p_1;
  Alpha_02=find();
  Alpha_01=Alpha_02;
  do
  {
    Alpha_01 = Alpha_01 ∩ Alpha_02;
    p_1=p_2;
    read_point(p_2);
    Alpha_02=find();
  } while(Alpha_01 ∩ Alpha_02 ≠ ∅);
  save_found_trend();
  p_0=p_1;
  p_1=p_2;
}
```

Figure 2: Pseudocode of the modified Sklansky and Gonzalez [35] algorithm for extracting trends

The bounding values of `Alpha_02` (γ_2, β_2), computed by function `find()` correspond to the slopes of two lines that (1) are tangent to the circle of radius ε around point $p_2 = (x_2, y_2)$ and (2) start at the point $p_0 = (x_0, y_0)$. Thus

$$\beta_2, \gamma_2 = arctg\left(\frac{\Delta x \cdot \Delta y - \pm\varepsilon\sqrt{(\Delta x)^2 + (\Delta y)^2 - \varepsilon^2}}{(\Delta x)^2 - \varepsilon^2}\right)$$

where $\Delta x = x_0 - x_2$ and $\Delta y = y_0 - y_2$.

The resulting linear ε-approximation of a group of points `p_0, ..., p_1` is either a single segment, chosen as, e.g., a bisector of the cone, or one that minimizes the distance (e.g., the sum of squared errors, SSE) from the approximated points, or the whole family of possible solutions, i.e., rays of the cone.

This method is effective and efficient as it requires only a single pass through the data. From now on we will identify (partial) *trends* with the line segments of the constructed piecewise linear function.

3 Linguistic summarization of time series

First, we will outline the very essence of our approach to the linguistic summarization of time series which is based on the two types of protoforms of linguistic summaries of trends:
- a short form:

$$\text{Among all segments, } Q \text{ are } P \tag{1}$$

e.g.: "Among all segments, *most* are *slowly increasing*".
- an extended form:

$$\text{Among all } R \text{ segments, } Q \text{ are } P \tag{2}$$

e.g.: "Among all *short* segments, *most* are *increasing*".
whose truth values are, respectively:

$$\mathcal{T}(\text{Among all } y\text{'s, } Q \text{ are } P) = \mu_Q\left(\frac{1}{n}\sum_{i=1}^{n}\mu_P(y_i)\right) \tag{3}$$

$$\mathcal{T}(\text{Among all } Ry\text{'s, } Q \text{ are } P) = \mu_Q\left(\frac{\sum_{i=1}^{n}\mu_R(y_i) \wedge \mu_P(y_i)}{\sum_{i=1}^{n}\mu_R(y_i)}\right) \tag{4}$$

where $\wedge$ is the minimum operation (more generally it can be another appropriate operator, notably a t-norm).

ISBN: 978-989-95079-6-8

For more information on this approach, and on the advantages of using protoforms as general forms of linguistic summaries, see Kacprzyk, Wilbik, Zadrożny [1, 2, 3, 4, 5, 6] or Kacprzyk, Wilbik [7, 8, 9].

4 A new fuzzy quantifier based method for the evaluation of similarity of time series

The degree of similarity of two time series is meant here as the degree to which, say "most" of the long, simultaneous segments are similar (i.e., "majority" of their features are similar).

The time series is represented by its segments (trends). Each segment is described by the following four values:
• starting time, called here a partition point,
• duration of the segment (expressed in time units),
• dynamics of change (expressed as the slope of the segment),
• variability occurring in the segment, describing the variability of differences between the trend (segment line) and the particular time series values over the duration of this trend.

Let us assume, we wish to compare two time series A and B. As the result of the segmentation procedure time series A was divided into k segments, represented by a_j, $j = 1, \ldots, k$ and time series B was divided into l segments, represented by b_j, $j = 1, \ldots, l$.

Next we create the set of partition points of the two time series considered, i.e. this set contains the time points when at least one segment has started. There are at most $k + l$ such partition points. Those times are sorted, so that the earlier ones come first.

Now between any consecutive partition points (e.g. p_i and p_{i+1}) there is only one segment (or a part of it) of each time series. We may denote those segments as a_{p_i} and b_{p_i}, i.e., a_{p_i} is the segment representing time series A that take place between p_i and p_{i+1}. Now we can compute the similarity of such simultaneous segments.

We compute the degree of similarity of two segments (trends), as the degree to which "most" of their features are similar. Here we consider for simplicity the three features only: duration, dynamics and variability. The values of the features of segment a_{p_i} are $dur_{a_{p_i}}$ for duration, $dyn_{a_{p_i}}$ for dynamics, and $var_{a_{p_i}}$ for variability. R_{dur}, R_{dyn} and R_{var} are the ranges of possible values of duration, dynamics and variability. The degree of similarity of segments a_{p_i} and b_{p_i} with respect to duration is computed as

$$sim^{dur}(a_{p_i}, b_{p_i}) = \mu^{dur}_{sim}\left(\frac{|dur_{a_{p_i}} - dur_{b_{p_i}}|}{R_{dur}}\right) \quad (5)$$

Similarly, separate functions can be defined for the similarity for each feature, here μ^{dyn}_{sim} and μ^{var}_{sim}. Let $\mu_{Q_1}(.)$ be the membership function of quantifier "most", i.e regular, nondecreasing and monotone. Thus the degree of similarity of segments a_{p_i} and b_{p_i} is computed as

$$\begin{aligned} sim(a_{p_i}, b_{p_i}) = \mu_{Q_1}\Bigg(& w_{dur}\left(\mu^{dur}_{sim}\left(\frac{|dur_{a_{p_i}} - dur_{b_{p_i}}|}{R_{dur}}\right)\right) + \\ &+ w_{dyn}\left(\mu^{dyn}_{sim}\left(\frac{|dyn_{a_{p_i}} - dyn_{b_{p_i}}|}{R_{dyn}}\right)\right) + \\ &+ w_{var}\left(\mu^{var}_{sim}\left(\frac{|var_{a_{p_i}} - var_{b_{p_i}}|}{R_{var}}\right)\right)\Bigg) \end{aligned} \quad (6)$$

where $w_{dur} + w_{dyn} + w_{var} = 1$, here all are equal, $w_{dur} = w_{dyn} = w_{var} = \frac{1}{3}$.

Next, to obtain the similarity of A and B, we need to aggregate above similarity values between the overlapping segments, and again we use a linguistic quantifier driven aggregation.

$$sim(A, B) = \mu_Q\left(\frac{\sum_{i=1}^{n} \frac{p_{i+1}-p_i}{T} sim(a_{p_i}, b_{p_i})}{\sum_{i=1}^{n} \frac{p_{i+1}-p_i}{T}}\right) \quad (7)$$

where $p_{i+1} - p_i$ is the difference between the two consecutive partition points p_i and p_{i+1}, and T is the total time span considered.

As $\sum_{i=1}^{n} p_{i+1} - p_i = T$ we may simplify this formula and we obtain

$$sim(A, B) = \mu_Q\left(\sum_{i=1}^{n} \frac{p_{i+1} - p_i}{T} sim(a_{p_i}, b_{p_i})\right) \quad (8)$$

5 Numerical results

The method proposed in this paper was tested on data on quotations of an investment (mutual) fund that invests at most 50% of assets in shares listed at the Warsaw Stock Exchange (WSE). We have used two benchmark time series, the index of WSE companies (WIG), that is the benchmark for this fund mentioned in the fund prospectus, and the index of 20 biggest and most liquid companies (WIG20). More information can be found on the WSE Web page (`http://www.gpw.pl`).

Data shown in Fig. 3 were collected from the beginning of 2007 until the end of 2008 (501 quotations); notice that one can see when world wide financial problems begun. They are measured in different units, and in order to compare them, we have scaled them using the following formula. The new value of the quotation

$$\bar{v}_i = \frac{v_i - v_0}{v_0} 10 + 10$$

where v_0 is the quotation value of the first observation, i.e. on January, 2, 2007. Those scaled time series may be interpreted as the amount of money that would be earned if we invested PLN 10 (PLN stands for the Polish Zloty) in the mutual fund or stocks of the index on January 2, 2007.

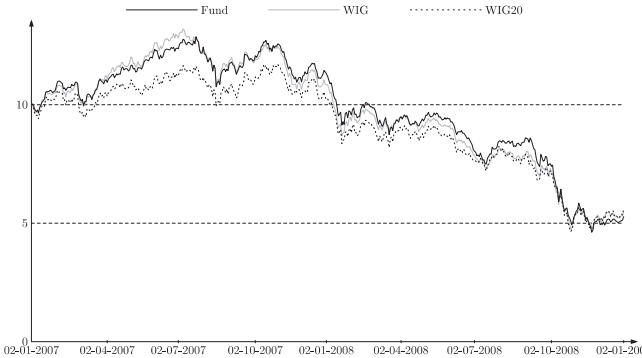


Figure 3: Daily quotations of an investment fund considered and the WIG and WIG20 indexes

The value of the mutual fund time series was equal PLN 5.28 at the end of the time span considered. The minimal value

ISBN: 978-989-95079-6-8

recorded was PLN 4.63 while the maximal one during this period was PLN 12.86.

The values of the benchmark time series were equal to PLN 5.46 and PLN 5.61 at the end of the time span considered for the WIG and WIG20, respectively. The minimum value of WIG recorded was PLN 4.71 while the maximum one during this period was PLN 13.20. The minimum and maximum values for WIG20 were PLN 4.62 and PLN 11.70, respectively.

Using the modified Sklansky and Gonzalez algorithm (cf. [2]) and $\varepsilon = 0.25$ we obtained 43 extracted trends (segments) for the fund time series, 51 for the WIG time series and 49 for the WIG 20 time series.

We can describe each time series independently using linguistic summaries of time series (cf. Kacprzyk, Wilbik, Zadrożny [1, 2, 3, 4, 5, 6] or Kacprzyk, Wilbik [7, 8, 9]). As those data sets serve only the purpose of example and are small, we applied the granulation with 3 labels for each feature only, namely *decreasing*, *constant* and *increasing* for the dynamics of change; *short*, *medium length* and *long* for duration and *low*, *moderate* and *high* for the variability.

The summaries obtained for the fund time series along with their truth values $\mathcal{T}$ are shown in Table 1.

Table 1: Linguistic summaries for the mutual fund time series

linguistic summary	$\mathcal{T}$
Among all y, most are constant	1
Among all medium y, most are constant	1
Among all moderate y, most are constant	1
Among all low y, most are constant	1
Among all medium and moderate y, most are constant	1
Among all medium and low y, most are constant	1
Among all long y, most are constant	1
Among all long and moderate y, most are constant	1
Among all short y, most are low	0.8107
Among all short y, most are constant	0.7819
Among all moderate y, most are medium	0.7462
Among all moderate y, most are medium and constant	0.7462
Among all long y, most are constant and moderate	0.7302
Among all long y, most are moderate	0.7302
Among all long and constant y, most are moderate	0.7302

The summaries obtained for the WIG index time series along with their truth values $\mathcal{T}$ are shown in Table 2.

Table 2: Linguistic summaries for the WIG index time series

linguistic summary	$\mathcal{T}$
Among all y, most are constant	1
Among all medium y, most are constant	1
Among all low y, most are constant	1
Among all moderate y, most are constant	1
Among all medium and low y, most are constant	1
Among all medium and moderate y, most are constant	1
Among all long y, most are constant	1
Among all short and moderate y, most are constant	1
Among all short y, most are constant	0.9950
Among all short and low y, most are constant	0.7856

The summaries obtained for the WIG20 index time series along with their truth values $\mathcal{T}$ are shown in Table 3.

Table 3: Linguistic summaries for the WIG20 index time series

linguistic summary	$\mathcal{T}$
Among all y, most are constant	1
Among all y, most are constant and low	1
Among all y, most are low	1
Among all constant y, most are low	1
Among all low y, most are constant	1
Among all medium y, most are constant	1
Among all medium y, most are constant and low	1
Among all medium y, most are low	1
Among all medium and constant y, most are low	1
Among all medium and low y, most are constant	1
Among all short y, most are low	1
Among all short and constant y, most are low	1
Among all long y, most are constant	1
Among all long y, most are constant and low	1
Among all long y, most are low	1
Among all long and constant y, most are low	1
Among all long and low y, most are constant	1
Among all short and low y, most are constant	0.9713
Among all short y, most are constant	0.9465
Among all short y, most are constant and low	0.8850
Among all constant y, most are medium	0.7333

The degree of similarity between the fund quotation time series and the WIG time series is equal 0.8622, while the degree of similarity of the fund quotation time series and the WIG20 time series is equal to 0.7529. This result clearly confirms our intuition. This may be viewed, on the one hand, as a proof of usefulness of our method, and on the other hand is quite natural because the investment fund considered in based on the WIG index. Luckily enough, the similarity with the WIG20 index, that of the biggest and most liquid companies, is also at a similarly high level.

6 Concluding remarks

We have proposed a new, human consistent method for the evaluation of similarity of time series, within our approach to the linguistic summarizatioon of time series that uses a fuzzy quantifier based aggregation of trends. The results obtained are very intuitively appealing. This is justified by the analysis of similarity between quotations of an investment fund and the two main indexes of the Warsaw Stock Exchange, WIG and WIG20. It also seems that a relation to natural language generation (NLG), along the lines of Kacprzyk and Zadrożny [37], may be very promising.

References

[1] J. Kacprzyk, A. Wilbik, and S. Zadrożny. Linguistic summarization of trends: a fuzzy logic based approach. In *Proceedings of the 11th International Conference Information Processing and Management of Uncertainty in Knowledge-based Systems*, pages 2166–2172, 2006.

[2] J. Kacprzyk, A. Wilbik, and S. Zadrożny. On some types of linguistic summaries of time series. In *Proceedings of the 3rd International IEEE Conference "Intelligent Systems"*, pages 373–378. IEEE Press, 2006.

[3] J. Kacprzyk, A. Wilbik, and S. Zadrożny. Linguistic summarization of time series by using the choquet integral. In P. Melin, O. Castillo, L.T. Aguilar, J. Kacprzyk, and W. Pedrycz, editors,

ISBN: 978-989-95079-6-8

Foundations of Fuzzy Logic and Soft Computing - IFSA 2007, pages 284–294. Springer-Verlag, Berlin and Heidelberg, 2007.

[4] J. Kacprzyk, A. Wilbik, and S. Zadrożny. Linguistic summaries of time series via an owa operator based aggregation of partial trends. In *Proceedings of the FUZZ-IEEE 2007 IEEE International Conference on Fuzzy Systems*, pages 467–472. IEEE Press, 2007.

[5] J. Kacprzyk, A. Wilbik, and S. Zadrożny. Mining time series data via linguistic summaries of trends by using a modified sugeno integral based aggregation. In *IEEE Symposium on Computational Intelligence and Data Mining CIDM 2007*, pages 467–472. IEEE Press, 2007.

[6] J. Kacprzyk, A. Wilbik, and S. Zadrożny. Linguistic summarization of time series using a fuzzy quantifier driven aggregation. *Fuzzy Sets and Systems*, 159(12):1485–1499, 2008.

[7] J. Kacprzyk and A. Wilbik. An extended, specificity based approach to linguistic summarization of time series. In *Proceedings of the 12th International Conference Information Processing and Management of Uncertainty in Knowledge-based Systems*, pages 551–559, 2008.

[8] J. Kacprzyk and A. Wilbik. A new insight into the linguistic summarization of time series via a degree of support: Elimination of infrequent patterns. In D. Dubois, M.A. Lubiano, H. Prade, M. A. Gil, P. Grzegorzewski, and O. Hryniewicz, editors, *Soft Methods for Handling Variability and Imprecision*, pages 393–400. Springer-Verlag, Berlin and Heidelberg, 2008.

[9] J. Kacprzyk and A. Wilbik. Linguistic summarization of time series using linguistic quantifiers: augmenting the analysis by a degree of fuzziness. In *Proceedings of 2008 IEEE World Congress on Computational Intelligence*, pages 1146–1153. IEEE Press, 2008.

[10] Subject: Poll results: Top data mining applications in 2008 and trends. http://www.kdnuggets.com/news/2009/n01/1i.html.

[11] Past performance does not predict future performance. www.freemoneyfinance.com/2007/01/past_performanc.html.

[12] Past performance is not everything. www.personalfn.com/detail.asp? date=9/1/2007&story=3.

[13] New year's eve:past performance is no indication of future return. stockcasting.blogspot.com/2005/12/new-years-evepast-performance-is-no.html.

[14] R. Myers. Using past performance to pick mutual funds. *Nation's Business*, Oct, 1997. findarticles.com/p/articles/mi_m1154/is_n10_v85/ai_19856416.

[15] J. C. Bogle. *Common Sense on Mutual Funds: New Imperatives for the Intelligent Investor*. Wiley, New York, 1999.

[16] U.S. Securities and Exchange Commission. Mutual fund investing: Look at more than a fund's past performance. www.sec.gov/investor/pubs/mfperform.htm.

[17] A 10-step guide to evaluating mutual funds. www.personalfn.com/detail.asp?date=5/18/2007&story=2.

[18] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra. Locally adaptive dimensionality reduction for indexing large time series databases. In *Proceedings of ACM SIGMOD Conference on Management of Data*, pages 151–162, Santa Barbara, CA, 2001.

[19] E. Keogh and C. A. Ratanamahatana. Exact indexing of dynamic time warping. *Knowledge Information Systems*, 7(3):358–386, 2005.

[20] P. Geurts. Pattern extraction for time series classification. In *PKDD '01: Proceedings of the 5th European Conference on Principles of Data Mining and Knowledge Discovery*, London, UK, 2001. Springer-Verlag.

[21] B. Chiu, E. Keogh, and S. Lonardi. Probabilistic discovery of time series motifs. In *Proceedings of the the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 493–498, 2003.

[22] A. Mueen, E. Keogh, Q. Zhu, S. Cash, and B. Westover. Exact discovery of time series motifs. In *SDM 2009, in press*, 2009.

[23] R. Agrawal, C. Faloutsos, and A. Swami. Efficient similarity search in sequence databases. In *Proceedings of the 4th International Conference on Foundations of Data Organization and Algorithms*, pages 69–84, 1993.

[24] G. Das, D. Gunopulos, and H. Mannila. Finding similar time series. In Jan Komorowski and Jan Zytkow, editors, *Principles of Data Mining and Knowledge Discovery (PKDD'97)*, pages 88–100. Springer-Verlag, 1997.

[25] N. Nakatsu, Y. Kambayashi, and S. Yajima. A longest common subsequence algorithm suitable for similar text strings. *Acta Informatica*, 18:171–179, 1982.

[26] K.-P. Chan and W.-C. Fu. Efficient time series matching by wavelets. In *Proceedings of the 15th International Conference on Data Engineering*. IEEE Computer Society, 1999.

[27] H. Shatkay and S. B. Zdonik. Approximate queries and representations for large data sequences. In *Proceedings 12th International Conference on Data Engineering*, pages 536–545, 1996.

[28] J. Kacprzyk and M. Fedrizzi. 'soft' consensus measures for monitoring real consensus reaching processes under fuzzy preferences. *Control and Cybernetics*, 15:309–323, 1986.

[29] J. Kacprzyk and M. Fedrizzi. A 'soft' measure of consensus in the setting of partial (fuzzy) preferences. *European Journal of Operational Research*, 34:315–325, 1988.

[30] J. Kacprzyk and M. Fedrizzi. A 'human-consistent' degree of consensus based on fuzzy logic with linguistic quantifiers. *Mathematical Social Sciences*, 18:275–290, 1989.

[31] J. Kacprzyk, M. Fedrizzi, and H. Nurmi. Group decision making and consensus under fuzzy preferences and fuzzy majority. *Fuzzy Sets and Systems*, 49:21–31, 1992.

[32] J. Kacprzyk and S. Zadrożny. Towards a general and unified characterization of individual and collective choice functions under fuzzy and nonfuzzy preferences and majority via the ordered weighted average operators. *International Journal of Intelligent Systems*, 24(1):4–26, 2009.

[33] E. Keogh, S. Chu, D. Hart, and M. Pazzani. An online algorithm for segmenting time series. In *Proceedings of the 2001 IEEE International Conference on Data Mining*, 2001.

[34] E. Keogh, S. Chu, D. Hart, and M. Pazzani. Segmenting time series: A survey and novel approach. In M. Last, A. Kandel, and H. Bunke, editors, *Data Mining in Time Series Databases*. World Scientific Publishing, 2004.

[35] J. Sklansky and V. Gonzalez. Fast polygonal approximation of digitized curves. *Pattern Recognition*, 12(5):327–331, 1980.

[36] E. Keogh and M. Pazzani. An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback. In *Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining*, pages 239–241, New York, NY, 1998.

[37] J. Kacprzyk and S. Zadrożny. Data mining via protoform based linguistic summaries: Some possible relations to natural language generation. In *2009 IEEE Symposium Series on Computational Intelligence Proceedings*, pages 217–224, Nashville, TN, 2009.

ISBN: 978-989-95079-6-8

Exchanging iterated expectations of random upper semicontinuous functions: an application to decision theory

Miguel López-Díaz Luis J. Rodríguez-Muñiz

Department of Statistics, Operations Research and Mathematical Teaching, University of Oviedo
c/ Calvo Sotelo s/n, E33071 Oviedo (Spain)
Email: {mld,luisj}@uniovi.es

Abstract— *In this paper we present a procedure to deal with a kind of single-stage decision problems with imprecise utilities. In this type of problems the product measurability of the utility function is not required. So that, the involved expectations are calculated by means of iterated integrals instead of integrals over product spaces.*

Keywords— Bayesian decision analysis, Iterated expectation, Kudo-Aumann's integral, Random upper semicontinuous function, Uncertainty modeling.

1 Introduction

There are situations in which product measurability of certain mappings is not satisfied, but iterated integrals are well-defined (see [10]). This type of situations appear in single-stage decision problems where we need conditions which allows us to exchange iterated expectations in order to perform a Bayesian analysis. This paper studies the case of Bayesian analysis of single-stage decision problems with imprecise utilities and non product-measurable utility funtion.

Several studies have been developed before to evaluate imprecise utilities, see for instance Watson *et al.* [27], Tong and Bonissone [26], Dubois aand Prade [7, 8, 9], Gil and Jain [11], Billot [1], Chen and Klein [3], Gil *et al.* [13], Krätschmer [16], Bordley [2], Rébillé [22] or Rodríguez-Muñiz and López-Díaz [25].

Here we model the imprecise utilities by means of fuzzy-valued utility functions (based on the concept of random upper semicontinuous functions or fuzzy random variable). However, only in [25] the non-product measurable case has been analyzed for this type of furrzy utilities, based on the theoretical results on iterated integrals of random upper semicontinuous functions given in [24].

In this paper we gather theoretical and applied results about how to deal with the type of problems referred above.

The paper is organized as follows: Preliminaries and notation constitutes Section 2, Section 3 contains theoretical results and Section 4 gather the statistical decision analysis results.

2 Preliminaries

Let us consider $\mathcal{K}_c$ the class of nonempty compact convex subsets of $\mathbb{R}$, endowed with a semilinear structure by means of the Minkowski addition and the product by a scalar. Also, consider the *Hausdorff metric* on $\mathcal{K}_c$ (see [6]). On a measurable space $(\Omega, \mathcal{A})$ we can define $S : \Omega \to \mathcal{K}_c$ a randon set as a $\mathcal{A}|\mathcal{B}_{d_H}$-measurable mapping ([14]).

A random set S is said to be *integrably bounded with respect to measure* μ, if $\|S\| \in L^1(\Omega, \mathcal{A}, \mu)$, where $\|S\|(\omega) = \sup_{x \in S(\omega)} \|x\|$. The *integral*, or *expected value* in case of μ being a probability, of S, is given by the *Kudo-Aumann integral* and it will be denoted either by $\int_\Omega S(\omega)\, d\mu(\omega)$ or $E(S|\mu)$ ([14]).

Let $\mathcal{F}_c$ denote the class of upper semicontinuous functions or *fuzzy sets* $U : \mathbb{R} \to [0,1]$ such that $U_\alpha \in \mathcal{K}_c$ for all $\alpha \in [0,1]$, where $U_\alpha = \{x \in \mathbb{R} : U(x) \geq \alpha\}$ for $\alpha \in (0,1]$, and $U_0 = \mathrm{cl}\,\{x \in \mathbb{R} : U(x) > 0\}$, cl denoting the topological closure.

The class $\mathcal{F}_c$ can be endowed with a semilinear structure, where addition and product by a scalar can be defined by means of Zadeh's extension principle ([28, 19]). On $\mathcal{F}_c$ we consider the d_∞ metric ([19]). The *magnitude* of $U \in \mathcal{F}_c$ is given by $\|U\| = d_\infty(U, \mathbf{1}_{\{0\}}) = d_H(U_0, \{0\})$.

Given a measurable space $(\Omega, \mathcal{A})$, a mapping $X : \Omega \to \mathcal{F}_c$ is said to be a *random upper semicontinuous function* (r.u.s.f. for short) if $X_\alpha : \Omega \to \mathcal{K}_c$ with $X_\alpha(\omega) = (X(\omega))_\alpha$ for all $\omega \in \Omega$, is a random set for all $\alpha \in [0,1]$ ([21, 4]).

A r.u.s.f. X is said to be *integrably bounded* with respect to a measure $\mu : \mathcal{A} \to \mathbb{R}$, if the mapping $\|X\| \in L^1(\Omega, \mathcal{A}, \mu)$, where $\|X\| : \Omega \to \mathbb{R}$ is given by $\|X\|(\omega) = \|X(\omega)\|$ for all $\omega \in \Omega$.

For an integrably bounded r.u.s.f., in [21] is defined its *integral*, denoted by $\int_\Omega X(\omega)\, d\mu(\omega)$ or $E(X|\mu)$, as the unique set in $\mathcal{F}_c$ such that $E(X|\mu)_\alpha = E\big(X_\alpha \big| \mu\big)$ for every $\alpha \in [0,1]$. When $\Omega = [a,b]$, we will use also the notation $\int_a^b X(\omega)\, d\mu(\omega)$.

If μ is a probability measure, an r.u.s.f. is also referred to as a *fuzzy random variable* and its integral as *the fuzzy expected value of* X.

It is possible to extend to upper semicontinuous functions the concept of *Hukuhara difference* or *Minkowski difference* for subsets ([15, 20]), so given $U, V \in \mathcal{F}_c$, its *Hukuhara difference*, denoted by $U -_h V$, is the set $W \in \mathcal{F}_c$ (if it exists) such that $U = V + W$.

Let T be a nonempty open subset of $\mathbb{R}$. A mapping $G : T \to \mathcal{F}_c$ is said to be *Hukuhara differentiable* at $t_0 \in T$ if there exists $G'(t_0) \in \mathcal{F}_c$, which is called the *Hukuhara differential of* G *at* t_0, such that

$$\lim_{\Delta t \to 0^+} d_\infty\left(\frac{G(t_0 + \Delta t) -_h G(t_0)}{\Delta t}, G'(t_0)\right)$$
$$= \lim_{\Delta t \to 0^+} d_\infty\left(\frac{G(t_0) -_h G(t_0 - \Delta t)}{\Delta t}, G'(t_0)\right) = 0.$$

The above definition ([20, 23]) is an extension of the Hukuhara's differentiability criterion for set-valued mappings ([15]).

ISBN: 978-989-95079-6-8

If a mapping G depends on more than one argument, we will make use of the usual symbol of partial derivative to indicate with respect to which argument the Hukuhara differential is considered.

Throughout the paper, for any set $\Omega \subset \mathbb{R}^k$ with $k \in \mathbb{N}$, $\mathcal{B}_\Omega$ will denote the Borel σ-field on Ω. Given $(\Omega, \mathcal{B}_\Omega)$ and $m_1, m_2 : \mathcal{B}_\Omega \to [0, \infty]$ two σ-finite measures, $m_1 \ll m_2$ will indicate that m_1 is absolutely continuous with respect to m_2, and $\frac{dm_1}{dm_2}$ will denote a Radon-Nikodym derivative of m_1 with respect to m_2. If it is supposed that there exists a continuous Radon-Nikodym derivative, then $\frac{dm_1}{dm_2}$ will denote this particular function.

As we will model the imprecise utilities by means of fuzzy-valued functions it should be necessary to rank fuzzy sets. For this purpose the ranking criterion introduced by De Campos and González [5] will be considered. Therefore, we will say that $U \in \mathcal{F}_c$ is *greater than or equal to* $W \in \mathcal{F}_c$ *in the* λ, μ*-average sense*, we will denote it by $U \geq_{\lambda,\mu} W$, if $V_\mu^\lambda(U) \geq V_\mu^\lambda(W)$, where $\lambda \in [0,1]$ represents a kind of degree of optimism/pessimism and μ is a measure on $[0,1]$ (see [18] for more details).

3 Theoretical results

We include in this section those theoretical results regarding the exchange of iterated integrals that are necessary to prove the applied results in Section 4.

The first result is about differentiating under the integral sign, by using Hukuhara derivative.

Proposition 3.1 *Let $(\Omega, \mathcal{A}, P)$ be a probability space with $\Omega \subset \mathbb{R}^k$, and let m denote the Borel measure on the interval $[a,b]$. For every $\omega \in \Omega$, let P_ω be a probability measure on $([a,b], \mathcal{B}_{[a,b]})$ with $P_\omega \ll m$, such that there exists a continuous Radon-Nikodym derivative.*

If $X : \Omega \times [a,b] \to \mathcal{F}_c$ satisfies the following conditions:

i) for every $\omega \in \Omega$, the mapping $X_\omega : [a,b] \to \mathcal{F}_c$, with $X_\omega(t) = X(\omega, t)$, is an integrably bounded r.u.s.f. with respect to P_ω, and X_ω is continuous $a.s.\,[P]$,

ii) there exists $h \in L^1(\Omega, \mathcal{A}, P)$, such that $\left\|X(\omega,t)\frac{dP_\omega}{dm}(t)\right\| \leq h(\omega)\ a.s.\,[P]$ for every $t \in [a,b]$, and the mapping $\omega \mapsto X(\omega,t)\frac{dP_\omega}{dm}(t)$ is continuous $a.e.\,[m]$,

iii) there exists $g \in L^1([a,b], \mathcal{B}_{[a,b]}, m)$ with $\left\|X(\omega,t)\frac{dP_\omega}{dm}(t)\right\| \leq g(t)\ a.e.\,[m]$ for every $\omega \in \Omega$,

then, the mapping

$$t \in [a,b] \mapsto \int_\Omega \Big(\int_a^t X(\omega,s)\, dP_\omega(s) \Big)\, dP(\omega)$$

is Hukuhara differentiable on (a,b), and for every $t \in (a,b)$ it holds that

$$\frac{\partial}{\partial t} \int_\Omega \left(\int_a^t X(\omega,s)\, dP_\omega(s) \right)\, dP(\omega)$$

$$= \int_\Omega X(\omega,t) \frac{dP_\omega}{dm}(t)\, dP(\omega).$$

Starting from last result, we can go one step further to prove a exchange of iterated integrals:

Theorem 3.2 *Let $(\Omega, \mathcal{B}_\Omega, P)$ be a probability space with $\Omega \subset \mathbb{R}^k$ and let m denote the Borel measure on the interval $T = [a,b]$. For every $t \in T$, let P_t be a probability measure on $(\Omega, \mathcal{B}_\Omega)$ such that $P_t \ll P$ and there exists a continuous Radon-Nikodym derivative. For every $\omega \in \Omega$, let P_ω be a probability on $(T, \mathcal{B}_T)$ such that $P_\omega \ll m$ and there exists a continuous Radon-Nikodym derivative.*

Let $X : \Omega \times T \to \mathcal{F}_c$ be a mapping satisfying that:

i) for every $t \in T$, X_t is an integrably bounded r.u.s.f. with respect to P_t,

ii) for every $\omega \in \Omega$, X_ω is an integrably bounded r.u.s.f. with respect to P_ω and it is continuous $a.s.\,[P]$,

iii) there exists $h_1 \in L^1(\Omega, \mathcal{B}_\Omega, P)$ such that $\left\|X(\omega,t)\frac{dP_\omega}{dm}(t)\right\| \leq h_1(\omega)\ a.s.\,[P]$ for every $t \in T$, and the mapping $\omega \mapsto X(\omega,t)\frac{dP_\omega}{dm}(t)$ is continuous $a.e.\,[m]$,

iv) there exists a mapping $g \in L^1([a,b], \mathcal{B}_{[a,b]}, m)$ such that for every $\omega \in \Omega$, $\|X(\omega,t)\frac{dP_\omega}{dm}(t)\| \leq g(t)\ a.e.\,[m]$ for every $\omega \in \Omega$,

v) the mapping $t \mapsto X(\omega,t)\frac{dP_t}{dP}(\omega)$ is continuous on T $a.s.\,[P]$,

vi) there exists $h_2 \in L^1(\Omega, \mathcal{B}_\Omega, P)$ such that $\|X(\omega,t)\frac{dP_t}{dP}(\omega)\| \leq h_2(\omega)\ a.s.\,[P]$ for every $t \in T$.

Let m' be a probability measure on $(T, \mathcal{B}_T)$ such that $m' \ll m$ and there exists a continuous Radon-Nikodym derivative. If for every $t \in T$, the equality

$$\frac{dP_\omega}{dm}(t) = \frac{dP_t}{dP}(\omega)\frac{dm'}{dm}(t) \quad a.s.\,[P]$$

holds, then

$$\int_\Omega \left(\int_a^t X(\omega,s)\, dP_\omega(s) \right) dP(\omega)$$

$$= \int_a^t \left(\int_\Omega X(\omega,s)\, dP_s(\omega) \right) dm'(s)$$

for every $t \in T$.

And also an unbounded version of previous theorem can be obtained.

Theorem 3.3 *Assume the conditions in Theorem 3.2 with the interval T being not necessarily bounded, and suppose that there exists $g' \in L^1(\Omega, \mathcal{B}_\Omega, P)$ such that*

$$\int_T \|X(\omega,s)\| dP_\omega(s) \leq g'(\omega)\ \ a.s.\,[P]\,.$$

Then, the following equality holds,

$$\int_\Omega \left(\int_T X(\omega,s)\, dP_\omega(s) \right) dP(\omega)$$

$$= \int_T \left(\int_\Omega X(\omega,s)\, dP_s(\omega) \right) dm'(s)\,.$$

It should be remarked that the conditions in Theorems 3.2 and 3.3 do not imply that X is an r.u.s.f. on the product measurable space as is illustrated in [24].

ISBN: 978-989-95079-6-8

4 Applied results

In this section we are using the theoretical results obtained in Section 3 to obtain applied results to the single-stage decision problem under non-product measurability conditions. Obviously, different conditions than in the Fubini-applicable case will be obtained, but it is again very important to underline that under the conditions stated here product measurability is not implied.

We first introduce the concept of fuzzy utility function considered in this communication. We will use the following notation: Θ is the state space and it will be considered an interval of $\mathbb{R}$, $\mathcal{B}_\Theta$ is the Borel σ-field on Θ, with m the Borel measure and $\mathbf{A}$ is the action space.

Definition 4.1 *A mapping $U : \Theta \times \mathbf{A} \to \mathcal{F}_c$ is said to be a fuzzy utility function on $\Theta \times \mathbf{A}$ if*

i) for every $a \in \mathbf{A}$, the projection $U_a : \Theta \to \mathcal{F}_c$ is an r.u.s.f. on $(\Theta, \mathcal{B}_\Theta)$,

ii) for every pair $a_1, a_2 \in \mathbf{A}$, a_1 will be considered preferred or indifferent to a_2 with respect to a probability distribution ξ on $(\Theta, \mathcal{B}_\Theta)$, if $E(U_{a_1}|\xi) \geq_{\lambda,\mu} E(U_{a_2}|\xi)$ (for fixed $\lambda \in [0,1]$ and measure μ).

The decision problem with fuzzy utilities will be denoted by $(\Theta, \mathbf{A}, U)$.

On the other hand, it will be considered a Bayesian context, so the existence of a probability distribution π on $(\Theta, \mathcal{B}_\Theta)$, *the prior distribution*, will be assumed. Then the "value" of the decision problem will be the fuzzy value $E(U_{a^\pi}|\pi)$, where a^π is a *prior Bayes action* in the λ, μ-average sense, this is, $a^\pi \in \mathbf{A}$ verifies $E(U_{a^\pi}|\pi) \geq_{\lambda,\mu} E(U_a|\pi)$ for all $a \in \mathbf{A}$.

Similar to the case of real-valued utilities, it is useful for increasing the expected utility to incorporate *sample information*. Let $\mathbf{X}$ be a statistical experiment characterized by a probability space $(\mathbb{X}, \mathcal{B}_\mathbb{X}, P_\theta)$, where $\theta \in \Theta$, $\mathcal{B}_\mathbb{X}$ is the Borel σ-field on $\mathbb{X} \subset \mathbb{R}^k$ and the experimental distribution P_θ depends on the true unknown state θ. We will denote by P the *marginal (predictive)* distribution of the experiment.

After the experiment is performed, if $\mathbf{X} = x$ is the available sample information, the fuzzy expected utility associated with an action $a \in \mathbf{A}$ is given by $E(U_a|\pi_x)$, where π_x is the *posterior distribution* of θ given $\mathbf{X} = x$, obtained on the basis of Bayes' formula. So, a *posterior Bayes action* is any $a^{\pi_x} \in \mathbf{A}$ such that $E(U_{a^{\pi_x}}|\pi_x) \geq_{\lambda,\mu} E(U_a|\pi_x)$ for every $a \in \mathbf{A}$.

In order to generalize the choice of an action for each possible sample, the concept of *decision rule,* as a mapping from $\mathbb{X}$ to $\mathbf{A}$ satisfying several conditions (based on Theorems 3.2 and 3.3) is formalized. These conditions will allow the proper extension of the normal and extensive forms of the Bayesian analysis.

Definition 4.2 *Let $(\mathbb{X}, \mathcal{B}_\mathbb{X}, P_\theta)$ be the probability space of a statistical experiment $\mathbf{X}$ associated with the decision problem $(\Theta, \mathbf{A}, U)$. A decision rule is a mapping $d : \mathbb{X} \to \mathbf{A}$ satisfying that*

i) for every $\theta \in \Theta$, $U(\theta, d()) : \mathbb{X} \to \mathcal{F}_c$ is an integrably bounded r.u.s.f. with respect to P_θ,

ii) for every $x \in \mathbb{X}$, $U(, d(x)) : \Theta \to \mathcal{F}_c$ is an integrably bounded r.u.s.f. with respect to π_x, moreover, it is continuous a.s. $[P]$,

iii) there exists $h_1 \in L^1(\mathbb{X}, \mathcal{B}_\mathbb{X}, P)$ such that $\left\|U(\theta, d(x))\frac{d\pi_x}{dm}(\theta)\right\| \leq h_1(x)$ a.s. $[P]$ for every $\theta \in \Theta$, and the mapping $x \mapsto U(\theta, d(x))\frac{d\pi_x}{dm}(\theta)$ is continuous a.e. $[m]$,

iv) there exists $g \in L^1(\Omega, \mathcal{B}_\Omega, m)$ such that for every $x \in \mathbb{X}$, it holds that $\|U(\theta, d(x))\frac{d\pi_x}{dm}(\theta)\| \leq g(\theta)$ a.e. $[m]$ for every $x \in \mathbb{X}$,

v) the mapping $\theta \mapsto U(\theta, d(x))\frac{dP_\theta}{dP}(x)$ is continuous on Θ a.s. $[P]$,

vi) there exists $h_2 \in L^1(\mathbb{X}, \mathcal{B}_\mathbb{X}, P)$ such that $\left\|U(\theta, d(x))\frac{dP_\theta}{dP}(x)\right\| \leq h_2(x)$ a.s. $[P]$ for every $\theta \in \Theta$,

vii) there exists $g' \in L^1(\mathbb{X}, \mathcal{B}_\mathbb{X}, P)$ with $\int_\Theta \|U(\theta, d(x))\| \, d\pi_x(\theta) \leq g'(x)$.

Now, on the one hand we can consider the *normal Bayesian analysis*. In this case we should find a *Bayes decision rule*, that is, a rule d_B such that

$$\int_\Theta \left(\int_\mathbb{X} U(\theta, d_B(x)) \, dP_\theta(x) \right) d\pi(\theta)$$

$$\geq_{\lambda,\mu} \int_\Theta \left(\int_\mathbb{X} U(\theta, d(x)) \, dP_\theta(x) \right) d\pi(\theta)$$

for every decision rule d. In this case, the "value" of the problem is

$$\int_\Theta \left(\int_\mathbb{X} U(\theta, d_B(x)) \, dP_\theta(x) \right) d\pi(\theta). \qquad (1)$$

On the other hand, we can consider the *extensive Bayesian analysis*. We should obtain for each sample information x a posterior Bayes action a^{π_x}, and consider the decision rule which associates with each x an action a^{π_x}. In this analysis, the "value" of the experiment $\mathbf{X}$ is quantified by the fuzzy expected terminal utility, defined as follows:

Definition 4.3 *Given $(\Theta, \mathbf{A}, U)$ a decision problem and $\mathbf{X} = (\mathbb{X}, \mathcal{B}_\mathbb{X}, P_\theta)$, an associated experiment, the fuzzy expected terminal utility of $\mathbf{X}$ is given by*

$$U_t(\mathbf{X}) = \int_\mathbb{X} \left(\int_\Theta U(\theta, a^{\pi_x}) \, d\pi_x(\theta) \right) dP(x). \qquad (2)$$

We are now showing the equivalence between the two forms of the Bayesian analysis in the sense that (1) and (2) are equal in the λ, μ-average sense. Previously, the following result for the exchange of the integrals is stated.

Theorem 4.4 *Let $(\Theta, \mathbf{A}, U)$ be a decision problem, let $\Theta \subset \mathbb{R}$ and let π be a prior probability on $(\Theta, \mathcal{B}_\Theta)$ such that $\pi \ll m$ with a continuous Radon-Nikodym derivative. Let $\mathbf{X} = (\mathbb{X}, \mathcal{B}_\mathbb{X}, P_\theta)$ be an associated experiment, and let P be the marginal distribution. For every $\theta \in \Theta$, suppose that $P_\theta \ll P$ and there exists a continuous Radon-Nikodym derivative. For every $x \in \mathbb{X}$, let π_x be the posterior distribution on $(\Theta, \mathcal{B}_\Theta)$ such that $\pi_x \ll m$ with a continuous Radon-Nikodym derivative.*

ISBN: 978-989-95079-6-8

If for every $\theta \in \Theta$*, it holds that* $\frac{d\pi_x}{dm}(\theta) = \frac{dP_\theta}{dP}(x)\frac{d\pi}{dm}(\theta)$ $a.s.\,[P]$*, then*

$$\int_{\mathbb{X}} \left(\int_{\Theta} U(\theta, d(x))\, d\pi_x(\theta) \right) dP(x)$$
$$= \int_{\Theta} \left(\int_{\mathbb{X}} U(\theta, d(x))\, dP_\theta(x) \right) d\pi(\theta)$$

whatever the decision rule $d : \mathbb{X} \to \mathbf{A}$ *may be.*

As a consequence, the following key result, which states the equivalence between the normal and extensive forms of Bayesian analysis, is obtained.

Theorem 4.5 *Assume the conditions of Theorem 4.4. Let us consider the mapping which associates with each sample* $x \in \mathbb{X}$ *a posterior Bayes action* a^{π_x}*. If this mapping satisfies the definition of decision rule, then it is a Bayes decision rule. Moreover,* $U_t(\mathbf{X})$ *is equal, in the* λ, μ*-average sense, to the fuzzy expected utility associated with any Bayes decision rule, this is*

$$U_t(\mathbf{X}) =_{\lambda,\mu} \int_{\mathbb{X}} \left(\int_{\Theta} U(\theta, d_B(x))\, d\pi_x(\theta) \right) dP(x)$$

Thus, the fuzzy expected terminal utility, so calculated, can be interpreted as the "value" of the decision problem once the experiment $\mathbb{X}$ is performed and one Bayes decision rule is calculated. This lead us not only to express the information of the problem by a value but also as a criterium to rank experiments in order to obtain the more informative.

5 Conclusions

By using a theoretical result about exchanging iterated integrals of $[0,1]^{\mathbb{R}}$-valued r.u.s.f., the model established in this paper provides a framework for single-stage decision problems in which both forms of Bayesian analysis (normal and extensive) are proved to be equivalent when imprecise utilities are not necessarily product measurable. Thus, these results together with those in [12, 17, 13] cover most of the situations which one can found when analyzing single-stage decision problems with imprecise utilities modeled by fuzzy random variables.

Acknowledgment

Authors want to acknowledge the financial support by Grant MTM2008-01519 from Ministry of Science and Innovation, Government of Spain.

References

[1] A. Billot. An existence theorem for fuzzy utility functions: A new elementary proof. *Fuzzy Sets and Systems*, 74 271-276, 1995

[2] F. Bordley. Reformulating decision theory using fuzzy set theory and Shafer's theory of evidence. *Fuzzy Sets and Systems*, 139: 243-266, 2003.

[3] C.B. Chen, C.M. Klein. A simple approach to ranking a group of aggregated fuzzy utilities. *IEEE Transactions on Systems, Man, and Cybernetics* 27: 26-35, 1997.

[4] A. Colubi, J.S. Domínguez-Menchero, M. López-Díaz, D.A. Ralescu. A $D_E[0,1]$ representation of random upper semicontinuous functions. *Proceedings of the American Mathematical Society*, 130: 3237-3242, 2002.

[5] L.M. De Campos, A. González. A subjective approach for ranking fuzzy numbers. *Fuzzy Sets and Systems*, 29: 145-153, 1989.

[6] G. Debreu. Integration of correspondences, *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability*, 1965/66, Vol. II: Contributions to Probability Theory, Part 1. University of California Press, Berkeley, pp. 351-372, 1967.

[7] D. Dubois, H. Prade. Additions of interactive fuzzy numbers. *IEEE Transactions on Automatic Control*, 26: 926-936, 1981.

[8] D. Dubois, H. Prade. *The use of fuzzy numbers in decision analysis. Fuzzy Information and Decision Processes*. North-Holland, Amsterdam, 309-321, 1982.

[9] D. Dubois, H. Prade. *Possibility Theory. An Approach to Computerized Processing of Uncertainty*. Plenum Press, New York, 1988.

[10] H. Friedman. A consistent Fubini-Tonelli theorem for nonmeasurable functions. *Illinois Jouranl of Mathematics*, 24: 390-395, 1980.

[11] M.A. Gil, P. Jain. Comparison of experiments in statistical decision problems with fuzzy utilities. *IEEE Transactions on Systems, Man, and Cybernetics* 22: 662-670, 1992.

[12] M.A. Gil, M. López-Díaz. Fundamentals and Bayesian analyses of decision problems with fuzzy-valued utilities. *International Journal of Approximate Reasoning*, 15: 203-224, 1996.

[13] M.A. Gil, M. López-Díaz, L.J. Rodríguez-Muñiz. An improvement of a comparison of experiments in statistical decision problems with fuzzy utilities. *IEEE Transactions on Systems, Man, and Cybernetics*, 28: 856-864, 1998.

[14] F. Hiai, H. Umegaki. Integrals, conditional expectations and martingales of multivalued functions. *Journal of Multivariate Analysis*, 7: 149-182, 1997

[15] M. Hukuhara. Intégration des applications mesurables dont la valeur est un compact convexe. *Funkcialaj Ekvacioj*, 10: 205-223, 1967.

[16] V. Krätschmer. Coherent lower previsions and Choquet integrals. *Fuzzy Sets and Systems*, 138: 469-484, 2003.

[17] M. López-Díaz, M.A. Gil. Reversing the order of integration in iterated expectations of fuzzy random variables, and statistical applications. *Journal of Statistical Planning and Inference*, 74: 11-29, 1998.

[18] M. López-Díaz, M.A. Gil. The λ-average value and the fuzzy expectation of a fuzzy random variable. *Fuzzy Sets and Systems*, 99: 347-352, 1998.

[19] M.L. Puri, D.A. Ralescu. Différentielle d'une fonction floue. *Comptes Rendus de l'Académie des Sciences. Série I. Mathématique* 293: 237-239, 1981.

[20] M.L. Puri, D.A. Ralescu. Differentials of fuzzy functions. *Journal of Mathematical Analysis and Applications* 91: 552-558, 1983.

[21] M.L. Puri, D.A. Ralescu. Fuzzy random variables. *Journal of Mathematical Analysis and Applications*, 114: 409-422, 1986.

[22] Y. Rébillé. Decision making over necessity measures through the Choquet integral criterion. *Fuzzy Sets and Systems* 157: 3025-3039, 2006.

ISBN: 978-989-95079-6-8

[23] L.J. Rodríguez-Muñiz, M. López-Díaz. Hukuhara derivative of the fuzzy expected value. *Fuzzy Sets and Systems*, 138: 593-600, 2003.

[24] L.J. Rodríguez-Muñiz, M. López-Díaz. On the exchange of iterated expectations of random upper semicontinuous functions. *Statistics and Probability Letters*, 77: 1628-1635, 2007.

[25] L.J. Rodríguez-Muñiz, M. López-Díaz. A new framework for the Bayesian analysis of single-stage decision problems with imprecise utilities. *Fuzzy Sets and Systems* 159: 3271-3280, 2008.

[26] R.M. Tong, P.P. Bonissone. A linguistic approach to decision-making with fuzzy sets. *IEEE Transactions on Systems, Man, and Cybernetics* 10: 716-723, 1980.

[27] S.R. Watson, J.J. Weiss, M.L. Donnell. Fuzzy decision analysis. *IEEE Transactions on Systems, Man, and Cybernetics*, 9: 1-9, 1979.

[28] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning. Parts I, II and III. *Information Science*, 8: 199-249; 8: 301-357; 9: 43-80, 1975.

ISBN: 978-989-95079-6-8

60 years "A Mathematical Theory of Communication" – Towards a "Fuzzy Information Theory"

Rudolf Seising

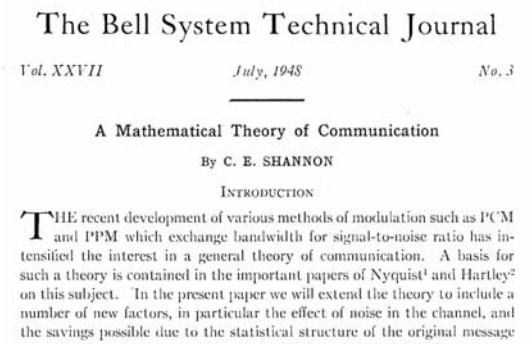
European Centre for Soft Computing
Mieres, Asturias, Spain
Email: rudolf.seising@softcomputing.es

Abstract*—When 60 years ago Shannon established "A Mathematical Theory of Communication" nobody could know the consequences for science and technology in the second half of the century. Shannon published his article in two parts in the July and October 1948 editions of the* Bell System Technical Journal*. However, it is very probable that this article wouldn't have become famous without the help of Weaver, whose popular text "The Mathematics of communication" re-interpreted Shannon's work for broader scientific audiences. Weaver's "preface" and Shannon's article were published together in the book* The Mathematical Theory of Communication *(1949) that represents the beginning of the then so-called "Information theory". However, in his "introduction" Weaver went over and above Shannon´s mathematical theory mentioning not only the technical but also the semantic and influential problems of communication. This classification is very similar to the foundations of the* Theory of Signs *(1938) that was established by Morris. This paper deals with the connectivity between this Information theory and the Theory of Fuzzy sets and systems that appeared in the first half of the 1950s.Then the paper focuses to the non-technical but philosophical aspects of information theory and it advocates a fuzzy information theory that has to be appropriate to cover the concept of information – particularly with regard to the philosophical aspects.*

***Keywords*—** communication, fuzziness, information, philosophy, semiotics, signs, signals

1 Introduction

60 years ago, in July 1948, the mathematician and electrical engineer Claude Elwood Shannon (1916-2001) published "A Mathematical Theory of Communication" in the *Bell System Technical Journal* [1] (Fig. 1) and in July of the following year "The Mathematics of Communication" by the mathematician and physicist Warren Weaver (1894-1978) appeared in the *Scientific American* [2] (Fig. 2).



The Bell System Technical Journal

Vol. XXVII July, 1948 No. 3

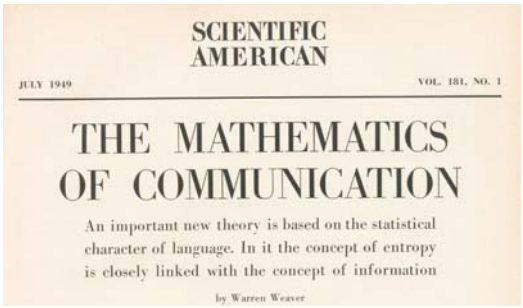
A Mathematical Theory of Communication

By C. E. SHANNON

INTRODUCTION

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist[1] and Hartley[2] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message

Figure 1: Title of Shannon's publication in July 1948 [1].



SCIENTIFIC AMERICAN

JULY 1949 VOL. 181, NO. 1

THE MATHEMATICS OF COMMUNICATION

An important new theory is based on the statistical character of language. In it the concept of entropy is closely linked with the concept of information

by Warren Weaver

Figure 2: Title of Weaver's publication in July 1949 [2].

However, other roots of the later so-called "Information Theory" can be found also in the *Cybernetics* of Norbert Wiener (1894-1964) [3], in the work of the Russian mathematician Andrei N. Kolmogorov (1903-1987) [4] and of the British statistician Ronald A. Fisher (1890-1962) [5].
Wiener wrote: "This idea occurred at about the same time to several writers, among them the statistician R. A. Fisher, Dr. Shannon of the Bell Telephone Laboratories, and the author. Fisher's motive in studying this subject is to be found in classical statistical theory; that of Shannon in the problem of coding information; and that of the author in the problem of noise and message in electrical filters. Let it be remarked parenthetically that some of my speculations in this direction attach themselves to the earlier work of Kolmogorov in Russia, although a considerable part of my work was done before my attention was called to the work of the Russian school." ([3], p. 18.)[1]
Shannon's "Mathematical Theory of Communication" and Wiener's *Cybernetics* had appeared simultaneously, but in both cases the publication was delayed by the war. However, Wiener mentioned the fact that, since Shannon was a Bell employee, his research projects were geared toward realizing and marketing his findings as quickly as possible, whereas Wiener, as a college professor, was able to approach his research freely. He had "found the new realm of communication ideas a fertile source of new concepts not only in communication theory, but in the study of the living organism and in many related problems".[2] In the manuscript

[1] Wiener cited Kolmogorov's article [4].
[2] Wiener in conversation with Bello (technology editor of the journal *Fortune*) on 13 October 1953, Box 4.179, MC22, Wiener Papers, MIT. Quoted from [6], p. 138.

ISBN: 978-989-95079-6-8

for his book *Invention: The Care and Feeding of Ideas*, which he never finished, Wiener wrote in 1957:
"Once I had alerted myself and the public in general to the statistical element in communication theory, conformation began to flow in from all sides. At the Bell Telephone Laboratories there was, and is, a young mathematical physicist by the name of Claude Shannon. He had already applied mathematical logic to the design of switching systems, and throughout all his work he has shown a love for the discrete problems, the problems with a small number of variable quantities, which come up in switching theory. I am inclined to believe that from the very start, a large part of his ideas in communication theory and its statistical basis were independent of mine, but whether they were or not, each of us appreciated the significance of the work of the other. The whole subject of communication began to assume a new statistical form, both for his sorts of problems and for mine. This is not the point for me to give the genealogy of every single piece of apparatus which this new statistical communication theory has fostered, but I can say that the impact of the work has gone from one end of communication theory to the other, until now there is scarcely a recent communication invention which has not been touched by statistical considerations. Thus, this whole wide spreading branch of science represents a subtle working out of concepts implicit in Gibbs and in the Lebesque-Borel team, but if I may say so, implicitly implicit, so that until some forty years had passed, no one could have seen the direction in which the earlier thought was bound to lead. This is, in my mind, a key example of a change in intellectual climate, and of the effect it has had both in discovery and in invention." ([7], p. 22f)
Today Shannon's name is associated almost unanimously with mathematical information and communication theory – these terms are often considered synonyms – although people had spoken at first of the "Wiener-Shannon Communication Theory".



Figure 3: C. E. Shannon, W. Weaver and N. Wiener.

However, it was not least Warren Weaver's article that popularized Shannon's theory and therefore it became – with little changes and the new title "Recent Contributions to the Mathematical Theory of Communication" – as an "introduction" to Shannon's text in their joint book-publication *The Mathematical Theory of Communication* [8]. Even more Weaver's manuscript was not only a brief and understandable sketch of Shannon's theory and their technical problems – in this text Weaver presented self-contained and more philosophical considerations on the semantic and influential problems with the concept of information that is in the core of Shannon's theory and that was called in the following years "Information theory". We will deal with these philosophical subjects in section 3. Beforehand we will give a few remarks on the genesis of the theory of fuzzy sets and systems, especially against the background of the developments of information theory in the 1950s.

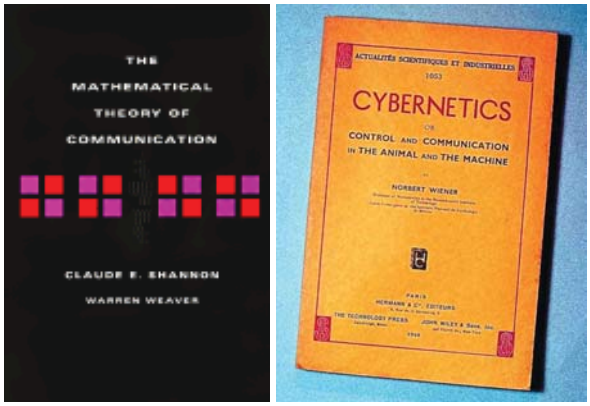


Figure 4: The books of Shannon and Weaver and Wiener.

2 Information theory and the theory of Fuzzy Sets and Systems

2.1 A Historical Sketch - in General

Inspired by Wiener's *Cybernetics*, Shannon's and Weaver's *Mathematical Theory of Communication* and the digital computer era that started during the wartime with the *Electronic Numerical Integrator and Computer* (ENIAC) and the *Electronic Discrete Variable Computer* (EDVAC), both designed by John P. Eckert (1919-1995) and John W. Mauchly (1907-1980), the young emigrant Lotfi A. Zadeh continued his studies in electrical engineering at MIT in Boston after his emigration from Iran into the USA in the 1940s. When he started his doctoral studies at the Columbia University in 1946, he became acquainted with these new milestones in science and technology. Shannon and Wiener delivered lectures in New York about the new theories they had developed during the War. In 1950 Zadeh acted as a moderator at a debate on digital computers at Columbia University, held between Shannon, Edmund Callis Berkeley (1923-1988), the author of the book *Giant Brains or Machines That Think* published in 1949 [9], and Francis Joseph Murray (1911-1996), mathematician and consultant to IBM.
15 years later Zadeh, who was then a professor of electrical engineering at Berkeley, established the new mathematical theory of Fuzzy Sets and Systems [10-12]. Already in 1962 he described the basic necessity of a new scientific tool to handle very large and complex systems in the real world: ''we need a radically different kind of mathematics, the mathematics of fuzzy or cloudy quantities which are not describable in terms of probability distributions. Indeed, the

ISBN: 978-989-95079-6-8

need for such mathematics is becoming increasingly apparent even in the realm of inanimate systems, for in most practical cases the a priori data as well as the criteria by which the performance of a man-made system are judged are far from being precisely specified or having accurately-known probability distributions'' [13].

The potential of the new techniques of the theory of Fuzzy sets and Fuzzy systems urged Ebrahim H. Mamdani, a professor of electrical engineering at Queen Mary College in London, to attempt the implementation of a fuzzy system under laboratory conditions. He expressed the intention to his doctor student Sedrak Assilian, who designed a fuzzy algorithm to control a small steam engine within a few days. The concepts of so-called linguistic variables and Zadeh's max-min composition were suitable to establish fuzzy control rules because *input*, *output* and *state* of the steam engine system range over fuzzy sets. Thus, Assilian and Mamdani designed the first real fuzzy application when they controlled the system by a fuzzy rule base system [14]. In 1974, Assilian completed his Ph. D. thesis on this first fuzzy control system [15].

The steam engine heralded the Fuzzy boom that started in the 1980s in Japan and later pervaded the Western hemisphere. Many fuzzy applications, such as domestic appliances, cameras and other devices appeared in the last two decades of the 20th century. Of greater significance, however, was the development of fuzzy process controllers and fuzzy expert systems that served as trailblazers for scientific and technological advancements of fuzzy sets and systems.[3]

Very little is known about the historical connectivity between the Information theory and the theory of Fuzzy Sets and Systems. In the following paragraph we review some links across the boundaries of these fields.

2.2 *In detail: Signals, Noise and Uncertainty*

Shannon intended to establish a general theory of communication and a basis for such a theory was already given in two former papers of other Bell-engineers, Harry Nyquist and Ralph V. L. Hartley [17, 18].[4] 20 years later, Shannon extended this theoretical basis. He included "new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information." ([1], p. 379) To illustrate this mathematical theory, Shannon had drawn a diagram of a general communication model (Fig. 5):

- An information source produces a series of messages that are to be delivered to the receiver side. Transmission can occur via a telegraph or teletypewriter system, in which case it is a series of letters. Transmission can also occur via a telephone or radio system, in which case it is a function of time $f(t)$, or else it is a function $f(x, y, t)$, such as in a black and white television system, or it consists of complicated functions.
- The transmitter transforms the message in some way so that it can produce signals that it can transmit via the channel. In telegraphy, these are dotdash codes; in telephony, acoustic pressure is converted into an electric current.
- The channel is the medium used for transmission. Channels can be wires, light beams and other options.
- The receiver must perform the opposite operation to that of the transmitter and in this way reconstructs the original message from the transmitted signal.
- The destination is the person or entity that should receive the message.

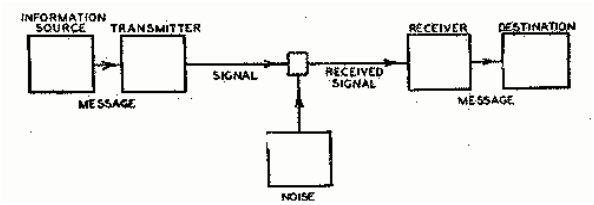


Figure 5: Shannon's diagram ([1], p. 381).

According to this, Shannon conceived of communication purely as the transmission of messages – completely detached from the meaning of the symbols!

In the second paragraph of his paper Shannon states: "The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point." ([1], p. 379) Eventually, in chapter 12 that is entitled "Equivocation and Channel Capacity", he applies his theoretical framework to this problem: "If the channel is noisy it is not in general possible to reconstruct the original message or the transmitted signal with *certainty* by any operation on the received signal *E*. There are, however, ways of transmitting the information which are optimal in combating noise." His solution of the problem is as follows: "We consider a communication system and an observer (or auxiliary device) who can see both what is sent and what is recovered (with errors due to noise). This observer notes the errors in the recovered message and transmits data to the receiving point over a "correction channel" to enable the receiver to correct the errors." ([1], p. 408) Shannon also indicated this situation schematically (Fig. 6).

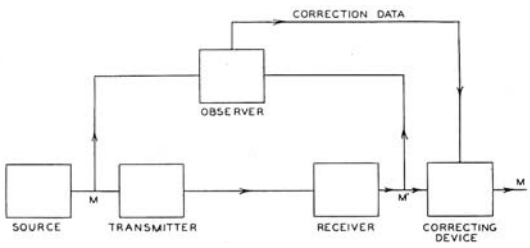


Figure 6: Shannon's diagram of a communication system with correction device ([1], p. 409).

[3] A more detailed presentation of the history of the theory of Fuzzy sets and systems can be found in [16].

[4] A more detailed presentation of the aspect of these papers is given in my paper for the 2009 annual meeting of NAFIPS [19].

ISBN: 978-989-95079-6-8

Shannon's introduction of the observer is of great theoretical interest but – of course – the system has to transmit the data on errors and the correction data via communication channels to the observer. When Warren Weaver explained the mathematics of Communication in the first version of his paper [2], he wrote: "If noise is introduced, then the received message contains certain distortions, certain errors. Certain extraneous material, that would certainly lead to increased uncertainty. […] To get the useful information in the received signal we must subtract the spurious portion. This is accomplished, in the theory, by establishing a quantity known as the "equivocation", meaning the amount of ambiguity introduced by noise." ([2], p. 13) And some paragraphs later he emphasized: "However clever one is with the coding process, it will always be true that after the signal is received there remains some undesirable uncertainty about what the message was". ([2], p. 13)

Let's see how Lotfi Zadeh dealt with this fundamental problem of communication in the 1950s!

In 1949, the year when the book of Shannon and Weaver appeared, Zadeh wrote the Ph. D. thesis on *Frequency Analysis of Variable Networks*, under supervision of Professor John Ralph Ragazzini, and in 1950 he was appointed to an assistant professor. He was enhanced by Wiener's cybernetics and Shannon's information theory. In February 1952, he presented *Some Basic Problems in Communication of Information* at the meeting of the Section of Mathematics and Engineering of the New York Academy of Sciences and in the following, one of these problems will be sketched, that deals with the *recovery process* of transmitted signals. In the proceedings of this meeting, Zadeh wrote: "Let $X=\{x(t)\}$ be a set of signals. An arbitrarily selected member of this set, say $x(t)$, is transmitted through a noisy channel Γ and is received as $y(t)$. As a result of the noise and distortion introduced by Γ, the received signal $y(t)$ is, in general, quite different from $x(t)$. Nevertheless, under certain conditions it is possible to recover $x(t)$ – or rather a time-delayed replica of it – from the received signal $y(t)$." ([20], p. 201)

In this paper, he didn't examine the case where $\{x(t)\}$ is an ensemble; he restricted his view to the problem to recover $x(t)$ from $y(t)$ "irrespective of the statistical character of $\{x(t)\}$"[5, p. 201]. Corresponding to the relation $y = \Gamma x$ between $x(t)$ and $y(t)$ he represented the recovery process of $y(t)$ from $x(t)$ by $x = \Gamma^{-1}y$, where $\Gamma^{-1}y$ is the inverse of Γ, if it exists, over $\{y(t)\}$.

Zadeh represented signals as ordered pairs of points in a signal space Σ, which is imbedded in a function space with a delta-function basis, and to measure the disparity between $x(t)$ and $y(t)$ he attached a distance function $d(x, y)$ with the usual properties of a metric. Then he considered the special case in which it is possible to achieve a perfect recovery of the transmitted signal $x(t)$ from the received signal $y(t)$. He supposed that "$X = \{x(t)\}$ consist of a finite number of discrete signals $x_1(t)$, $x_2(t)$, …, $x_n(t)$, which play the roles of symbols or sequences of symbols. The replicas of all these signals are assumed to be available at the receiving end of the system. Suppose that a transmitted signal x_k is received as y. To recover the transmitted signal from y, the receiver evaluates the 'distance' between y and all possible transmitted signals x_1, x_2, …, x_n, by the use of a suitable distance function $d(x, y)$, and then selects that signal which is 'nearest' to y in terms of this distance function (Fig. 7). In other words, the transmitted signal is taken to be the one that results in the smallest value of $d(x, y)$. This in brief, is the basis of the reception process." ([20], p. 201) By this process the received signal x_k is always 'nearer' – in terms of the distance functional $d(x, y)$ – to the transmitted signal $y(t)$ than to any other possible signal x_i, i.e. $d(x_k, y) < d(x_i, y)$, $i \neq k$,*for all k and i.*

But at the end of his reflection of this problem Zadeh conceded that "in many practical situations it is inconvenient, or even impossible, to define a quantitative measure, such as a distance function, of the disparity between two signals. In such cases we may use instead the concept of neighborhood, which is basic to the theory of topological spaces" ([20], p. 202). Spaces such as these, Zadeh surmised, could be very interesting with respect to applications in communication engineering. Therefore, this problem of the recovery process of transmitted signals which is a special case of Shannon's *fundamental problem of communication*, the problem "of reproducing at one point either exactly or approximately a message selected at another point", was one of the problems that initiated Zadeh's thoughts about not precisely specified quantitative measures, i. e. or cloudy or fuzzy quantities. – About 15 years later he proposed his new 'concept of neighborhood' which is now basic to the theory of fuzzy systems!

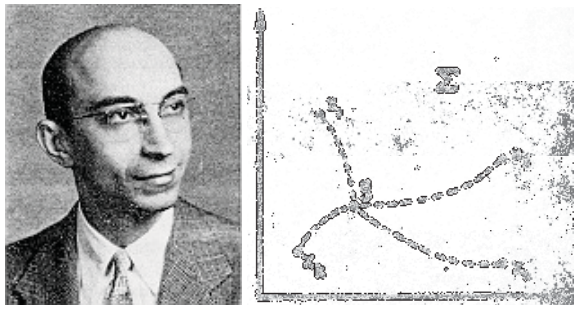


Figure 7: Lotfi A. Zadeh (born 1921) and his illustration: "Recovery of the input signal by means of the comparison of the distances between the received signal y and all possible transmitted signals" [20].

3 Fuzziness and Information Theory

In this section we concentrate our attention from the problem of communication to some philosophical considerations in the first half of the 20th century by Charles William Morris (1901-1976) and Warren Weaver. However, Morris was an engineer by training and he did his Ph. D. in philosophy under George Herbert Mead (1863-1931), the founder of social psychology, and Weaver was a mathematician and physicist. Nevertheless, both scientists wrote philosophical

ISBN: 978-989-95079-6-8

works on the subjects of signs and their transmission. Moreover, there is a distinct similarity between the two theoretical works that has to be underlined.

3.1 Theory of Signs, semiotics, and Information theory

The study of sign processes, or signification and communication is called "semiotics" since the 1930s. Great work to formalize this field was done by members of the Vienna Circle but also by the philosophers and linguists Charles Sanders Peirce (1839–1914), Ferdinand de Saussure (1857–1913) and Louis Hjelmslev (1899–1965), but here we will limit our considerations to the fundamental work of semiotics by Morris.

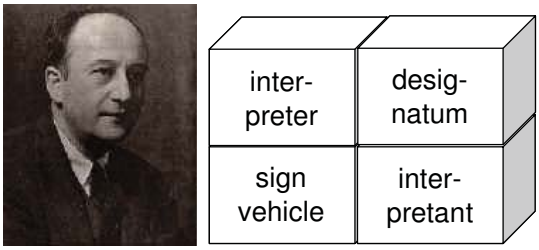


Figure 8: Ch. W. Morris; an illustration of the four components of the semiosis process.

Morris was in touch with some members of the Vienna Circle and he was a member of the Unity of Science movement. When he intended in his 1938 published *Foundations of the Theory of Signs* [21] a science of signs "on a biological basis and specifically with the framework of the science of behavior", he defined *semiotics* as a *universal* theory of signs and an *interdisciplinary* undertaking. In his view, the mission of semiotics as a science of signs is analyzing language as a social system of signs. Language is a system of signs which produces dispositions to social behavior, and in order to understand the uses and effects of signs we have to understand that and how signs influence social behavior. The process by which a sign-vehicle may function as a sign is called *semiosis*.[5] In Morris' foundation there are four components of the semiosis (Fig. 8):

1) the *sign vehicle* − this is the object or event which functions as a sign,
2) the *designatum* − this is the kind of object or class of objects which the sign designates,
3) the *interpretant* − this is the disposition of an interpreter to initiate a response-sequence as a result of perceiving the sign, and
4) the *interpreter* − this is the person for whom the sign-vehicle functions as a sign.

He also divided *semiotics* into three interrelated disciplines:

1) syntactics − the study of the methods by which signs may be combined to form compound signs,
2) semantics − the study of the signification of signs,
3) pragmatics − the study of the origins, uses, and effects of signs.

3.2 Weaver's three levels in Information theory

It seems that Warren Weaver was familiar with Morris' 10 years old classification of the semiosis when he wrote his paper on Shannon's Mathematical theory of communication [2]. Already in the third paragraph of his paper he wrote "In communication there seem to be problems at three levels: 1) technical, 2) semantic, and 3) influential. The technical problems are concerned with the accuracy of transference of information from sender to receiver. They are inherent in all forms of communication, whether by sets of discrete symbols (written speech), or by a varying two-dimensional pattern (television). The semantic problems are concerned with the interpretation of meaning by the receiver, as compared with the intended meaning of the sender. This is a very deep and involved situation, even when one deals only with the relatively simple problems of communicating through speech. […] The problems of influence or effectiveness are concerned with the success with which the meaning conveyed to the receiver leads to the desired conduct on his part. It may seem at the first glance undesirable narrow to imply that the purpose of all communication is to influence the conduct of the receiver. But with any reasonably broad definition of conduct, it is clear that communication either affects conduct or is without any discernible and provable effect at all." ([2], p. 11)

In the revised version of the paper that was published in [8], Weaver explained the trichotomy of the communication problem in extenso and he divided it into three levels:

- **Level *A*** contains the purely technical problem involving the exactness with which the symbols can be transmitted,
- **Level *B*** contains the semantic problem that inquires as to the precision with which the transmitted signal transports the desired meaning,
- **Level *C*** contains the pragmatic problem pertaining to the effect of the symbol on the destination side: What influence does it exert?

He underscored very clearly the fact that Shannon's theory did not even touch upon any of the problems contained in levels *B* and *C*, that the concept of information therefore must not be identified with the "meaning" of the symbols: "In fact, two messages, one of which is heavily loaded with meaning and the other of which is pure nonsense, can be exactly equivalent, from the present viewpoint, as regards information." [8] However, there is plenty of room for fuzziness in the levels *B* and *C*. The interpretation of meaning of signs, e. g. linguistic signs, names, words, is obviously a fuzzy process, and influence or effectiveness the exerted to the receiver's side is a fuzzy process, too. We will have this fuzziness at the back of our mind following Weaver's continuing considerations.

[5] The term *semiosis* was introduced by Ch. S. Peirce to describe "a process that interprets signs as referring to their objects" [22].

ISBN: 978-989-95079-6-8

3.3 Fuzziness in the diagram of a communication system

It is quite plain: Weaver went over and above Shannon´s theory: "The theory goes further. Though ostensibly applicable only to problems at the technical level, it is helpful and suggestive at the levels of semantics and effectiveness as well." Weaver stated, that Shannon's formal diagram of a communication system (Fig. 5) "can, in all likelihood, be extended to include the central issues of meaning and effectiveness. […] One can imagine, as an addition to the diagram, another box labeled "Semantic Receiver" interposed between the engineering receiver (which changes signals to messages) and the destination. This semantic receiver subjects the message to a second decoding the demand on this one being that it must match the statistical semantic characteristics of the message to the statistical semantic capacities of the totality of receivers, or of that subset of receivers which constitutes the audience one wishes to affect.

Similarly one can imagine another box in the diagram which inserted between the information source and the transmitter, would be labeled "Semantic Noise" (not to be confused with "engineering noise". This would represent distortions of meaning introduced by the information source, such as a speaker, which are not intentional but nevertheless affect the destination, or listener. And the problem of semantic decoding must take this semantic noise into account. It is also possible to think of a treatment or adjustment of the original message that would make the sum of message meaning plus semantic noise equal to the desired total message meaning at the destination." ([2], p. 13)

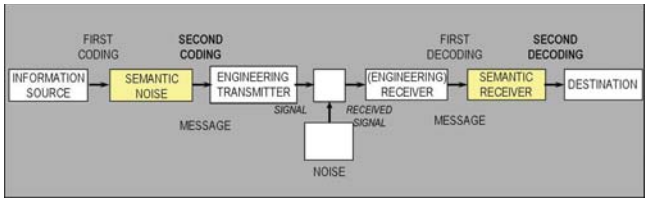


Figure 9: Shannon's diagram with Weaver's addition of the two boxes "Semantic Receiver" and "Semantic Noise".

Fig. 9 shows Weaver's two additional "fuzzy" boxes in Shannon's diagram of a communication system. We will interpret the "first coding" between the information source and the "Semantic Noise" as a *fuzzification* and the "second decoding" between the "Semantic Receiver" and the destination as a *defuzzification*.

4 Outlook

In future work we will proceed with this generalization of Weaver's philosophical considerations on Shannon's Mathematical theory of communication to a "Fuzzy information theory".

References

[1] C. E. Shannon, A Mathematical Theory of Communication, *The Bell System Technical Journal*, 27, pp. 379-423 & 623-656, 1948.

[2] W. Weaver, The mathematics of communication, *Scientific American*, 181: 11-15, 1949.

[3] N. Wiener, *Cybernetics or Control and Communications in the Animal and the Machine*. Cambridge, Massachusetts: MIT Press 1948.

[4] A. N. Kolmogorov, Interpolation and Extrapolation. In: *Bulletin de l'Académie des Sciences de l'URSS*, Serie Mathématique, Vol. 5, pp. 3-14, 1941.

[5] R. A. Fisher, *Statistical Methods for Research Workers.* Edinburgh: Oliver and Boyd, 1925.

[6] L. E. Kay, *Who Wrote the Book of Life?: A History of the Genetic Code*. Stanford, Calif.: Stanford University Press 2000.

[7] N. Wiener, *Invention. The Care and Feeding of Ideas*, London: The MIT Press, 1993.

[8] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Urbana, Illinois: University of Illinois Press, 1949.

[9] E. C. Berkeley, *Giant Brains or Machines that Think*. New York, London: John Wiley & Sons, Chapman & Hall, 1949.

[10] L. A. Zadeh, Fuzzy Sets, *Information and Control*, 8, 1965, pp. 338-353.

[11] L. A. Zadeh, Fuzzy Sets and Systems. In: Fox, J. (ed.): *System Theory*. Microwave Research Institute Symposia Series XV, Brooklyn, New York: Polytechnic Press 1965, p. 29-37.

[12] R. E. Bellman, R. Kalaba, L. A. Zadeh, Abstraction and Pattern Classification. *Journal of Mathematical Analysis and Applications* 13 1966, pp. 1-7.

[13] L. A. Zadeh, From Circuit Theory to System Theory. *Proceedings of the IRE*, Vol. 50, 1962, pp. 856-865.

[14] E. H. Mamdani, S. Assilian: An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller. In: *International Journal of Man-Machine Studies*, Vol. 7, No. 1, pp. 1-13, 1975.

[15] S. Assilian, *Artificial Intelligence in the Control of Real Dynamic Systems*. Ph.D. Thesis, No. DX193553, University London, August 1974.

[16] R. Seising, *The Fuzzification of Systems. The Genesis of Fuzzy Set Theory and Its Initial Applications – Developments up to the 1970s* (Studies in Fuzziness and Soft Computing Vol. 216) Berlin, New York, [et al.]: Springer 2007.

[17] H. Nyquist, Certain Factors Affecting Telegaph Speed, *The Bell System Technical Journal* 3, 1924, pp. 324-346.

[18] R. V. L. Hartley, Transmission of information, *The Bell System Technical Journal* 7, 1928, pp. 535-563.

[19] R. Seising, *On two 60 years old theories and the Theory of Fuzzy Sets and Systems: Cybernetics and Information Theory*. Proceedings of the 2009 Annual Meeting of the North American Fuzzy Information Society – NAFIPS-09 Annual Conference, University of Cincinnati, Ohio, USA, June 14-17, 2009 (to appear)

[20] L A. Zadeh, Some Basic Problems in Communication of Information *The New York Academy of Sciences,* Series II, Vol. 14, No. 5, 1952, pp. 201-204

[21] Ch. W. Morris, *Foundations of the Theory of Signs,* (International Encyclopedia of Unified Science, ed. Otto Neurath, vol. 1 no. 2.) Chicago: University of Chicago Press, 1938. Rpt, Chicago: University of Chicago Press, 1970-71. Reprinted in Charles Morris, *Writings on the General Theory of Signs* (The Hague: Mouton, 1971), pp. 13-71.

[22] N. Houser and Chr. Kloesel (eds.), *The Essential Peirce. Selected Philosophical Writings*, Vol. 2 (1893-1913), Peirce Edition Project, Bloomington/Indianapolis 1998 pp. 411.

ISBN: 978-989-95079-6-8

On classical, fuzzy classical, quantum, and fuzzy quantum systems

Rudolf Seising

European Centre for Soft Computing
Mieres, Asturias, Spain
Email: rudolf.seising@softcomputing.es

***Abstract*—*In this paper we consider physical systems and the concept of their states in the context of the theory of fuzzy sets and systems. In section 1 we give a brief sketch on the fundamental difference between the theories of classical physics and quantum mechanics. In section 2 and 3 we introduce very shortly systems and their states in classical and quantum mechanics, respectively. Section 4 presents the concept of fuzzy systems. We propose to fuzzify the classical systems in section 5 and quantum systems in section 6. In section 7 we start to consider a fuzzy interpretation of the uncertainty principle.*

Keywords— Fuzzy sets and systems, quantum mechanics, system theory, philosophy.

1 Introduction

Due to the scientific revolution brought about by the discovery of quantum mechanics in the first third of the 20th century, a basic change took place in the relationship between the exact scientific theory of physics and the phenomena observed in basic experiments. Systems of quantum mechanics – "quantum systems" – do not behave like systems of classical theories in physics – their elements are not particles and they are not waves, they are different. This change led to a new mathematical conceptual fundament in physics.

Niels Bohr, Max Born, Louis de Broglie, Paul A. M. Dirac, Werner Heisenberg, Pacual Jordan, Wolfgang Pauli, Erwin Schrödinger, John von Neumann and others introduced new objects and theoretical terms to the new mathematical theory of atomic physics and rather quantum physics, then so-called "quantum mechanics" that differ significantly from those in classical physics. Their properties are completely new and not comparable with those of observables in classical theories such as Newton's mechanics or Maxwell's electrodynamics.

The new theoretical term is the quantum mechanical state function ψ that is an element of the abstract Hilbert space H. The theory of quantum mechanics is completely abstract: it is a theory of mathematical state functions that have no exact counterpart in reality. This means that per se ψ is not observable but, nonetheless, we can experiment with a quantum mechanical object having a state function in order to measure its position value, and we can also experiment with this object in order to measure its momentum value. However, we cannot conduct both experiments simultaneously and thus are not able to get both values for the same point in time. But we can predict these values as outcomes of experiments at this point in time. Since predictions are targeted on future events, we cannot valuate them with the logical values "true" or "false," but must use probabilities. The probability distribution to measure a certain position, e.g. value $\boldsymbol{r}$ (t) = (x (t), y (t), z (t)) at point t in time is given by $|\psi(\boldsymbol{r}, t)|^2$ and the probability distribution to measure a certain momentum value $\boldsymbol{p}$ at time point t is given by $|\psi(\boldsymbol{p}, t)|^2$, where $\psi(\boldsymbol{x}, t)$ or $\psi(\boldsymbol{p}, t)$, are representtations of the abstract Hilbert space element ψ in the position or momentum representation respectively.

2 Classical Systems

In classical physics, the state of a "system" or an "object" is represented by a set of observables. For example, in Newtonian mechanics the state of an object (a particle with mass m) is given by the pair of 3-component vector values of the object's position vector $\boldsymbol{r}$ and its momentum vector $\boldsymbol{p}$. These two vectors imply all other properties of the object that are relevant in the Newtonian theory of mechanics. We can formulate that the state of a physical object is the collection of all the object's properties P_i. In order to represent these properties P_i in terms of the physical theory, we must determine the formally possible functions F_i in this mechanical theory, and in order to know the object's properties at a given point in time t, we must measure the values of these functions F_i. Thus, the representation of the "state of a classical object" is related to the measurement process or the perception process of the observer.

Due to the possible errors of measurement and the systematic errors occurring in every experiment, we can attribute their probability of this being the real value to all measured values of observables. Thus, the state of an object in Newtonian mechanics is given by the pair of the probability distributions of position $\boldsymbol{r}$ and momentum $\boldsymbol{p}$.

3 Quantum Systems

3.1 The concept of state in quantum mechanics

The state of a quantum mechanical system is much more difficult to determine than that of classical systems as we cannot measure sharp values for both variables simultaneously. This is the meaning of Heisenberg's uncertainty principle.

However, we can experiment with quantum mechanical objects in order to measure a position value, and we can also experiment with these systems in order to measure their momentum value. But: we cannot conduct both experiments

ISBN: 978-989-95079-6-8

simultaneously and thus are not able to get both values for the same point in time respectively. We can predict these values as outcomes of experiments at this point in time. Since predictions are targeted on future events, we cannot valuate them with logical values "true" or "false", but with probabilities.
Accordingly, in quantum mechanics, we have to use a modified concept of the state: the state of a quantum mechanical system consists of the probability distributions of all the object's properties that are formally possible in this physical theory.

- Max Born ([1, 2]) proposed an interpretation of this non-classical peculiarity of quantum mechanics – the quantum mechanical wave function is a "probability-amplitude": The absolute square of its value equals the probability of it having a certain position or a certain momentum if we measure the position or momentum respectively. The higher the probability of the position value, the lesser that of the momentum value and vice versa.
- In 1932, John von Neumann published the *Mathematical Foundations of Quantum Mechanics* [3], in which he defined the quantum mechanical wave function as a one-dimensional subspace of an abstract Hilbert space, which is defined as the state function of a quantum mechanical system (or object). Its absolute square equals the probability density function of it having a certain position or a certain momentum in the position or momentum representation of the wave function respectively.

Unfortunately there is no joint probability distribution for events in which both variables have a certain value simultaneously, as there is no classical probability space that comprises these events. Such pairs would describe classical states. Thus, the quantum mechanical system's state function embodies the probabilities of all properties of the object, but it delivers no joint probability distribution for all these properties. Therefore we claim here: "We need a radically different kind of mathematics, the mathematics of […] quantities which are not describable in terms of [classical] probability distributions.[1]

3.2 Quantum logic and Quantum probability theory

After the establishment of quantum mechanics there appeared some approaches to achieve a new logic and later also a new probability theory to handle quantum mechanical propositions and quantum mechanical events.

- 1936, Garett Birkhoff and John von Neumann proposed the introduction of a "quantum logic", as the lattice of quantum mechanical propositions is not distributive, and therefore not Boolean [5].
- In 1963, George Whitelaw Mackey attempted to provide a set of axioms for the propositional system of predictions of experiment's outcomes. He was able to show that this system is an orthocomplemented partially ordered set. [6]

In these logico-algebraic approaches, the "probabilities" of evaluating the predictions of the properties of a quantum mechanical system do not satisfy Kolmogorov's well-known axioms. The double-slit experiment shows that they are not additive and together with their non-distributivity it is indicated that the probabilistic structure of quantum mechanics is more complicated than that of the classical probability space as it was defined by Kolmogorov.

- Already in the 1960s, the philosopher and statistician Patrick Suppes discussed the "probabilistic argument for a non-classical logic of quantum mechanics" [7, 8]. He introduced the concept of a "quantum mechanical σ-field" as an "orthomodular partial ordered set" covering the classical σ-fields as substructures.
- In the 1980s, a "quantum probability theory" was proposed and developed by Stanley Gudder and Imre Pitowski [9, 10].[2]

The quantum mechanical lattice of predictions is Suppes' "quantum mechanical σ-field", which can be restricted to a Boolean lattice corresponding to a given observable. The quantum probabilities became classical probabilities again, only applying to predictions of compatible observables.
The theory of Fuzzy sets and systems pertains to "quantities which are not describable in terms of probability distributions" and theory of Quantum mechanics pertains to quantities which are not describable in terms of classical probability distributions.
Quantum logic and Quantum probability theory represent important approaches to manage quantum mechanical uncertainties within limits of usual mathematics and the developments in the last decades show numerous and also very difficult results. Thus, quantum logic and quantum probability theory are new theories in classical mathematics that became more and more complex. On the other hand there was the new theory of fuzzy sets and systems available at the same time and the question arose in the 1980s and 1990s of whether fuzzy sets could be useful in the interpretation of quantum mechanics. However, at that time this approach was not successful. The disappointing results may have stemmed from the fact that fuzzy set theory was not as well accepted as a mathematical tool the as it is today and from a lack of interest in using the new theory on the part of theoretical physicists. Moreover, until recently there was also no interest in fuzzy set theory in the philosophy of science. But now the theory of fuzzy sets is broadly accepted, particularly in applied sciences and technology and in the history of science the theory of fuzzy sets attracts

[1] This is analogue to Zadeh's requirement: ". . . we need a radically different kind of mathematics, the mathematics of fuzzy or cloudy quantities which are not describable in terms of probability distributions" in his 1962 article [4], p. 857.

[2] These developments regarding a theory of probabilistic structures of quantum mechanics became very complex, as the reader can see in the authors Ph. D thesis [11].

ISBN: 978-989-95079-6-8

attention [12, 13, 14]. In physics, the results of new experiments (Alain Aspect's test of Bell's inequality in 1982 [15, 16] and Anton Zeilinger's experiments on quantum teleportation since 1997 [17] have sparked a new debate on the interpretation of quantum mechanics and there is growing interest in the theory of fuzzy sets in the field of scientific history [29]. In the next sections we will evolve some ideas towards a "fuzzy view" on quantum systems.

4 Fuzzy Systems

In his talk on "A New View on System Theory," for the *Symposium on System Theory* that took place in Brooklyn in April 1965, Lotfi Zadeh defined fuzzy systems as follows:

Definition: A system S is a fuzzy system if input $u(t)$, output $y(t)$, or state $x(t)$ of S or any combination of them ranges over fuzzy sets. ([18], p. 33)[3]

He explained that "these concepts relate to situations in which the source of imprecision is not a random variable or a stochastic process but rather a class or classes which do not possess sharply defined boundaries." ([18], p. 29

Eight years later, in "Outline of a New Approach to the Analysis of Complex Systems and Decision Processes," he introduced "linguistic variables" that are variables whose values may be sentences in a specific natural or artificial language. [19] To illustrate, the values of the linguistic variable "age" might be expressible as *young*, *very young*, *not very young*, *somewhat old*, *more or less young*.

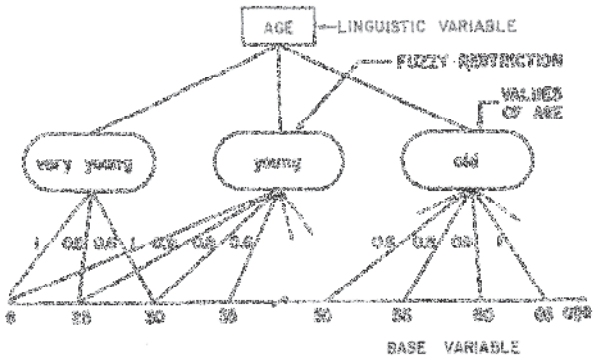


Figure 1: Example of the linguistic variable "Age" [20].

These values are formed with the label *old*, the negation *not*, and the hedges *very*, *somewhat*, and *more or less*. In this sense the variable "age" is a linguistic variable (see Fig. 1).

Linguistic variables became a proper tool for reasoning without exact values. Since in many cases, it is either impossible or too time-consuming (and therefore too expensive) to measure or compute exact values, the concept of linguistic variables has been successful in many fuzzy application systems, e.g., in control and (medical) decision making. In the next section we seek to apply the concept of linguistic variables in quantum mechanics where exact values of observables do not exist. The situation is that outcomes of a physicist's experiments have to be values of observables, i.e., an observing physicist assigns sharp values (and their classical probability distributions of the classical variables (e.g. its position $\boldsymbol{r}$ or its momentum $\boldsymbol{p}$) due to the possible errors of measurement and the systematic errors occurring in every experiment)[4] to a quantum theoretical object. This value is not sufficient to determine the quantum object's state. This is only one representation – among many others – and none of these representations of the state of the quantum system is complete!

5 Fuzzy Classical Systems

In this section a new approach to the interpretation of uncertainty in quantum mechanics using the methodology of fuzzy sets and systems is outlined.

We will interpret the "fuzzy state variable" of a physical system as a vector of linguistic variables instead of numerical variables. This interpretation yields to the interpretation of the "fuzzy state" of a classical physical object as a pair of the two linguistic variables – position and momentum

In a manner similar to Zadeh's extension of systems to fuzzy systems, the definition of a linguistic variable operating on a fuzzy set, and assignment of membership degrees and elements of the term set of the linguistic variable, the "fuzzy state" of a physical system is interpreted as a vector of linguistic variables instead of numerical variables.

A concrete system a has a certain number of properties P^i, $i \in \{1, \ldots, n\}$. In classical physics we attach classical numerical variables (observables) $V_P{}^i$ to these properties that can be measured. To use the methods of fuzzy set theory, now we attach can find also a linguistic variable $LV_P{}^i$ to representing the property P^i. These linguistic variables operate on fuzzy sets and assign membership degrees and elements of a term set, for example:

$$T(LV_P{}^i) = \{\text{very small, small, big, very big, … etc .}\}$$

We can imagine the n-tuple $\boldsymbol{LVP_n} = \langle LV_P{}^1, LV_P{}^2, \ldots, LV_P{}^n \rangle$ to be a "linguistic vector" in an n-dimensional Cartesian space. The value $\boldsymbol{LVP_n}$ (a, t) for a system a at time point t is called the "linguistic state" or "fuzzy state" of this system at this time. During this time, the linguistic state of a system moves in the "linguistic state space" or "fuzzy state space" $\Sigma_{\boldsymbol{n}}{}^L(\mathrm{a}) = \{\boldsymbol{LVP_n}(\mathrm{a}, t) \mid t \in T\}$ of the system.

In the case of a classical particle in Newtonian mechanics, the "fuzzy state" is the pair (2-tupel) $(LV_{\boldsymbol{r}}, LV_{\boldsymbol{p}})$ of the two linguistic variables position $LV_{\boldsymbol{r}}$ and momentum $LV_{\boldsymbol{p}}$ that operate on fuzzy sets and assign membership degrees and elements of a term set, for example:

$$T(LV_{\boldsymbol{r}}) = \{\text{very small, small, null, big, very big, … etc.}\}$$

and

$$T(LV_{\boldsymbol{p}}) = \{\text{very small, small, null, big, very big, … etc.}\}$$

Usually the shape of the fuzzy set's membership functions is subjectively chosen or dependant on the problem. In a very special case, the membership function may have the shape of the classical Gaussian probability distribution and thus the

[3] We assume that the reader is familiar with the basic principles of the theory of fuzzy sets.

[4] In the next sections we sometimes will omit this completion about the classical errors of measurement and the systematic errors.

ISBN: 978-989-95079-6-8

fuzzy state variable yields the probabilities of measuring the observables position ***r*** and momentum ***p*** due to the calculation of errors. However, in general, membership functions of fuzzy sets do not represent probability distributions of measurement errors or randomness, but more general uncertainties that are deeply rooted in the absence of the theoretical concept's strict boundaries.

We know already that classical concepts such as position and momentum, having strict boundaries in Newtonian mechanics, do not have such boundaries in the theory of quantum mechanics. Therefore this pair of concepts does not match the quantum mechanical state variable – to represent the quantum mechanical "state", position and momentum are in use with some uncertainty. This concept of uncertainty in quantum mechanics is often represented by classical probability, but, in the strict sense, the quantum mechanical uncertainty is different from the concept of classical probability.

6 Fuzzy Quantum Systems

Let's try to extend our approach of "fuzzy states" of systems to quantum systems to include the assumption that the classical theoretical concepts are not the right concepts, but that we have no better concepts to interpret the outcomes of classical experiments. Thus, we can use fuzzy sets and linguistic variables to convert classical observables to systems of quantum mechanics.

In the case of a quantum system, the "fuzzy quantum state" is an infinite dimensional vector $\boldsymbol{LVP_\infty}\ (a,t)$ of linguistic variables in the abstract Hilbert space, with an infinite tuple $\boldsymbol{LVP_\infty} = \langle LV_P^1, LV_P^2, \ldots, LV_P^n, \ldots \rangle$ of linguistic variables LV_i, but not all linguistic variables LV_i and LV_j are compatible, i.e. they are in an uncertainty relation with each other, e.g., $LV_1 = LV_{\boldsymbol{r}}$ (position observable) and $LV_j = LV_{\boldsymbol{p}}$ (momentum observable).

In general, at one point in time we can measure one of these LV_i and this measurement reduces the membership function to a numerical value of this observable. This is an effect that is known in usual quantum mechanics as the "collapse" of the quantum mechanical state function.

Let's use the fuzzy state concept for quantum systems! After the measurement of an observable, say V_P^i, there is no "collapse" of the fuzzy quantum state. We still have it's complete representation as the infinite tuple of linguistic variables $\boldsymbol{LVP_\infty} = \langle LV_P^1, LV_P^2, \ldots, LV_P^n, \ldots \rangle$. However, after the measurement of observable V_P^i, this component of the fuzzy quantum state is not any longer a linguistic but a numeric term: the measurement, just a number (associated with a unit).

7 Fuzzy Uncertainty Principle

First we consider the classical uncertainty principle that was found by Heisenberg in 1925. This uncertainty principle states that the values of certain pairs of conjugate variables[5] (e.g. position ***r*** and momentum ***p***) cannot both be known with arbitrary precision. That is, the more precisely one variable is known, the less precisely the other is known. We already noticed that this is not the uncertainty of the measurement of particular observables of a system.

Heisenberg's uncertainty principle was very often formulated in usual mathematics (classical probability theory and statistics) as follows: every quantum state has the property that the *root-mean-square deviation* of the position ***r*** from its mean (the standard deviation of the ***r***-distribution) times the root-mean-square deviation of the momentum ***p*** from its mean (the standard deviation of ***p***) can never be smaller than a small fixed fraction of Planck's constant:

$$\Delta r \cdot \Delta p \geq \tfrac{\hbar}{2}, \tag{1}$$

where $\Delta r = \sqrt{\left\langle (r - \langle r \rangle)^2 \right\rangle}$ and $\Delta p = \sqrt{\left\langle (p - \langle p \rangle)^2 \right\rangle}$.

Any measurement of the position with accuracy $\Delta\boldsymbol{r}$ collapses the quantum state making the standard deviation of the momentum $\Delta\boldsymbol{p}$ larger than $\frac{\hbar}{2}$.

This view is using classical probability theory what is adequate for each one of the two variables ***r*** and ***p*** but what is not appropriate to the combination of the two "pictures". There is no joint probability distribution for the two variables ***r*** and ***p*** and therefore the product of their *root-mean-square deviations* has no meaning in classical probability theory. That is, what is meant by the logico-algebraic result that there is no Boolean lattice of quantum mechanics.

Let's consider the uncertainty principle in our approach of fuzzy quantum states! In our view position and momentum, ***r*** and ***p***, are associated with linguistic variables, $LV_{\boldsymbol{r}}$ and $LV_{\boldsymbol{p}}$. Hence we can consider $\Delta r = \sqrt{\left\langle (r - \langle r \rangle)^2 \right\rangle}$ and $\Delta p = \sqrt{\left\langle (p - \langle p \rangle)^2 \right\rangle}$ as being linguistic variables as well.

Because Heisenberg's uncertainty relation (1) is a proposition of the possible values of $\Delta\boldsymbol{r}\cdot\Delta\boldsymbol{p}$, we can also consider this product of the two linguistic variables $\Delta\boldsymbol{r}$ and $\Delta\boldsymbol{p}$ as being a linguistic variable ***S*** ("action"[6]).

Heisenberg's inequality claims that $\boldsymbol{S} = \Delta\boldsymbol{r}\cdot\Delta\boldsymbol{p}$ can never be smaller than $\frac{\hbar}{2}$, i.e. the minimal value of ***S*** equals $\frac{\hbar}{2}$, but this result was never derived by Heisenberg himself. Moreover, in his famous 1930 lecture in Chicago he refined his principle to the following inequality [21]:

$$\Delta x \Delta p \gtrsim h. \tag{2}$$

The formulation (1) of the uncertainty relations was proved by E. H. Kennard in 1927 but in this prove it was postulated that $\Delta\boldsymbol{r}$ and $\Delta\boldsymbol{p}$ are the standard deviations of position and momentum, therefore this formulation uses classical probability theory [21]

What does this mean? – The value of the linguistic variable ***S*** at a given point of time is the composition of the values of the linguistic variables $\Delta\boldsymbol{r}$ and $\Delta\boldsymbol{p}$ but to get both of the two

[5] In physics, conjugate variables are pair of variables mathematically defined in such a way that they become Fourier transform duals of one-another.

[6] "Action" is the name of this particular quantity (product of position and momentum or of energy and time) in physics.

ISBN: 978-989-95079-6-8

values is not possible in the theory of quantum mechanics. (Remark: To compute the value of $\Delta \boldsymbol{r}$ or $\Delta \boldsymbol{p}$ we have to measure the value of $\boldsymbol{r}$ or $\boldsymbol{p}$ respectively.)
Thus, it is only possible to know one of the linguistic values of $\Delta \boldsymbol{r}$ or $\Delta \boldsymbol{p}$ and therefore it is not possible to compute the value of the linguistic variable $\boldsymbol{S} = \Delta \boldsymbol{r} \cdot \Delta \boldsymbol{p}$.

8 Conclusions

During the last decade the literature on combining the theories of quantum mechanics and fuzzy sets was growing, e.g. [23- 29].[7] Of course there might be many more papers on the meeting of the two theories but unfortunately – as one of the reviewers to this paper, "it seems to be sometimes the application of ideas from quantum theory to fuzzy sets than the other way around". This paper was written to plea for applications of ideas from Fuzzy set theory to quantum mechanics. Of course, we gave only some preliminary ideas to establish a fuzzy approach to physics and particularly to quantum mechanics.
We presented a rough idea of what the methodologies of fuzzy sets and systems, along with linguistic variables, could contribute to the project to represent phenomena in physics that can not be represented in classical mechanics. Quantum mechanics is a very successful theory to represent these phenomena but there are difficulties with the using of the concept of probability. It seems that the concepts of fuzzy sets and linguistic variables can be more appropriate for the representation of quantum phenomena in the framework of quantum mechanics than the classical concept of probability. Our definition of the fuzzy state of a quantum mechanical system avoids the difficulties that arise in classical probability theory in attempts to define this state. On the other hand, with this fuzzy approach new difficulties arise, e.g. with the experimental side of physics and also with its interpretative side. Of course, scientists should take such problems seriously. At the other hand, these are the usual and well-known problems of all fuzzy-approaches in the field of science and technology of the last 40 year. From the point of view of a historian and philosopher of science this is also a big challenge!

References

[1] M. Born, Zur Quantenmechanik der Stoßvorgänge, *Zeitschrift für Physik*, 37: 863-867, 1926

[2] M. Born, Das Adiabatenprinzip in der Quantenmechanik, *Zeitschrift für Physik*, 40: 167-191, 1926.

[3] J. von Neumann, *Mathematical Foundations of Quantum Mechanics*, Princeton Univ. Press, Princeton, New Jersey, 1955.

[4] L. A. Zadeh: From Circuit Theory to System Theory, *Proceedings of the IRE*, 50, 5: 856-865, 1962.

[5] G. Birkhoff, J. von Neumann, The Logic of Quantum Mechanics, *Annals of Mathematics*, Series 2, 37: 823 ff, 1936.

[6] G. W. Mackey, *Mathematical Foundations of Quantum Mechanics*, New York: W. A. Benjamin, 1963.

[7] Thank you very much to the anonymous referee who mentioned these papers in his/her review to this paper.

[7] P. Suppes, Probability Concepts in Quantum Mechanics, *Philosophy of Science*, 28: 378-389, 1961.

[8] P. Suppes, The Probabilistic Argument for a Non-Classical Logic of Quantum Mechanics, *Philosophy of Science*, 33, 14-21, 1966.

[9] S. P. Gudder, *Quantum Probability*, Academic Press: San Diego, 1988.

[10] I. Pitowski, *Quantum Probability, Quantum Logic*, Lecture Notes in Physics, 321, Berlin: 1989.

[11] A. Aspect, P. Grangier, and G. Roger, Experimental Realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: A New Violation of Bell's Inequalities, *Physical Review Letters*, 49, 2: 91-94, 1982.

[12] R. Seising, From Vagueness in Medical Thought to the Foundations of Fuzzy Reasoning in Medical Diagnosis. *Artificial Intelligence in Medicine* 38: 237-256, 20'06.

[13] R. Seising, *The Fuzzification of Systems. The Genesis of Fuzzy Set Theory and Its Initial Applications – Its Development to the 1970s*. Berlin [et al.]: Springer (Studies in Fuzziness and Soft Computing 216) 2007.

[14] R. Seising, On the Absence of Strict Boundaries – Vagueness, Haziness, and Fuzziness in Philosophy, Medicine, and Science, *Applied Soft Computing* (Special Issue: Forging the Frontiers) 8, 3: 1232-1242, 2008.

[15] A. Aspect, Dalibard, G. Roger, Experimental test of Bell's inequalities using time-varying analyzers, *Physical Review Letters,* 49: 1804-1807, 1982.

[16] Dik Bouwmeester, Jan-Wei Pan, Klaus Mattle, Manfred Eibl, Harald Weinfurter, Anton Zeilinger, Experimental Quantum Teleportation, *Nature,* 390: 575 ff, 1997.

[17] R. Seising, *Probabilistische Strukturen der Quantenmechanik*, Frankfurt: Lang 1996. (in German)

[18] L. A. Zadeh, Fuzzy Sets and Systems. In: J. Fox (ed.): *System Theory*, Microwave Res. Inst. Symp. Series XV, Brooklyn, New York: Polytechnic Press, 1965, 29-37.

[19] L. A. Zadeh, Outline of a New Approach to the Analysis of Complex Systems and Decision Processes, *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3, 1: 28-44, 1973.

[20] L. A. Zadeh, On the Analysis of Large-Scale Systems. Memorandum No. ERL-M418, Electronic Research Laboratory, College of Engineering, University of California, Berkeley, January 8, 1974.

[21] W. Heisenberg, *Physikalische Prinzipien der Quantentheorie* Leipzig: Hirzel, 1930. English translation *The Physical Principles of Quantum Theory*, Chicago: University of Chicago Press, 1930.

[22] E. H. Kennard, Zur Quantenmechanik einfacher Bewegungstypen *Zeitschrift für Physik*, 44: 326, 1927.

[23] S. Ishikawa, A quantum mechanical approach to a fuzzy theory, *Fuzzy Sets and Systems*, 90, 3: 277-306, 1997.

[24] Mirko Navara, Boolean representations of fuzzy quantum spaces, Fuzzy Sets and Systems, 87, 2: 201-207, 1997.

[25] Anatolij Dvurecenskij, Fuzzy set representations of some quantum structures, Fuzzy Sets and Systems, 101, 1: 67-78, 1999.

[26] Stan Gudder, Non-disturbance for fuzzy quantum measurements, Fuzzy Sets and Systems, 155, 1: 18-25, 2005.

[27] G.G. Rigatos, S.G. Tzafestas, Quantum learning for neural associative memories, Fuzzy Sets and Systems, 157, 13: 1797-1813, 2006.

[28] G. Cattaneo, A. Leporati, R. Leporini, Quantum conservative many-valued computing, Fuzzy Sets and Systems, 159, 9: 1001-1030, 2008.

[29] Roberto Giuntini, Hector Freytes, Antonio Ledda, Francesco Paoli, A discriminator variety of Gödel algebras with operators arising in quantum computation, Fuzzy Sets and Systems, 160, 8: 1082-1098, 2009.

ISBN: 978-989-95079-6-8

Possibility and Necessity Evaluations based on Ordinal Comparability

Masaaki Ida

National Institution for Academic Degrees and University Evaluation
1-29-1, Gakuen-nishimachi, Kodaira, Tokyo 187-8587 Japan
Email: ida@niad.ac.jp

Abstract— *Subjective uncertainty is one of the most essential subjects for evaluation that was the reason Fuzzy theory was proposed. Related research areas are widely spread such as decision making, data analysis, information retrieval, psychology, and human computer interaction and so on. As a basis for various researches mathematical interpretation of evaluation methods occupies an important position. In Possibility theory, possibility and necessity evaluations have been utilized as ways for evaluating a vague achievement in an ambiguous situation. However, mathematical interpretations for these evaluations have been insufficient. In this article, possibility and necessity evaluations are examined based on the ordinal comparability in social choice theory. Equivalent conditions to the evaluations are deduced so that axiomatic interpretations for the evaluations become possible, and order relations related to the two evaluations are deduced.*

Keywords— evaluation, measure, necessity, ordinal comparability, ordinal utility, possibility.

1 Introduction

In Possibility theory [1, 2, 3, 4, 5], subjective value of attainment or satisfaction of an evaluation item in state x is represented by the membership function $\mu(x)(0 \leqq \mu(x) \leqq 1)$. And occurrence degree of subjective possibility is represented by the membership function $\pi(x)(0 \leqq \pi(x) \leqq 1)$. Tuple of these functions, $(\mu(x), \pi(x))$, can be regarded as a representation of state x with two kinds of uncertainties (vagueness and ambiguity) and be applied for evaluation and decision making.

An evaluation value for (μ, π) is defined as follows:

Definition 1

$$\Pi(\mu, \pi) = \sup_x \min(\mu(x), \pi(x)) \tag{1}$$

This can be regarded as an extension of possibility measure for fuzzy sets. This is also interpreted as a fuzzy integral for fuzzy set on possibility measure.

In Possibility theory, necessity measure is defined as dual for possibility evaluation:

Definition 2

$$N(\mu, \pi) = \inf_x \max(\mu(x), 1 - \pi(x)) \tag{2}$$

These two evaluations (possibility and necessity) also can be interpreted as optimistic (positive) and pessimistic (negative) evaluations respectively.

In this paper, we investigate generalized possibility and necessity evaluations. As seen in these definitions, generally in Fuzzy theory, *max* operation and *min* operation are used frequently. For example, they are used in the synthetic rule of fuzzy reasoning.

Moreover, some restrictions, e.g., *normality*, are assumed on membership functions in many cases. A continuity and convexity are sometimes assumed on membership functions. [3].

Definition 3 *A membership function $\pi(x)$ is normal means that there exists x such that $\pi(x) = 1$.*

In order to perform the mathematical interpretation on possibility and necessity evaluations and to examine the validity of the evaluations, it is necessary to consider on the property of functions $\mu(x)$ and $\pi(x)$ determined subjectively. For example, whether it is a cardinal utility function or ordinal utility function.

In Possibility theory, considerations for improving qualitative possibilistic criteria are discussed. Axiomatic approach has been made in order to examine decision theoretic foundation [6]. Generalized concordance rules that request comparisons within possible values for components and comparisons between sets of attributes is proposed [7]. The injection of lexicographic ingredients in the possibilistic criteria has been studied [8].

It is necessary to consider the conditions at the time of unifying the functions and performing synthetic evaluation, e.g., evaluation level comparability on evaluation items. In social choice theory or voting theory, the comparability of utility functions was introduced with ordinal number based on extended ordering [9, 10]. and examination has been made about how synthetic evaluation should be performed [11].

This paper newly defines possibility and necessity evaluations based on ordinal comparability and deduces equivalent conditions to them. Therefore, the applicable conditions of possibility and necessity evaluations are clarified. Moreover, relations between these two evaluations are discussed.

Hereafter, Section 2 examines necessary and sufficient condition of the evaluations based on extended ordering. In Section 3, interpretations of the conditions for evaluations, possibility, and necessity evaluation, are discussed. Section 4 considers order relations between possibility evaluation and necessity evaluation.

2 Extended ordering and preference relation

We denote that X is a set of states or a set of alternatives. We define *preference relation* as a binary relation.

Definition 4 *Preference relation R on X ($x, y \in X, xRy$: x is preferred to y) is a binary relation which is reflective*

ISBN: 978-989-95079-6-8

($\forall x \in X : xRx$), transitive ($\forall x, y, z \in X : (xRy \wedge yRz) \rightarrow xRz$). and total ($\forall x, y \in X : (x \neq y) \rightarrow (xRy \vee yRx)$). Strong preference relation P and indifference relation I are defined by: $xPy \equiv (xRy \wedge \neg(yRx))$, and $xIy \equiv (xRy \wedge yRx)$.

We denote H as a set of evaluation items. In social choice theory [9], X is called *a set of social states*, and H is called *a set of individuals*. In this paper, we make fundamental investigations on preference relation, so that the number of elements of the set H is restricted to two.

We consider $(x, i) \in X \times H$. A binary relation $\tilde{R}$ on $X \times H$ is assumed to be a preference relation. This relation is called *extended ordering* [9]. In similar meaning, it is also called extended sympathy [10].

Definition 5 *Preference relation $(x, i)\tilde{R}(y, j)$ means that an evaluation item i on a state x is preferred to j on a state y. Strong preference relation $\tilde{P}$ and indifference relation $\tilde{I}$ on $X \times H$ are defined by: $(x, i), (y, j) \in X \times H$ and $\tilde{R}$, $(x, i)\tilde{P}(y, j) \equiv ((x, i)\tilde{R}(y, j) \wedge \neg((y, j)\tilde{R}(x, i)))$, $(x, i)\tilde{I}(y, j) \equiv ((x, i)\tilde{R}(y, j) \wedge (y, j)\tilde{R}(x, i))$.*

Here, we consider generalized function $f : \tilde{R} \rightarrow R$, which corresponds the preference relation $\tilde{R}$ ($\tilde{P}$ and $\tilde{I}$) on $X \times H$ to the preference relation R (P and I) on X. The function f is called *Generalized Social Welfare Function* (GSWF) in social choice theory [9].

In the following section, higher evaluation on state x is denoted as $2(x) \in \{1, 2\}$ and lower evaluation is denoted as $1(x) \in \{1, 2\}$. This means that $(x, 2(x))\tilde{R}(x, 1(x))$.

Definition 6 *GSWF f^P is defined as follows. For all $x, y \in X$,*
(1) if for all $i \in \{1, 2\}$ $(x, i(x))\tilde{I}(y, i(y))$, then xIy,
(2) if $(x, 1(x))\tilde{I}(y, 1(y))$ and $(x, 2(x))\tilde{P}(y, 2(y))$, then xPy,
(3) if $(x, 1(x))\tilde{P}(y, 1(y))$, then xPy.

GSWF f^P corresponds to *leximin rule*, because we compare preference from lower level to higher level [9, 11].

Next, we consider following five conditions on the two preference relations $\tilde{R}$ and R. We denote $(x, i)\tilde{R}(y, i)\tilde{R}(z, i)$ for short as $(x, i)\tilde{R}(y, i)$ and $(y, i)\tilde{R}(z, i)$. We denote for $\tilde{P}$ and $\tilde{I}$ in similar way.

(C1) *Preference relation $\tilde{R}$ can take any logically possible relation.*

(C2) *If the restrictions of $\tilde{R}$ and $\tilde{R}'$ on any pair in X are the same, then the restrictions of R and R' on that pair are also same.*

(C3) *For any x, y in X, if $(x, 1)\tilde{R}(y, 1) \wedge (x, 2)\tilde{R}(y, 2)$, then xRy. If one of the two relations is a strong preference relation $\tilde{P}$, then xPy.*

(C4) *For any x, y in X, if $(y, 2)\ \tilde{P}\ (x, 2)\ \tilde{R}\ (x, 1)\ \tilde{P}\ (y, 1)$ or if $(y, 1)\ \tilde{P}\ (x, 1)\ \tilde{R}\ (x, 2)\ \tilde{P}\ (y, 2)$, then xPy.*

(C5) *For any x, y in X, if $(x, 1)\tilde{I}(y, 2) \wedge (x, 2)\tilde{I}(y, 1)$ and $\neg((x, 1)\tilde{I}(x, 2))$, then xIy.*

We note that if $(x, 1)\tilde{R}(y, 1) \wedge (x, 2)\tilde{R}(y, 2)$ and $(y, 1)\tilde{R}(x, 1) \wedge (y, 2)\tilde{R}(x, 2)$, then xIy by (C3). Furthermore we note that the conditions (C3), (C4), and (C5) are mutually independent.

For GSWF f^P and the five conditions, following theorem is stated [12].

Theorem 1 *Assume that $|X| > 2$, only GSWF that satisfies (C1), (C2), (C3), (C4), and (C5) is f^P.*

Next, from the view of symmetry, we consider a different condition only for (C4) in the above five conditions. We define the following converse condition (C4').

(C4') *For any x, y in X, if $(x, 2)\ \tilde{P}\ (y, 2)\ \tilde{R}\ (y, 1)\ \tilde{P}\ (x, 1)$ or if $(x, 1)\ \tilde{P}\ (y, 1)\ \tilde{R}\ (y, 2)\ \tilde{P}\ (x, 2)$, then xPy.*

In this case we consider the followings.

Definition 7 *GSWF f^N is defined as follows. For all $x, y \in X$,*
(1) for any $i \in \{1, 2\}$, if $(x, i(x))\tilde{I}(y, i(y))$, then xIy,
(2) if $(x, 2(x))\tilde{I}(y, 2(y))$ and $(x, 1(x))\tilde{P}(y, 1(y))$, then xPy,
(3) if $(x, 2(x))\tilde{P}(y, 2(y))$, then xPy.

For GSWF f^N and the five conditions, following theorem is stated.

Theorem 2 *Assume $|X| > 2$, only GSWF that satisfies conditions (C1), (C2), (C3), (C4'), (C5) is f^N.*

By the above two theorems, conditions equivalent to two GSWFs, f^P and f^N, were obtained. Based on these results we can discuss what kind of conditions are assumed when each GSWF is applied for evaluation.

3 Interpretation of possibility and necessity evaluation based on extended ordering

3.1 Ordinal utility function

In this section, as an application of the knowledge obtained in the previous section to Possibility theory, we consider the case that $H = \{\mu, \pi\}$. In the case we define GSWF f^P as a possibility evaluation.

We can say from (1) and the definition of f^P that if xPy for f^P, then $\min(\mu(x), \pi(x)) \geqq \min(\mu(y), \pi(y))$.

In the following, we discuss the interpretation of possibility evaluation based on the theorem obtained in the previous section.

We consider that $(x, i) \in X \times H$ corresponds to (x, μ) and (x, π). The relations between the preference relation $\tilde{R}$ and membership functions are defined as follows;
$(x, \mu)\tilde{R}(y, \mu) \equiv \mu(x) \geqq \mu(y)$,
$(x, \pi)\tilde{R}(y, \pi) \equiv \pi(x) \geqq \pi(y)$,
$(x, \mu)\tilde{R}(x, \pi) \equiv \mu(x) \geqq \pi(x)$,
$(x, \pi)\tilde{R}(x, \mu) \equiv \pi(x) \geqq \mu(x)$.
We denote strong preference relation and indifference relation as "$>$" and "$=$" respectively.

ISBN: 978-989-95079-6-8

We should note that membership functions μ and π can be regarded as **ordinal level comparable utility functions** based on extended ordering.

3.2 Possibility evaluation

In this subsection we discuss the conditions in Theorem 1.

Conditions (C1) and (C2) are naturally acceptable conditions.

Condition (C3) (if $\mu(x) \geqq \mu(y)$ and $\pi(x) \geqq \pi(y)$, then xRy) is so-called Pareto condition, so that this is naturally acceptable for the functions μ and π.

However, (C4) (if $\mu(y) > \mu(x) \geqq \pi(x) > \pi(y)$ or $\pi(y) > \pi(x) \geqq \mu(x) > \mu(y)$, then xRy) means that we must compare two evaluation elements of μ and π on one state.

And condition (C5) (if $\mu(x) = \pi(y)$ and $\pi(x) = \mu(y)$, then xIy) means that in indifferent two states, two evaluation elements of μ and π can be exchanged as same.

Therefore conditions (C4) and (C5) for the functions μ and π are restrictions on two evaluations, then these restrictions are strongly imposed on setting these two membership functions.

If conditions (C1), (C2), (C3), (C4), (C5) are acceptable for μ and π, then by Theorem 1, we can utilize the function f^P as a preference functions on $x, y \in X$.

On the other hand, in the case that we use the function f^P as a preference relation, the conditions (C1), (C2), (C3), (C4), (C5) should be satisfied.

Consequently, for $x, y \in X$ we clarify the conditions for evaluation of (1) based on comparing $\min(\mu(x), \pi(x))$ and $\min(\mu(y), \pi(y))$.

3.3 Necessity evaluation

Similarly we can discuss necessity evaluation. In case of necessity, for π in (2) we can see $1 - \pi(x)$. Therefore, as a converse order relation we defined an ordinal utility function $\pi^*(x)$ as follows.

Definition 8 *$\pi^*(x)$ is defined by:*

$$\pi^*(x) < \pi^*(y) \quad \equiv \quad \pi(x) > \pi(y), \tag{3}$$

$$\pi^*(x) = \pi^*(y) \quad \equiv \quad \pi(x) = \pi(y). \tag{4}$$

Similarly as possibility evaluation we can discuss the interpretation of necessity evaluation based on Theorem 2 obtained in the previous section.

We consider that $(x, i) \in X \times H$ corresponds to (x, μ) and (x, π^*) in the same way as possibility. The relation between the preference relation $\tilde{R}$ and membership functions are defined as follows;

$(x, \mu)\tilde{R}(y, \mu) \equiv \mu(x) \geqq \mu(y)$,
$(x, \pi^*)\tilde{R}(y, \pi^*) \equiv \pi^*(x) \geqq \pi^*(y)$,
$(x, \mu)\tilde{R}(x, \pi^*) \equiv \mu(x) \geqq \pi^*(x)$,
$(x, \pi^*)\tilde{R}(x, \mu) \equiv \pi^*(x) \geqq \mu(x)$.

We denote strong preference relation and indifference relation as "$>$" and "$=$" respectively.

We should note that membership functions μ and π^* can be regarded as ordinal comparable utility functions based on extended ordering. As similar with possibility, we can discuss μ, π^*, and conditions (C1), (C2), (C3), (C4'), (C5).

4 Order relation between possibility evaluation and necessity evaluation

In this section, as discussed in earlier sections we assume that $H = \{\mu, \pi\}$ and that extended ordering on (x, μ) and (x, π) satisfies conditions (C1), (C2), (C3), (C4), and (C5) for preference relations $\tilde{R}$ and R.

Definition 9 *$x^P \in X$ is defined by: for all $y(\neq x^P) \in X$, $x^P R y$.*

Similarly we assume that $H = \{\mu, \pi^*\}$ and that extended ordering on (x, μ) and (x, π^*) satisfies conditions (C1), (C2), (C3), (C4'), and (C5) for preference relations $\tilde{R}^*$ and R^*. We discriminate $\tilde{R}, R$ and $\tilde{R}^*, R^*$.

Definition 10 *$x^N \in X$ is defined by: for all $y(\neq x^N) \in X$, $y R^* x^N$.*

We should recall that $\mu(x)$ and $\pi(x)$ are ordinal functions.

In Possibility theory, optimistic evaluation and pessimistic evaluation are discussed. In this section, order relation between possibility evaluation and necessity evaluation is discussed.

We obtain the following result with respect to $\mu(x^P)$ and $\mu(x^N)$.

Lemma 3

$$\mu(x^P) \geqq \mu(x^N). \tag{5}$$

(Proof)
Assuming $\mu(x^P) < \mu(x^N)$ we deduce contradiction. In this case, $x^P \neq x^N$. If $\pi(x^P) \leqq \pi(x^N)$, then $(x^N, \mu)\tilde{P}(x^P, \mu)$ and $(x^N, \pi)\tilde{R}(x^P, \pi)$. Therefore, by (C3) of GSWF f^P, $x^N P x^P$. This contradicts the definition of x^P. Next, we consider the case that $\pi(x^P) > \pi(x^N)$. In this case, $\pi^*(x^P) < \pi^*(x^N)$ and $\mu(x^P) < \mu(x^N)$, then $(x^N, \mu)\tilde{P}^*(x^P, \mu)$ and $(x^N, \pi)\tilde{P}^*(x^P, \pi)$. By (C3) of GSWF f^N, $x^N P^* x^P$, so that this contradicts the definition of x^N.

For order relations with respect to $\max(\mu(x^N), \pi^*(x^N))$ and $\min(\mu(x^P), \pi(x^P))$, we obtain the following results.

Lemma 4 *If $\pi^*(x) > \pi(x)$, then*

$$\max(\mu(x^N), \pi^*(x^N)) > \min(\mu(x^P), \pi(x^P)). \tag{6}$$

Lemma 5 *If there exists x' which satisfies the condition $\pi(x') \geqq \mu(x') \geqq \pi^*(x')$, then*

$$\min(\mu(x^P), \pi(x^P)) \geqq \mu(x') \geqq \max(\mu(x^N), \pi^*(x^N)). \tag{7}$$

(Proof)
By the definition of x^P, we have $\min(\mu(x^P), \pi(x^P)) \geqq \min(\mu(x'), \pi(x')) \geqq \mu(x')$. Similarly with respect to x^N, $\max(\mu(x^N), \pi^*(x^N)) \leqq \max(\mu(x'), \pi^*(x')) \leqq \mu(x')$. Consequently we obtain the result.

Next we discuss in the framework of conventional Possibility theory, i.e., we consider $\pi^*(x) = 1 - \pi(x)$. x^P is most preferable on X with respect to GSWF f^P. Then x^P takes the optimal value of (1) in Definition 1. Similarly, x^N is most unpreferable on X with respect to GSWF f^N. Then x^N takes the optimal value of (2) in Definition 2.

ISBN: 978-989-95079-6-8

Some assumptions are sometimes taken in Possibility theory. In this section we assume that the membership function $\pi(x)(0 \leqq \pi(x) \leqq 1)$ satisfies normality condition, i.e., there exists the $x' \in X$ such that $\pi(x') = 1$. Then we obtain the following result by Lemma 5.

Theorem 6 *If $\pi(x)$ is normal, then*

$$\Pi(\mu, \pi) \geqq N(\mu, \pi). \tag{8}$$

(Proof) By the definition of normality there exists x' that $\pi^*(x') = 0$ and $\pi(x') = 1$, and $0 \leqq \mu(x') \leqq 1$, then the prerequisite condition of Lemma 5 is satisfied.

5 Conclusion

This paper described the mathematical foundation of subjective uncertainty. Axiomatic interpretations on possibility evaluation and necessity evaluation, and the relation between these evaluations were examined based on the extended ordering in social choice theory.

The obtained results are as follows:

In section 2, based on extended ordering we obtained the equivalent conditions for two kinds of GSWF f^P and f^N; conditions (C1), (C2), (C3), (C4), (C5), and conditions (C1), (C2), (C3), (C4'), (C5).

In section 3, we examined generalized possibility evaluation and necessity evaluation, and discussed the conditions obtained in section 2.

In section 4, we defined x^P and x^N, and discussed the order relations between possibility evaluation and necessity evaluation, and obtained some results on the relations.

By these results obtained in this paper, mathematical interpretations and applicable conditions of the evaluations became clear. These are important when the evaluations are applied in various practical situations.

References

[1] L.A. Zadeh. Fuzzy Sets as a Basis for a Theory of Possibility, *Fuzzy Sets and Systems*, 1:3–28, 1978.

[2] D. Dubois, and H. Prade. *Possibility Theory*, Plenum Press, 1988.

[3] H. Tanaka, and P. Guo. *Possibilistic Data Analysis for Operations Research*, Physica-Verlag, 1999.

[4] T. Murofushi, and M. Sugeno. An Interpretation of Fuzzy Measures and the Choquet Integral as an Integral with Respect to a Fuzzy Measure, *Fuzzy Sets and Systems*, 29:201–227, 1989.

[5] G. Klir. *Uncertainty and Information*, Wiley, 2006.

[6] D. Dubois, H. Prade, and R. Sabbadin. Decision-theoretic Foundations of Qualitative Possibility Theory, *European Journal of Operational Research*, 128:459–476, 2001.

[7] D. Dubois, H. Fargier, P. Perny, and H. Prade. A Characterization of Generalized Concordance Rules in Multicriteria Decision Making. *International Journal of Intelligent Systems*, 18:751–774, 2003.

[8] H. Fargier and R Sabbadin. Qualitative decision under uncertainty: Back to expected utility. *Artificial Intelligence*, 164:245–280, 2005.

[9] A. Sen. *Collective Choice and Social Welfare*, Holden-Day, 1970.

[10] K.J. Arrow, A. Sen, and K. Suzumura. *Handbook of Social Choice and Welfare*, Elsevier, 2002.

[11] P.J. Hammond. Equity, Arrow's Conditions, and Rawls' Difference Principle, *Econometrica*, 44(4):793–804, 1976.

[12] M. Ida. Possibility Evaluation based on Extended Ordering, *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, 19:41–46, 2007.

ISBN: 978-989-95079-6-8

Fuzzy answer set programming with literal preferences

Jeroen Janssen[1*] Steven Schockaert[2†] Dirk Vermeir[1] Martine De Cock[3‡]

1. Department of Computer Science, Vrije Universiteit Brussel, Belgium
2. Department of Applied Mathematics and Computer Science, Universiteit Gent, Belgium
3. Institute of Technology, University of Washington, Tacoma, USA
Email: {jeroen.janssen,dvermeir}@vub.ac.be, steven.schockaert@ugent.be, mdecock@u.washingon.edu

Abstract— In the current approaches to fuzzy answer set programming (FASP) one can state preferences amongst rules to denote their relative importance. However, in many situations we need more complex preferences such as those in the answer set optimization framework proposed by Brewka for crisp answer set programming. Unfortunately, these complex preferences do not readily fit into the current FASP approaches. In this paper we propose a language to state such preferences and show that programs with these preferences can be translated into equivalent general fuzzy answer set programming (gFASP) programs. This not only provides an implementation method, but also shows that this extension can be added as syntactic sugar on top of general fuzzy answer set programming solvers.

Keywords— Logic Programming, Fuzzy Logic, Answer Set Programming, Preference-based Logic Programming.

1 Introduction

Answer set programming [1] is a logic programming language based on the stable model semantics [2]. Roughly speaking, in answer set programming a program contains a set of *rules* of the form $a \leftarrow b_1, \ldots, b_n, \mathit{not}\ c_1, \ldots, \mathit{not}\ c_m$ describing a problem whose solution is generated by an *answer set solver*. For example, if one is interested in finding solutions to the problem of assigning a seat s to either of two persons p_1 and p_2, the following program could be used:

$$\begin{array}{lcl} sit(p_1, s) & \leftarrow & \mathit{not}\ sit(p_2, s) \\ sit(p_2, s) & \leftarrow & \mathit{not}\ sit(p_1, s) \end{array}$$

This program has two *answer sets*, which correspond to the solutions of the modeled problem, $S_1 = \{sit(p_1, s)\}$ and $S_2 = \{sit(p_2, s)\}$.

In recent work, answer set programming and, more general, logic programming have been extended to deal with fuzzy [3, 4], probabilistic [5] and many-valued logics [6, 7, 8, 9], allowing for a more flexible description of both the problem domain and the proposed solutions of the problem domain. Specifically, in [4] degrees are attached to answer sets describing the quality of prospective solutions. For example, in case one is interested in only allowing solutions in which people are seated close to friends, one could write a program containing fuzzy *constraints*, as follows:

$$0 \leftarrow sit(p_1, s_1) \wedge sit(p_2, s_2) \wedge \mathit{friend}(p_1, p_2) \wedge \sim near(s_1, s_2)$$

where $sit(p, s)$ means that person p sits on seat s, $\mathit{friend}(p_1, p_2)$ that person p_1 is a friend of p_2 (to a certain degree), $near(s_1, s_2)$ denotes the proximity of seats s_1 and s_2 and $\sim$ is a negator. The constraint above would then attach a higher degree of quality to answer sets in which friends are close together. Note that $\sim near(s_1, s_2)$ should be used and not $near(s_1, s_2)$ since the satisfaction of the constraint rule should increase with the value of $near(s_1, s_2)$. More details can be found in Section 2.

Although the quality degree allows to express preference amongst solutions, the current proposals only allow this preference to be based on the degree of rule fulfillment. However, in many practical situations, we would like to express preferences based on the truth degrees of literals. Furthermore, we might want to express that the importance of some preferences depends on the context, i.e. we would like to have conditional preferences. For example, we wish to be able to state that if we are sitting on a seat, we prefer it to be by the window, but only consider this important when the scenery is nice, as follows:

$$scene : (sit(p, s) \rightarrow nearWin(s)) \triangleleft nicescenery \qquad (1)$$

where $nearWin(s)$ denotes that seat s is near the window, the $nicescenery$ literal determines the weight of importance of the rule and $\rightarrow$ is an implicator. The exact semantics of such rules are given in Section 3.

In the crisp case, Brewka proposed rules like (1) in [10]. However, to the best of our knowledge no such proposal exists for fuzzy answer set programs, which we remedy in this paper. The main contributions of this paper are the following:

- In Section 3 we propose a generalisation of the *Preference Description Language* (PDL) of Brewka to the fuzzy case.
- In Section 4 we show how programs containing these preferences can be translated to standard general fuzzy answer set programming (gFASP) programs [11], providing an implementation method and hence also showing that this extension is actually syntactic sugar on the general fuzzy answer set programming language.

2 General Fuzzy Answer Set Programming

General fuzzy answer set programming (gFASP) programs over a complete lattice $\mathcal{L}$ consist of *rules*, which are objects of the form

$$r\colon a \leftarrow f(b_1, \ldots, b_n)$$

*Funded by a joint Research Foundation–Flanders (FWO) project

†Postdoctoral fellow of the Research Foundation–Flanders (FWO)

‡On leave from Ghent University

ISBN: 978-989-95079-6-8

where r is the label of the rule, a and b_i, $1 \leq i \leq n$, are either elements from $\mathcal{L}$ or propositional symbols called *literals* (where $b_i = b_j$ for $i \neq j$ is possible), f is a (computable) function from $\mathcal{L}^n$ to $\mathcal{L}$ such that for $b_k, b'_k \in \mathcal{L}$ it holds that if $b_k \leq b'_k$, either $f(b_1, \ldots, b_k, \ldots, b_n) \leq f(b_1, \ldots, b'_k, \ldots, b_n)$ or $f(b_1, \ldots, b_k, \ldots, b_n) \geq f(b_1, \ldots, b'_k, \ldots, b_n)$ and $\leftarrow$ denotes a residual implicator over $\mathcal{L}$. If $a \in \mathcal{L}$, the rule is called a *constraint*. The head a of a rule r is denoted as r_h and the body $f(b_1, \ldots, b_n)$ is denoted as r_b.

The semantics are given by *interpretations*, i.e. fuzzy sets over the literals of a program. These are extended to give meaning to rules in a straightforward fashion, i.e. if I is an interpretation, $[a \leftarrow f(b_1, \ldots, b_n)]_I = [a]_I \leftarrow [f(b_1, \ldots, b_n)]_I$, where $[a]_I = I(a)$ if a is a literal and $[a]_I = a$ if $a \in \mathcal{L}$, and $[f(b_1, \ldots, b_n)]_I = f([b_1]_I, \ldots, [b_n]_I)$.

Models of a program P are defined using an *aggregator expression* $\mathcal{A}_P$ which maps *rule propositions* (corresponding to rule labels) to a value in a quasi-ordered set $\mathcal{Q}$. For example, given a program P_1 with two rules r_1 and r_2, a possible aggregator would be $\mathcal{A}_{P_1} = r_1 + r_2$. Such an expresssion can be evaluated by an interpretation, e.g. for I an interpretation of P we obtain $[\mathcal{A}_{P_1}]_I = \gamma([r_1]_I) + \gamma([r_2]_I)$, with γ an order-preserving $\mathcal{L} \rightarrow \mathcal{Q}$ mapping. A *k-model*, $k \in \mathcal{Q}$, is then an interpretation for which $[\mathcal{A}_P]_I \geq k$.

k-Answer sets of a program P are k-models that contain the maximal amount of knowledge inferrable from a program through forward chaining, without resorting to external hypotheses, i.e. they are models for which the truth value of every literal is supported by the application of a rule in the program.

As an example, consider program P_{seat}, describing seating arrangements, with the following rulebase $\mathcal{R}_{P_{seat}}$ over the lattice $([0,1], \leq)$:

$$
\begin{array}{rrll}
gen\colon & sit(P,S) & \leftarrow_g & 1 \\
c_1\colon & 0 & \leftarrow_g & sit(P,S) \wedge sit(P',S) \wedge P \neq P' \\
c_2\colon & 0 & \leftarrow_g & sit(P,S) \wedge sit(P,S') \wedge S \neq S' \\
cr\colon & 0 & \leftarrow_g & sit(P,S) \wedge (1 - sit(P,S)) \\
u\colon & unh(P) & \leftarrow_g & sit(P,S) \wedge friend(P,P') \\
 & & & \wedge sit(P',S') \wedge (1 - near(S,S')) \\
q\colon & 0 & \leftarrow_l & unh(P) \\
s\colon & s(P) & \leftarrow_g & sit(P,S) \\
a\colon & 0 & \leftarrow_g & 1 - s(P)
\end{array}
$$

where $\wedge$ is the minimum t-norm over $([0,1,\leq)$ and $\leftarrow_l$, $\leftarrow_g$ are resp. the Łukasiewicz and Gödel implicator. The aggregator is defined as $\mathcal{A}_{P_{seat}} = \inf_{r \in \mathcal{R}_{P_{seat}} \setminus \{gen\}} r$. Due to the fact that the gen-rule is not incorporated in the aggregation expression, this rule generates random truth degrees for $sit(P,S)$. To ensure that $sit(P,S)$ is either 0 or 1, we use the cr (crispify) rule. Rules c_1 and c_2 guarantee that a seat is occupied by a single person, resp. that a person only occupies one seat. The a and s rules ensure that everyone has a seat and the u and q rules are used to prefer solutions where friends are seated close together. Note that due to grounding a rule such as $s(P) \leftarrow_g sit(P,S)$ actually denotes a set of rules $\{s(p_i) \leftarrow_g sit(p_i, s_j)\}$ for a set of persons $(p_i)_{i \in I}$ and a set of seats $(s_j)_{j \in J}$.

If we have three seats s_1, s_2 and z, only two of which are near, viz. $near(s_1, s_2)^{\cdot 8}$ and $near(s_2, s_1)^{\cdot 8}$ and there are three persons a, b and c connected by friendship degrees $friend(a,b)^{\cdot 8}$, $friend(a,c)^{\cdot 5}$, and $friend(b,c)^0$, then from program P_{seat} one could e.g. derive the .2-answer set $A_1 = \{sit(a,s_1)^1,\ sit(b,z)^1,\ sit(c,s_2)^1,\ unh(a)^{\cdot 8},\ \ldots\}$ and the (better) .5-answer set $A_2 = \{sit(a,s_1)^1,\ sit(b,s_2)^1,\ sit(c,z)^1,\ unh(a)^{\cdot 5}, \ldots\}$.

For more details on the gFASP framework, we refer the reader to [11].

3 Complex Preferences on Literals

3.1 Fuzzy preference rules: syntax

In this section, we build a fuzzy preference framework based on a *fuzzy preference description language* ($FPDL$) which is an extension of the language PDL proposed by Brewka in [10]. Formally, a gFASP program with literal preferences consists of a tuple (P, e), where P is a gFASP program and e is an expression from $FPDL$. This expression determines a preference ordering on answer sets based on truth degrees of literals.

The basic elements of this language are *fuzzy preference rules*. Note that although they are called *rules*, they should not be confused with the rules of a fuzzy answer set program introduced in Section 2: the former are used for creating preference expressions (the e in (P, e)), whereas the latter are used to create (general) fuzzy answer set programs (the P in (P, e)). Formally the fuzzy preference rules are defined over a set of literals L and the complete lattice $\mathcal{L}$ as expressions of the following form:

$$r : g(a_1, \ldots, a_n) \triangleleft f(b_1, \ldots, b_m) \tag{2}$$

In this rule, the a_is ($1 \leq i \leq n$) and b_js ($1 \leq j \leq m$) are literals from L, the function g is a mapping from $\mathcal{L}^n$ to $\mathcal{L}$, f is a mapping from $\mathcal{L}^m$ to $\mathcal{L}$ and r is the label of the rule. For convenience, for a given fuzzy preference rule r we also refer to $g(a_1, \ldots, a_n)$ as r_h (the *head* of the rule) and to $f(b_1, \ldots, b_m)$ as r_b (the *body* of the rule).

Intuitively, the g-function denotes an expression determining the *suitability* of an answer set with respect to a certain preference and the f-function denotes the *applicability* of the rule, meaning the conditions under which this preference rule is considered important or relevant. For example, consider the following preference rule:

$$scene : (sit(p,s) \rightarrow nearWin(s)) \triangleleft nicescenery$$

This rule states that by default we (with "we" being person p) would like to have a seat near the window. However, we only consider this requirement important to the degree that the scenery is actually nice. The reason for choosing g to be an implication in this rule is that we want to state that this specific rule is also fulfilled (thus the answer set considered is still "suitable") if we are not sitting at this seat. Note that there is a difference between the $\rightarrow$ in the head and the $\triangleleft$ of the rule: if we would rewrite the scene rule as

$$scene' : (sit(p,s) \wedge nicescenery \rightarrow nearWin(s)) \triangleleft 1$$

it would denote that the nicer the scenery, the closer to the window we would like to sit rather than that we only find it important to sit near the window when the scenery is nice.

ISBN: 978-989-95079-6-8

This can also be seen from the fact that if $\rightarrow$ is a residual implicator and $nicescenery$ is true to a degree of 0.5, the $scene$ rule is only fully fulfilled when $nearWin(s)$ is 1 (or $sit(p,s)$ is 0), whereas in the $scene'$ rule it suffices that $nearWin(s)$ is 0.5 (or $sit(p,s)$ is 0).

Intuitively we thus prefer answer sets with a better suitability score, but can forgive the answer set a low suitability score if the applicability degree is rather low. Formally the semantics of rules are dependent on a function $suit$ and app defined for a given fuzzy preference rule r and an arbitrary answer set A as

$$suit(A,r) = [g(a_1,\ldots,a_n)]_A = g(A(a_1),\ldots,A(a_n))$$

$$app(A,r) = [f(b_1,\ldots,b_m)]_A = f(A(b_1),\ldots,A(b_m))$$

In the approach of Brewka [10], the semantics of preference rules are defined by attaching a *penalty* score in $\mathbb{R}$ to each applicable rule. Such a penalty reflects how good an answer set fulfills a preference rule, where lower penalties correspond to better fulfillment. Inapplicable rules are not considered important and are therefore attached penalty 0, i.e. they are always optimally fulfilled.

Unfortunately, the same approach is not possible when generalizing from crisp to fuzzy answer set programming: in the fuzzy case the applicability value is in $[0,1]$ and by reducing the values $suit(A,r)$ and $app(A,r)$ to a single number (i.e. penalty score in $\mathbb{R}$), we lose the important semantic distinction between suitability of answer sets and applicability of preference rules. In the following sections, we define an alternative semantics for fuzzy preference rules which does adhere to this distinction. We will first give the semantics for rules with a crisp applicability score and then define penalty expressions built from rules; starting from the semantics thus obtained we then introduce the semantics for rules with a fuzzy applicability score using the recently introduced concept of gradual number [12]; finally we introduce the general penalty expressions built from rules with fuzzy applicabilities.

3.2 Rules with crisp applicability

Suppose r is a rule with a crisp applicability condition (i.e. $app(A,r)$ is either 1 or 0), then we define the penalty score as:

$$pen^c(A,r) = \begin{cases} \Xi(suit(A,r)) & \text{if } app(A,r) = 1 \\ 0 & \text{otherwise} \end{cases}$$

where Ξ is a decreasing $\mathcal{L}$ to $[0,+\infty[$ mapping, such that $\Xi(1) = 0$, in the examples in this paper usually assumed to be $\Xi(x) = 1 - x$.

As an example, consider the following rule:

$$VIP_p : goodSeat(p) \triangleleft vip(p)$$

where $goodSeat(p) \in [0,1]$ reflects how good the seat of person p is and $vip(p) \in \{0,1\}$ denotes whether person p is a VIP. This rule encodes that we find it important that VIPs have good seats. If $A(goodSeat(p)) = 0.8$ and $A(vip(p)) = 1$, we know from the definition of pen^c that $pen^c(A,r) = 1 - 0.8 = 0.2$. Hence, answer set A has a low penalty score and thus represents a seating configuration that satisfies our wishes.

Starting from rules with crisp applicability, we can define the set of complex penalty expresssions $FPDL^p$ as follows:

1. If r is a fuzzy preference rule whose applicability condition takes only crisp values, r is in $FPDL^p$;
2. If $e_1, e_2, \ldots, e_k$ are in $FPDL^p$, then so are $(sum\ e_1,\ldots,e_k)$ and $(max\ e_1,\ldots,e_k)$.

We extend the definition of pen^c to cover complex penalty expressions in a straightforward way, e.g.

$$pen^c(A, sum\ e_1, ..., e_k) = pen^c(A,e_1) + ... + pen^c(A,e_k)$$

For example, suppose we have two VIP_p rules for persons p_1 and p_2, then we can combine these into one penalty expression summing the penalties as

$$e_v : (sum\ VIP_{p_1}, VIP_{p_2})$$

If we then have an answer set A for which $pen^c(A, VIP_{p_1}) = 0.2$ and $pen^c(A, VIP_{p_2}) = 0.5$, the total penalty of this expression is $pen^c(A, (sum\ VIP_{p_1}, VIP_{p_2}) = 0.7$.

3.3 Rules with vague applicability

In the general case, where applicability conditions are vague, we define penalties as gradual numbers in the sense of [12], i.e. as mappings from $]0,1]$ to $\mathbb{R}$. For each threshold λ in $]0,1]$, we convert preference rules r of the form (2) to preference rules r_λ with a crisp applicability condition:

$$r_\lambda : g(a_1,\ldots,a_n) \triangleleft (f(b_1,\ldots,b_m) \geq \lambda) \tag{3}$$

The (gradual) penalty of A w.r.t. a preference rule r is then given by the gradual number $pen(A,r)$, defined for each λ in $]0,1]$ as

$$pen(A,r)(\lambda) = pen^c(A, r_\lambda) \tag{4}$$

In Figure 1, one can see an example of the penalty thus induced by an answer set A on the rule VIP_p where $A(vip(p)) = app(A, VIP_p) = 1/2$ and $A(goodSeat(p)) = suit(A, VIP_p) = 0.75$. Note that for each rule the set $\{\lambda \mid pen(A,r)(\lambda) = \Xi(suit(A,r))\}$ is the half-open interval $]0, app(A,r)]$ and the set $\{\lambda \mid pen(A,r)(\lambda) = 0\}$ is the half-open interval $]app(A,r), 1]$. Hence, an answer set is only accountable (gets a penalty) w.r.t. a rule to the extent that the rule is applicable for the answer set.

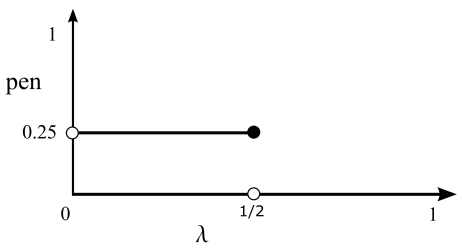


Figure 1: Example rule penalty

These penalties can easily be extended to complex preference expressions from a set $FPDL^{gp}$, defined analogously as $FPDL^p$ above, but starting from arbitrary preference rules. To define the semantics for expressions from $FPDL^{gp}$, we rely on the semantics of corresponding expressions from

ISBN: 978-989-95079-6-8

$FPDL^p$. Specifically, for e in $FPDL^{gp}$ and λ in $\mathcal{L}$, we define

$$pen(A, e)(\lambda) = pen^c(A, e^\lambda)$$

where e^λ is obtained from e by replacing all occurrences of preference rules r by the corresponding preference rules r_λ. Note that in principle, aggregation strategies that operate directly on gradual penalties could be defined as well.

Figure 2 depicts $pen(A, e)$ (the solid line) and $pen(A', e)$ (the dashed line) for an expression of the form $e = (sum\ dance, win)$ where the rule

$$dance : (sit(p, s) \rightarrow nearDanceFl(s)) \triangleleft nearFriends(p)$$

encodes that person p would like to sit near the dancefloor, but only finds this important if he is sitting close to his friends. The other rule is defined as

$$win : (sit(p, s) \rightarrow nearWin(s)) \triangleleft nearColleagues(p)$$

and denotes that person p likes a window seat, but only finds this important if he is sitting close to his colleagues. The answer sets A and A' take on the following applicability and suitability values:

app	$dance$	win
A	0.375	0.625
A'	0.75	0.5

$suit$	$dance$	win
A	0.18	0.82
A'	0.8	0.95

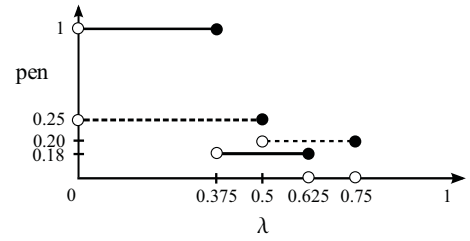


Figure 2: Example complex penalties

Now, given Figure 2, which answer set is preferred: A or A'? This depends on the application: if one only wants to look at the higher (more important) lambdas, we would prefer A as for $\lambda = 0.75$ this answer set has penalty 0; however if the global penalty score over all lambdas is important, we would prefer A' as A has a high penalty for values lower than 0.375 (viz. 1). Hence many options are viable, depending on the application. Furthermore there is a need to state preferences amongst expressions themselves so that we can e.g. state that only satisfying the $dance$ rule is better than only satisfying the win rule. In the next section we show how such orderings of answer sets based on the penalties and the preferences amongst preference expressions can be stated.

3.4 Preference strategies

Using penalties, we can define preference orderings on answer sets, which are referred to as *strategies*. Various such strategies can be specified. A natural strategy if we are comparing answer sets on the basis of one expression is to define the preference ordering $\leq$ between answer sets, based on an ordering $\preceq$ of gradual numbers. This strategy is denoted by $(\preceq\ e)$, and the corresponding preference ordering $\leq$ is defined for answer sets A_1 and A_2 as

$$A_1 \leq A_2 \equiv pen(A_1, e) \succeq pen(A_2, e)$$

Several partial and total orderings between gradual numbers can be used to this end. When the applicability values are interpreted as priorities, the total ordering $\preceq_1$ can be used, where $g_1 \preceq_1 g_2$ for gradual numbers g_1 and g_2 whenever

1. $g_1 = g_2$; or
2. $g_1(\lambda) < g_2(\lambda)$ for some λ in $]0, 1]$, and $g_1(\lambda') = g_2(\lambda')$ for all $\lambda' > \lambda$.

A useful partial ordering is the pointwise extension of the natural ordering $\leq$ on $\mathbb{R}$, denoted by $\preceq_2$, i.e. $g_1 \preceq_2 g_2$ iff $g_1(\lambda) \leq g_2(\lambda)$ for all λ in $]0, 1]$. Finally, an averaging strategy $\preceq_3$ could be used where $g_1 \preceq_3 g_2$ iff $\int_0^1 g_1(\lambda)w(\lambda)\,\mathrm{d}\lambda \leq \int_0^1 g_2(\lambda)w(\lambda)\,\mathrm{d}\lambda$. Intuitively, we may want to give more weight to the values of g_1 and g_2 for larger values of λ. This is encoded by using an increasing $]0, 1]$–$]0, 1]$ mapping w, defined, for instance, by $w(\lambda) = \lambda$ for all λ in $]0, 1]$.

As an example, consider Figure 2 with expression e from Section 3.3 again. From the picture one can see that if $\preceq_1$ is used, answer set A would be preferred over A' as for $\lambda = 0.75$ we have that $pen(A, e)(\lambda) < pen(A', e)(\lambda)$ and for any $\lambda > 0.75$ it holds that $pen(A, e)(\lambda) = 0 = pen(A', e)(\lambda)$. Hence $\preceq_1$ encodes a form of priority where we discriminate answer sets on their penalties for rules with a high applicability. Using $\preceq_2$, we would obtain that A and A' are incomparable. Thus $\preceq_2$ takes a more global approach, only considering preference when one of the two answer sets globally imposes a lower penalty score. With $\preceq_3$ and $w(\lambda) = \lambda$ for each $\lambda \in]0, 1]$ then, we obtain that A' would be preferred over A, due to the fact that A has a very high penalty at lower λ scores and is close to the penalty of A at the more important λ values. We can thus conclude that $\preceq_3$ can be used when one is not only interested in the fact that one penalty is higher than the other, but also in the exact value of these penalties.

The $\preceq_i$ $(1 \leq i \leq 3)$–strategies are not sufficient however: if we have two preference rules r_1 and r_2 and we want r_1 to be more important than r_2, we cannot readily encode this using a combination of the sum or max expressions and one of the orderings on the gradual penalties. To this end, we introduce other strategies, encoded as expressions from the set $FPDL$. First we have $(\preceq\ e) \in FPDL$ for $e \in FDPL^{gp}$ and $\preceq$ a partial ordering between gradual numbers. In analogy with Brewka [10], we also consider the following expressions:

1. If $e_1, \ldots, e_k$ are in $FPDL$, then $(pareto\ e_1, \ldots, e_k)$ and $(lex\ e_1, \ldots, e_k)$ are in $FPDL$
2. If $e_1, \ldots, e_k$ are in $FPDL^{gp}$ then $(rinc\ e_1, \ldots, e_k)$ and $(rcard\ e_1, \ldots, e_n)$ are in $FPDL$

The semantics of $(pareto\ e_1, \ldots, e_k)$ and $(lex\ e_1, \ldots, e_k)$ are defined analogously to the crisp case. In particular, the ordering $\leq$ induced by $(pareto\ e_1, \ldots, e_k)$ is defined by

$$A_1 \leq A_2 \equiv A_1 \leq_1 A_2 \wedge \ldots \wedge A_1 \leq_k A_2$$

ISBN: 978-989-95079-6-8

where $\preceq_i$ is the ordering induced by e_i. As an example, consider the expression $(pareto\ \ (\preceq_1\ \ dance), (\preceq_1\ \ win))$ with $dance$ and win as defined in Section 3.3. The preference ordering induced by this expression does not favor either of the expressions $(\preceq_1\ \ dance)$ and $(\preceq_1\ \ win)$.

Similarly, the ordering induced by $(lex\ e_1, \ldots, e_k)$ is given by

$$A_1 \leq A_2 \equiv \forall j \in 1..k \cdot A_1 \preceq_j A_2 \\ \vee \exists j \in 1..k \cdot A_1 \prec_j A_2 \wedge \forall j' < j \cdot A_1 \preceq_{j'} A_2$$

Expression $(lex\ (\preceq_1\ \ dance), (\preceq_1\ \ win))$ states that we first need to look whether we can prefer one of two answer sets A and A' by the expression $(\preceq_1\ \ dance)$. If this is not the case, we look whether $(\preceq_1\ \ win)$ can be used to prefer one of the two. Intuitively we thus prefer answer sets where person p is close to the dancefloor, even if that means he is sitting far from the window and close to his colleagues.

The $pareto$ and lex strategies both take arbitrary $FPDL$ expressions as their arguments. The next two strategies we introduce, $rinc$ and $rcard$, however, have expressions from $FPDL^{gp}$ as arguments. The difference with the $\preceq_i$ ($1 \leq i \leq 3$) strategies introduced earlier is that $rinc$ and $rcard$ base the preferences on more than one penalty expression and thus allow to have some interplay between rules and penalty expressions.

The semantics of $(rinc\ e_1, \ldots, e_k)$ and $(rcard\ e_1, \ldots, e_k)$ are defined in terms of the interval-valued fuzzy set P^a_A in the finite universe $\{1, \ldots, k\}$, where A is an answer set, $a \in [0, +\infty[$ and $L^a_A(i) = \{\lambda \mid pen(A, e_i)(\lambda) = a\}$:

$$P^a_A(i) = \begin{cases} [\inf L^a_A(i), \sup L^a_A(i)] & \text{if } L^a_A(i) \neq \emptyset \\ [0, 0] & \text{otherwise} \end{cases}$$

Intuitively, P^a_A is the set of (indices of) subexpressions e_i that attach a penalty a to an answer set A. Due to penalties being gradual numbers and because of the specific form of $pen(A, e)(\lambda)$ for any penalty expression e, P^a_A is an interval-valued fuzzy set.

The preference ordering induced by $(rinc\ \ e_1, \ldots, e_k)$ is defined as $A_1 \leq A_2$ iff

1. $P^a_{A_1} = P^a_{A_2}$ for all a in $[0, +\infty[$; or
2. $P^a_{A_1} \subset P^a_{A_2}$ for some a in $[0, +\infty[$, and $P^b_{A_1} = P^b_{A_2}$ for all $b < a$

where $\subset$ is an inclusion measure on interval-valued fuzzy sets, such as those proposed in [13]. As an example, consider $(rinc\ \ dance, win, boss)$ with $dance$ and win as in Section 3.4 and $boss$ the rule

$$boss : (sit(p, s) \rightarrow nearSpeechStand(s)) \triangleleft nearBoss(p)$$

Furthermore, we have answer sets A and A' with the following applicability and suitability values

app	r_1	r_2	r_3
A	0.25	0.5	0.5
A'	0.75	0.25	1

$suit$	r_1	r_2	r_3
A	0.5	0.5	0.25
A'	1	0.1	0.5

By definition of $rinc$ we first need to check whether we can discriminate between A and A' on the basis of P^0_A and $P^0_{A'}$. From the definition we easily obtain that:

P^0	r_1	r_2	r_3
A	$[0.25, 1]$	$[0.5, 1]$	$[0.5, 1]$
A'	$[0, 1]$	$[0.25, 1]$	$[0, 0]$

Note that $P^0_{A'}(1) = [0, 1]$ as the penalty of A' is 0 over the whole interval. Likewise $P^0_{A'}(3) = [0, 0]$ as the rule is fully applicable and penalty 0 is thus never reached. We use the inclusion measure $A \subseteq B \equiv \forall x \in X \cdot A(x) \leq B(x)$ for interval-valued fuzzy sets A and B over a universe X where $[a, b] \leq [a', b'] \equiv b \leq b' \wedge (b - a) \leq (b' - a')$ in this example, i.e. we take both the more important high λs and the amount of lambdas into account. By this definition $P^0_A(2) < P^0_{A'}(2)$ and $P^0_{A'}(3) < P^0_A(3)$ meaning that using $rinc$ the answer sets A and A' are incomparable.

For $(rcard\ e_1, \ldots, e_k)$, which uses a cardinality based approach, the ordering is given by $A_1 \leq A_2$ iff

1. $|P^a_{A_1}| = |P^a_{A_2}|$ for all a in $[0, +\infty[$; or
2. $|P^a_{A_1}| < |P^a_{A_2}|$ for some a in $[0, +\infty[$, and $|P^b_{A_1}| = |P^b_{A_2}|$ for all $b < a$

where $|U|$ denotes a measure of cardinality for the interval-valued fuzzy set U, such as those defined in [14]. Note that $rcard$ induces a partial order, whereas $rinc$ induces a total order.

As an example, consider $(rcard\ dance, win, boss)$, A and A' defined as before and a cardinality measure of the form $|A| = \sum_{x \in X} f(A(x))$ on an interval-valued fuzzy set A. When interval-valued membership degrees are used to encode uncertainty or bipolarity, $f_1([a, b]) = \frac{a+b}{2}$ is a natural choice. However, in our case, we need a notion of cardinality which is increasing in both b and $b - a$: a given penalty value is more representative for a penalty expression when it corresponds to more λ values, and to higher λ values. A suitable function is therefore given by $f_2([a, b]) = b(b - a)$. To further justify this choice, we provide the outcomes of both f_1 and f_2 in this example. For $rcard$ we obtain with f_1 that $|P^0_A| = \frac{4.25}{2} > \frac{2.25}{2} = |P^0_{A'}|$ whereas with f_2 we get $|P^0_A| = 1.75 = |P^0_{A'}|$, hence with f_1 answer set A would be preferred over A', whereas with f_2 no difference can be made by looking at penalty 0 and thus we must look for a higher penalty that does make a difference. The first penalty we should consider is 0.5 as this is the first penalty actually occurring for some rules. Computing $P^{0.5}$ gives:

$P^{0.5}$	r_1	r_2	r_3
A	$[0, 0.25]$	$[0, 0.5]$	$[0, 0]$
A'	$[0, 0]$	$[0, 0]$	$[0, 1]$

From this table we see that with f_2 we get $|P^0_A| = 0.25 * 0.25 + 0.5 * 0.5 = 0.3125$ and $|P^0_{A'}| = 1$, hence $A \leq A'$ and thus A' is preferred over A.

4 Translating Literal Preferences to a gFASP Program

In this section, we show that the gFASP framework we summarized in Section 2 is sufficiently general to encode $FPDL$–

ISBN: 978-989-95079-6-8

rules, an observation which immediately leads to an implementation method. Consider the gFASP program P, comprised of a rule base $\mathcal{R}_P$ and an aggregator expression $\mathcal{A}_P$ over the quasi-order $\mathcal{Q}_P$, and the expression e from $FPDL$. Specifically, we will show that a set of rules $\mathcal{R}_{P'}$ can be found, and an aggregator expression $\mathcal{A}_{P'}$ over a quasi-order $\mathcal{Q}_{P'}$, such that for any two answer sets A_1 and A_2

$$[\mathcal{A}_{P'}]_{A_1} \leq [\mathcal{A}_{P'}]_{A_2} \equiv [\mathcal{A}_P]_{A_1} \leq [\mathcal{A}_P]_{A_2} \wedge A_1 \leq_e A_2$$

Intuitively, $\mathcal{A}_P$ encodes which rules are more important, and how answer sets should be penalized when some rules are violated, while e encodes preferences between the actual solutions, i.e. preferences on the truth values of the literals.

As a first step, we need to provide the means to refer to the value of a certain function over literals $f(a_1, \ldots, a_n)$ from the aggregator expression. This is done by creating a constraint of the form $c_f : 0 \leftarrow \sim f(a_1, \ldots, a_n)$ where $\sim$ is an involutive negator and $\leftarrow$ is an implicator satisfying $0 \leftarrow x = \sim x$. It is easy to see that, in this way, the interpretation of rule proposition c_f, which we can refer to in the aggregator expression, will always be equal to the corresponding interpretation of $f(a_1, \ldots, a_n)$. Since the aggregator expression is increasing in all rule propositions, this procedure only works if we prefer answer sets with higher values of $f(a_1, \ldots, a_n)$. If, on the other hand, lower values of $f(a_1, \ldots, a_n)$ are preferred, we introduce the constraint $c_f : 0 \leftarrow f(a_1, \ldots, a_n)$ and use $\sim c_f$ in the aggregator.

Let $\mathcal{R}_e = \{r_1, \ldots, r_n\}$ be the set of preference rules occurring in e. Given that the preference of answer sets increases with increasing suitability of the rules and decreasing applicability, the aforementioned procedure leads to the rulebase $\mathcal{R}_{P'}$, defined by

$$\mathcal{R}_{P'} = \mathcal{R}_P \cup \{g_i : 0 \leftarrow \sim r_{ih} \mid i \in 1..n\} \cup \{f_i : 0 \leftarrow r_{ib}\}$$

The second part of our embedding consists of constructing a suitable aggregator expression $\mathcal{A}_{P'}$ and corresponding quasi-order $\mathcal{Q}_{P'}$. The aggregator $\mathcal{A}_{P'}$ is straightforwardly obtained from $\mathcal{A}_P$ by extending it with the various suitability and applicability values, i.e.:

$$\mathcal{A}_{P'} = (\mathcal{A}_P, (g_1, \ldots, g_n, \sim f_1, \ldots, \sim f_n))$$

The corresponding quasi-order is given by $\mathcal{Q}_{P'} = (\mathcal{Q}_P \times \mathcal{L}^{2n}, \leq_{\mathcal{Q}_{P'}})$ where

$$(a, b) \leq_{\mathcal{Q}_{P'}} (a', b') \equiv (a \leq_{\mathcal{Q}_P} a') \wedge (b \leq_e b')$$

Note how we can discriminate between answer sets on the basis of their rule fulfillment, using the first element of tuples from $\mathcal{Q}_{P'}$, as well as on the basis of the fulfillment of preference expression e, using the second element. The ordering $\leq_e$ is defined in the same way as the orderings on preference expressions, taking as arguments tuples with applicability and suitability scores, instead of answer sets.

The following proposition shows that the proposed embedding indeed preserves the intended semantics:

Proposition 1 *Let P be a gFASP program and e a fuzzy preference expression from FPDL. For the program P', obtained in the way described above, it holds for any two answer sets A_1 and A_2 that*

$$[\mathcal{A}_{P'}]_{A_1} \leq [\mathcal{A}_{P'}]_{A_2} \equiv [\mathcal{A}_P]_{A_1} \leq [\mathcal{A}_P]_{A_2} \wedge A_1 \leq_e A_2$$

5 Concluding remarks

In this paper, we generalized Brewka's answer set optimization framework to the fuzzy domain. The resulting formalism combines the flexibility of fuzzy answer set programming with the expressiveness of conditional preferences. We emphasized the conceptual difference between applicability and suitability of preference rules. This difference, which collapses in the two-valued case, played a crucial role in our fuzzification, leading to a generalization of penalty values using the notion of gradual numbers.

Although our extension is independent of any particular approach to fuzzy answer set programming, we showed that the recently introduced gFASP framework can provide a convenient implementation. One of the core elements in this approach is the use of an aggregator expression to discriminate between answer sets. We have shown how fuzzy preference expressions can be seen as particular choices of this aggregator expression. Hence, fuzzy preference expressions should be regarded as a convenient way to encode complex quasi-orders between answer sets, in a truely declarative way, rather than as a fundamental extension to this earlier work.

References

[1] C. Baral. *Knowledge Representation, Reasoning and Declarative Problem Solving*. Cambridge University Press, 2003.

[2] M. Gelfond and V. Lifschitz. The stable model semantics for logic programming. In *Proc. of ICLP1988*, pages 1081–1086. MIT Press, 1988.

[3] T. Lukasiewicz and U. Straccia. Tightly integrated fuzzy description logic programs under the answer set semantics for the semantic web. In *Proc. of RR2007*, pages 289–298. Springer-Verlag, 2007.

[4] D. Van Nieuwenborgh, M. De Cock, and D. Vermeir. An introduction to fuzzy answer set programming. *Annals of Mathematics and Artificial Intelligence*, 50(3-4):363–388, 2007.

[5] T. Lukasiewicz. Probabilistic logic programming. In *Proc. of ECAI1998*, pages 388–392. J. Wiley and Sons, 1998.

[6] U. Straccia. Annotated answer set programming. In *Proc. of IPMU2006*, 2006.

[7] C.V. Damásio, J. Medina, and M. Ojeda-Aciego. Sorted multi-adjoint logic programs: termination results and applications. In *Proc. of JELIA2004*, pages 260–273. Springer-Verlag, 2004.

[8] C. V. Damásio and L. M. Pereira. Antitonic logic programs. In *Proc. of LPNMR2001*, pages 379–392. Springer-Verlag, 2001.

[9] U. Straccia. Managing uncertainty and vagueness in description logics, logic programs and description logic programs. In *Reasoning Web, 4th International Summer School, Tutorial Lectures*, number 5224, pages 54–103. Springer Verlag, 2008.

[10] G. Brewka. Complex preferences for answer set optimization. In *Proc. of KR2004*, pages 213–223, 2004.

[11] J. Janssen, S. Schockaert, D. Vermeir, and M. De Cock. General fuzzy answer set programs. Accepted for WILF2009.

[12] J. Fortin, D. Dubois, and H. Fargier. Gradual numbers and their application to fuzzy interval analysis. *Fuzzy Systems, IEEE Transactions on*, 16(2):388–402, 2008.

[13] C. Cornelis and E.E. Kerre. Inclusion measures in intuitionistic fuzzy set theory. In *Proc. of ECSQARU2003*, pages 345–356, 2003.

[14] G. Deschrijver and P. Král'. On the cardinalities of interval-valued fuzzy sets. *Fuzzy Sets and Systems*, 158(15):1728–1750, 2007.

ISBN: 978-989-95079-6-8

Two Results About Optimization of Fuzzy Variable Functions.

Andrés David Báez-Sánchez
Ph.D. Student
Department of Applied Mathematics
Institute of Mathematics, Statistics and Scientific Computation, Campinas State University
Campinas, SP. , Brazil
Email: davidbaez@ime.unicamp.br

Abstract— *We discuss some optimization problems for fuzzy variable functions and show two interesting results. First result is related to conditions for existence of global optimal solutions for some general class of optimization problems on $\mathbb{E}^n$, the space of compacts, upper semicontinuous and normal fuzzy sets of $\mathbb{R}^n$. The second result is about preservation of local optimality points, via Zadeh's principle of extension.*

Keywords— Coercitivity of fuzzy variable functions, compacity in fuzzy metric spaces, fuzzy optimization, optimization of fuzzy variable functions, principle of extension.

1 Introduction

In most optimization problems, the coefficients, parameters and variable decisions involved are assumed to be precise elements of some real space $\mathbb{R}^n$. However in many real situations there is not such certainty about those values and is necessary to consider some models or methods to deal with optimization problems under uncertainty. Two main approaches have been deeply developed when we consider uncertainty in the parameters, in the constraints or in the values of the objective function, these are stochastic programming and fuzzy optimization. We may refer [1] or [2] for stochastic programming and [3], [4] or [5] for an overview of fuzzy optimization.

By the other hand, in connection with applications and modelling using fuzzy theory, a different kind of fuzzy optimization problems has been appeared in which some of the decision variables are itself fuzzy, this means, optimization of fuzzy variable functions. Some situations like this can be found in problems of depend or constraints programming as is shown in [6] or [7]), in estimation and regression models for fuzzy random variables as in [8], [9], [10], [11] or [12], or in problems of optimal control for fuzzy differential equations as in [13] or [14].

In most of the cases those fuzzy variable optimization problem was faced successfully using some *ad hoc* techniques or properties, some heuristic algorithm, or considering just specific classes of fuzzy numbers such as triangular or L-R fuzzy numbers. We believe that fuzzy variable optimization problems have not been thoroughly studied yet, and they deserve a systematic study in itself, because a general approach to this issue can leave a fertile field for their application and theoretical developments.

In this work we are dealing with fuzzy quantities in $\mathbb{R}^n$, specifically we are considering $\mathbb{E}^n$ the spaces of compacts, upper semicontinuous and normal fuzzy sets of $\mathbb{R}^n$. It is well known that $\mathbb{E}^n$ with usual arithmetic between fuzzy numbers is not a vector space, so many of general results in optimization theory are not availables immediately on this context, precisely because most of them use the vector space structure. Taking advantage of the special characteristics and properties of $\mathbb{E}^n$ we will be able to show two interesting theoretical results in optimization of fuzzy variable functions.

In section 2 we set the two problems under consideration and we recall some basic notation and definitions. In section 3 we recall some results about compacity in $\mathbb{E}^n$. Section 4 presents our first main result related to conditions for existence of optimal global solutions for some general class of optimization problems on $\mathbb{E}^n$. In section 5 we recall the Zadeh's principle of extension and some related properties and in section 6 we present our second main result related to preservation of local optimality points, via Zadeh's principle of extension.

2 Preliminaries

We are going to consider first the following optimization problem:

$$\min f(u) \qquad \text{subject to} \qquad u \in D \tag{1}$$

with $D \subset \mathbb{E}^n$ and $f : \mathbb{E}^n \to \mathbb{R}$. Later we will consider a given function $f : \mathbb{R}^n \to \mathbb{R}$ and study the problem of optimal local points of the function $\hat{f} : \mathbb{E}^n \to \mathbb{E}$ defined by Zadeh's principle of extension from f.

Besides of the space $\mathbb{E}^n$ we will also use the metric space $K(\mathbb{R}^n)$ of compacts subsets of $\mathbb{R}^n$ with the Hausdorff metric h defined by $h(A,B) = \max\{d(A,B), d(B,A)\}$ where A, B are compacts subsets of $\mathbb{R}^n$, $d(A,B) = \sup_{a\in A} d(a,B)$ and $d(a,B) = \inf_{b\in B} d(a,b)$.

Let u be an element of $\mathbb{E}^n$. For every $\alpha \in (0,1]$ the α-level of u is defined by $u_\alpha = \{x \in \mathbb{R}^n : u(x) \geq \alpha\}$ and $u_0 = supp(u) = \overline{\{x \in \mathbb{R}^n : u(x) > 0\}}$.

ISBN: 978-989-95079-6-8

The supremum metric d_∞ on $\mathbb{E}^n$ is defined in terms of Hausdorff metric between α-level sets:

$$d_\infty(u,v) = \sup_{\alpha\in[0,1]} h(u_\alpha, v_\alpha),$$

where $u, v \in \mathbb{E}^n$.

3 Compacity in $\mathbb{E}^n$

Definition 3.1. *A set $U \subset \mathbb{E}^n$ is said to be **compact-supported** if there is a compact subset K of $\mathbb{R}^n$ such that $supp(u) \subset K$ for every $u \in U$.*

If a subset U of $\mathbb{E}^n$ is compact-supported then is bounded. We explain this in the next theorem to be used later.

Theorem 3.2. *Let U be a subset of $\mathbb{E}^n$. If U is not compact-supported then exist a sequence $\{u^k\} \subset U$ such that $||u^k|| := d_\infty(u^k, 0) \to \infty$ as $k \to \infty$.*

Proof: If U is not compact-supported then for every compact set $K \subset \mathbb{R}^n$ there exist some $u^K \in U$ such that $supp(u^K) \not\subset K$. If we take $K = \overline{B_k(0)}$ the closed ball with center 0 an radius k, then for each $k \in \mathbb{N}$ there is a $u^k \in U$ such that $supp(u^k) \not\subset \overline{B_k(0)}$.

We have that $d(0, \overline{B_k(0)}) = 0$ and $d(\overline{B_k(0)}, 0) = k$ so $h(0, \overline{B_k(0)}) = k$, and by the other hand we have that at least one element x of $supp(u^k)$ is not an element of $\overline{B_k(0)}$ so $d(x, 0) > k$ and this implies that $d(supp(u^k), 0) = d((u^k)_0, 0) > k$ and furthermore $h((u^k)_0, 0) > k$.

By definition of supremum metric we have that

$$d_\infty(u^k, 0) = \sup_{\alpha\in[0,1]} h((u^k)_\alpha, 0) > k,$$

so, $||u^k|| := d_\infty(u^k, 0) \to \infty$ as $k \to \infty$ □

We consider also the level-set mapping $u_{(\cdot)} : [0,1] \to K(\mathbb{R}^n)$ which associate with each α the level-set u_α.

Definition 3.3. *A set $D \subset \mathbb{E}^n$ is called **equicontinuous** if the family of level-sets functions $\{u_{(\cdot)} : u \in D\}$ is a family of equicontinuous functions of [0,1] in $K(\mathbb{R}^n)$.*

The following is a well known result of metric spaces writen in terms of our fuzzy optimization problem:

Theorem 3.4. *Let D be a compact subset of $\mathbb{E}^n$ and f a continuous function on D. Then exists an optimal global solution for problem 1.*

An also well known corollary it follows, but first we need an additional notation. For $c \in \mathbb{R}$ and $f : D \to \mathbb{R}$ the set $L_{f,D}(c)$ is defined by

$$L_{f,D}(c) := \{x \in D : f(x) \leq c\}$$

Corollary 3.5. *If f is continuous and exist $c \in \mathbb{R}$ such that $L_{f,D}(c)$ is not empty and compact then exists an optimal global solution for problem 1.*

Proof: If $L_{f,D}(c)$ is not empty and compact then previous theorem implies that the function f reaches a minimum at some $x^* \in L_{f,D}(c)$ when f is restricted to $L_{f,D}(c)$. By definition if $x \notin L_{f,D}(c)$ then $f(x) > c$. But $f(x^*) \leq c < f(x)$ so x^* is an optimal global solution of problem 1 □

So far what we have just write are old compacity results. Those results can be easy to understand but maybe hard to apply because until now we have not talked about the meaning of compacity on $(\mathbb{E}^n, d_\infty)$.

An interesting characterizations of compacity in $(\mathbb{E}^n, d_\infty)$ in terms of concepts previously defined is the following.

Theorem 3.6. *A subset D of $(\mathbb{E}^n, d_\infty)$ is relatively compact if D is compact-supported and equicontinuous.*

Proof: See Corollary 10 in [15] □

Corollary 3.7. *If a subset D of $(\mathbb{E}^n, d_\infty)$ is closed, compact-supported and equicontinuous then D is compact.*

4 A Coercitivity Condition

In this section we will prove our first main result. We need to recall some definitions.

Definition 4.1. *A sequence $\{x^k\} \subset \mathbb{E}^n$ is called critic in relation to a set D if $\{x^k\} \subset D$ and $||x^k|| := d_\infty(0, x^k) \to \infty$ or $x^k \to x$ with $x \in \overline{D}$ but $x \notin D$.*

Definition 4.2. *A function f is coercive in a set D if for all sequence $\{x^k\}$ critics in relation to D, $\limsup_{k\to\infty} f(x^k) = +\infty$.*

Previous ones are well known definitions on normed vector spaces but even if $\mathbb{E}^n$ is not exactly a vector space, this concepts can be used in this context because of the special characteristics of $\mathbb{E}^n$. Let's see how this can be done.

Theorem 4.3. *If f is continuous and coercive in D and exist $c \in \mathbb{R}$ such that $L_{f,D}$ is not empty and equicontinuous, then problem 1 has an optimal global solution.*

Proof: We will show that under this hypothesis $L_{f,D}(c)$ is compact and then we use corollary 3.5.

ISBN: 978-989-95079-6-8

If $L_{f,D}(c)$ is not compact then because corollary 3.7 and hypothesis, $L_{f,D}(c)$ is not closed or is not compact-supported.

If $L_{f,D}(c)$ is not closed, as we assume that $L_{f,D}(c)$ is not empty, then exist a sequence $\{u^k\} \subset L_{f,D}(c)$ such that $u^k \to u$ and $x \notin L_{f,D}(c)$. If $u \in D$ then continuity of f in D and the fact that $f(u^k) \leq c$ for all $k = 0, 1, ...$, allow us to conclude $f(u) \leq c$ but $u \notin L_{f,D}(c)$ by definition of u, so $u \notin D$.

This means that $\{u^k\}$ is a critic sequence in relation to D and by coercitivity of f in D we have that $f(u^k) \to \infty$, which is a contradiction because $f(u^k) \leq c$.

Now, if $L_{f,D}(c)$ is not compact-supported then for theorem 3.2 there is a sequence $\{u^k\} \subset L_{f,D}(c)$ such that $||u^k|| := d_\infty(u^k, 0) \to \infty$ and again because of coercitivity of f we conclude that $f(u^k) \to \infty$ which is again a contradiction because $f(u^k) \leq c$.

So $L_{f,D}(c)$ has to be compact and the desired result it follows from corollary 3.5 □

5 Zadeh's Principle of Extension and Ordering in $\mathbb{E}$

We recall now for the case of $f : \mathbb{R}^n \to \mathbb{R}$ one of the cornerstones of fuzzy theory.

Zadeh's Principle of Extension: Given a function $f : \mathbb{R}^n \to \mathbb{R}$ we define the Zadeh's extension of f by:

$$\hat{f} : \mathbb{E}^n \to \mathbb{E},$$

$$\hat{f}(u)(x) = \begin{cases} \sup_{z \in f^{-1}(x)} u(z), & \text{if } f^{-1}(x) \neq \emptyset, \\ 0, & \text{if } f^{-1}(x) = \emptyset, \end{cases}$$

for each $u \in \mathbb{E}^n$.

Next theorem is now a well known result and establishes conditions on f in order to $\hat{f}$ be a well-defined and continuous function.

Theorem 5.1. *If $f : \mathbb{R}^n \to \mathbb{R}$ is continuous then $\hat{f} : (\mathbb{E}^n, d_\infty) \to (\mathbb{E}, d_\infty)$ is a well-defined continuous function and*

$$[\hat{f}(u)]_\alpha = f(u_\alpha),$$

for all $u \in \mathbb{E}^n$ and $\alpha \in [0, 1]$.

Proof: Proof of this result is an easy adaptation of theorems 2.1 and 3.3 at [16] and a famous result of Nguyen at [17] □

A usual partial order on $\mathbb{E}^n$ is defined by α-cuts inclusion in the following way:

$$u \leq v \Leftrightarrow u_\alpha \subseteq v_\alpha \quad \text{for all } \alpha \in [0, 1]],$$

with $u, v \in \mathbb{E}^n$.

Nevertheless, even that this order has been used succesfully for theoretical developments as in [18], it is still a little unsatisfactory at least in the case $n = 1$ because in this case this order is not a generalization of usual order on $\mathbb{R}$. In fact, number 2 is not longer greater than 1, they are not even comparable using this order.

There is, however, a generalization of natural order on $\mathbb{R}$ that we are going to consider now. Before, is necessary to note that if $u, v \in \mathbb{E}$ this implies that all their α-cuts are bounded closed interval of real line, this means that for all $\alpha \in [0, 1]$

$$u_\alpha = [a_\alpha, b_\alpha], v_\alpha = [c_\alpha, d_\alpha]$$

with $a_\alpha, b_\alpha, c_\alpha, d_\alpha \in \mathbb{R}$.

Definition 5.2. *Given $u, v \in \mathbb{E}$ and $\alpha \in [0, 1]$ we consider $u_\alpha = [a_\alpha, b_\alpha]$ and $v_\alpha = [c_\alpha, d_\alpha]$. We define a partial order relation on $\mathbb{E}$ by:*

$$u \leq_1 v \Leftrightarrow \quad a_\alpha \leq c_\alpha \quad \textit{and} \quad b_\alpha \leq d_\alpha$$

for all $\alpha \in [0, 1]$.

It is easy to show that previous order is in fact a partial order on $\mathbb{E}$ and that it is indeed a generalization of natural order on $\mathbb{R}$.

6 Local Optimality Preservation

Now we are going to prove our second main result. We are still considering the order defined in definition 5.2 and the supremum metric on $\mathbb{E}^n$

Theorem 6.1. *If $f : \mathbb{R}^n \to R$ is continuous and has a local minimum (maximum) at x^* then $\hat{f} : (\mathbb{E}^n, d_\infty) \to (\mathbb{E}, d_\infty, \leq_1)$ has also a local minimum (maximum) at x^*.*

Proof: We will just make the proof for local minimum. The other part is identical.

If f has a local minimum at x^* then exist a closed ball $B_r(x^*) \subset \mathbb{R}^n$ such that $f(x^*) \leq f(x)$ for all $x \in B_r(x^*)$.

Because $\hat{f}$ is just an extension of f to $\mathbb{E}^n$ and the order $\leq_1$ is a generalization or usual order of $\mathbb{R}$ to $\mathbb{E}$ then obviously $\hat{f}(x^*) \leq_1 \hat{f}(x)$ for all $x \in B_r(x^*)$ so we need to find a neighborhood of x^* in $(\mathbb{E}^n, d_\infty)$ with the same property.

We consider simply the closed ball $\tilde{B}_r(x^*) \subset (\mathbb{E}^n, d_\infty)$ with the same radius and center than the previous one, this means

$$\begin{aligned} \tilde{B}_r(x^*) =& \{u \in \mathbb{E}^n : d_\infty(u, x^*) \leq r\} \\ =& \{u \in \mathbb{E}^n : \sup_{\alpha \in [0,1]} h(u_\alpha, x^*) \leq r\} \\ =& \{u \in \mathbb{E}^n : \forall \alpha \in [0, 1] \quad h(u_\alpha, x^*) \leq r\} \\ \tilde{B}_r(x^*) =& \{u \in \mathbb{E}^n : \forall \alpha \in [0, 1] \quad u_\alpha \subset B_r(x^*), \} \end{aligned} \tag{2}$$

ISBN: 978-989-95079-6-8

and we have that $B_r(x^*) \subset \tilde{B}_r(x^*)$ when elements of $B_r(x^*)$ are considered as fuzzy as well.

If $u \in \tilde{B}_r(x^*)$, we have that for all $\alpha \in [0,1]$, $u_\alpha \subset B_r(x^*)$ so $f(u_\alpha) \subset f(B_r(x^*))$ and by definition of $B_r(x^*)$ this implies that for all $\alpha \in [0,1]$, $f(x^*) \leq y$ for all $y \in f(u_\alpha)$.

By theorem 5.1 we have that $[\hat{f}(u)]_\alpha = f(u_\alpha)$ so, for all $\alpha \in [0,1]$

$$f(x^*) \leq y \quad \text{for all} \quad y \in [\hat{f}(u)]_\alpha. \tag{3}$$

If we write $[\hat{f}(u)]_\alpha$ as an interval, this means, $[\hat{f}(u)]_\alpha = [a_\alpha, b_\alpha]$ for some $a_\alpha, b_\alpha \in \mathbb{R}$ the previous expression in the specific case of a_α and b_α becomes

$$\forall \alpha \in [0,1], \quad f(x^*) \leq a_\alpha \quad \text{and} \quad f(x^*) \leq b_\alpha. \tag{4}$$

Because $f(x^*) = \hat{f}(x^*)$ and because $f(x^*)$ considered as a fuzzy set has α-level sets always equals to $[f(x^*), f(x^*)]$, the previous considerations and the expression 4 implies that for $u \in \tilde{B}_r(x^*)$

$$\hat{f}(x^*) \leq_1 \hat{f}(u),$$

so $\hat{f}$ has a local minimum at x^*. □.

7 Conclusions and Future Work

In this work has been shown that a classical result of vector space optimization about coercitivity can be extended also for optimization problems on $\mathbb{E}^n$, even if this spaces is not a vector space. We believe that many other results can be generalized in this way and we will continue to study this subject and its applications.

This kind of problem can be seen also as general metric spaces optimization problems, a subject that is lately getting attention again (See [19]), however because of the fact that $\mathbb{E}^n$ is indeed more than just a metric space (after all it has some arithmetic structure) we will try to take advantage of its particular structure to obtain better results and applications.

We also believe that there are interesting questions about the way that a fuzzy extension of a function and orderings are related to optimal points and others qualitative characteristics of a function. We will keep study this subject deeply and we hope to find interesting interpretations and future applications.

Acknowledgment

I want to thank my advisor professor Marko Rojas-Medar and my dear colleague Johanna Garzón. This work was supported by CNPq, National Council for Scientific and Technological Development - Brazil.

References

[1] J.R. Birge, F. Louveaux, *Introduction to Stochastic Programming*, Springer-Verlag, (1997).

[2] A. Prékopa, *Stochastic Programming*, Kluwer Academic Publishers, (1995).

[3] C.R. Bector, S. Chandra, Fuzzy Mathematical Programming and Fuzzy Matrix Games,*Studies in Fuzziness and Soft Computing*, Springer-Verlag, Berlin, (2005).

[4] R. Slowinski (Ed.)A. Prkopa, *Fuzzy Sets in Decision Analysis, Operations Research and Statistics*, (The Handbooks of Fuzzy Sets) Springer (1998).

[5] Y.J. Lai, C.L. Hwang, Fuzzy Multiple Objective Decision Making: Methods and Applications, *Lecture Notes in Economics and Mathematical Systems* 404, Springer-Verlag, (1994).

[6] J. Gao, J. Zhao, X. Ji, Fuzzy Chance-Constrained Programming for Capital Budgeting Problem with Fuzzy Decisions, *L. Wang and Y. Jin (Eds.): FSKD 2005, LNAI 3613*, pp. 304-311, Springer-Verlag Berlin Heidelberg, (2005).

[7] B. Liu, Dependent-chance programming with fuzzy decisions, *IEEE Transactions on Fuzzy Systems*, 7, 354-360, (1999).

[8] P. Diamond, R. Körner, Extended Fuzzy Linear Models and Least Squares Estimates, *Computers Math. Applic.* Vol. 33, No. 9, pp. 15-32, (1997).

[9] R. Körner, W. Näther, Linear regression with random fuzzy variables: extended classical estimates, best linear estimates, least squares estimates, *Journal of Information Sciences* 109 95–118, (1998).

[10] A. Wünsche, W. Näther, Least-squares fuzzy regression with fuzzy random variables, *Fuzzy Sets and Systems* 130, 43-50, (2005).

[11] A. Colubi, et. al., A determination coefficient for fuzzy random variables in a fuzzy arithmetic-based linear model. *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE, London, United Kingdom)*, pp. 473-476, (2007).

[12] G. González-Rodríguez, et. al., Estimation of a simple linear regression model for fuzzy random variables, *Fuzzy Sets and Systems* 160, 357-370, (2009).

[13] P. Diamond, P. Kloeden, Optimization under uncertaintly. *Bouchon-Meunier, B., Zadeh, L.A., Yager, R.R. (eds.) IPMU 1990. LNCS 521*, pp. 247-249. Springer, Heidelberg (1991)

ISBN: 978-989-95079-6-8

[14] Y.C. Kwun, et. al., The Existence of Fuzzy Optimal Control for the Semilinear Fuzzy Integrodifferential Equations with Nonlocal Conditions, *B.-Y. Cao (Ed.): Fuzzy Information and Engineering (ICFIE), ASC*(40) , pp. 657665, (2007).

[15] Greco, G.H., Moschen M.P.: Supremum metric and relatively compact sets of fuzzy sets. *Nonlinear Analysis* **64**, 1325–1335 (2006).

[16] H. Román-Flores, et. al., A note on Zadeh's Extension, *Fuzzy Sets And Systems* 117, 327–331, (2002).

[17] H.T. Nguyen, A note on the extension principle for fuzzy sets, *J. Math. Anal.* Appl. 64, 369–380, (1978).

[18] M.A. Rojas-Medar, et.al., Fuzzy quasilinear spaces and applications, *Fuzzy Sets and Systems* 152 2, 173–190, (2005).

[19] M. McAsey, L. Mou L, A multiplier rule on a metric spaces, *J. Math. Anal. Appl.*, 337, 1064–1071 (2008)

[20] A. Izmailov, M. Solodov, *Otimização Vol. 1* IMPA, (2005).

[21] P. Diamond, P. Kloeden *Metric Spaces of Fuzzy Sets: Theory and Applications* World Scientific Publishing Co., Inc., River Edge, (1994).

ISBN: 978-989-95079-6-8

Industrial Monitoring by Evolving Fuzzy Systems

Lara Aires[1], João Araújo[2], António Dourado[1]

[1] Center for Informatics and Systems, Department of Informatics Engineering,
University of Coimbra, Polo II 3030-290 Coimbra Portugal
[2] Galp Sines Refinery, Process Control,
Sines, Portugal
Email: lccaires@gmail.com, dourado@dei.uc.pt, joao.araujo@galpenergia.com

Abstract *— Industrial monitoring of complex processes with hundreds or thousands of variables is a hard task faced in this work through evolving fuzzy systems. The Visbreaker process of the Sines Oil Refinery is the case studied. Firstly dimension reduction is performed by multidimensional scaling, obtaining the process evolution in a three dimensional space. Then an evolving fuzzy system (eFS) is developed to detect eventual malfunction of sensors. This eFS takes the data in the three reduced dimensions as antecedents and classifies the process state into normal and abnormal states. A software platform- the eFSLab (Evolving Fuzzy Systems Laboratory) - , with which this work has been developed, is presented and discussed. Several strategies for rule creation and evolution of rules, for Takagi-Sugeno (and for Mamdani system obtained from these) , are implemented in eFSLab. The obtained eFS shows a promising performance in the case studied, classifying in some simulations the state of the process into abnormal-normal condition in about 95% of the cases, with a number of rules between 5 and 8.*

Keywords — Evolving fuzzy systems; multidimensional scaling; industrial monitoring; fuzzy logic software.

1 Introduction

The growing complexity of industrial processes, the market competition, the more strict quality and safety rules, enhance the needs for sophisticated maintenance functions, including online process monitoring, failure detection and failure diagnosis [1] [11] [12]. This turns to be a complex problem mainly because of its non-linear high dimensional characteristics, since industrial systems are described using a large number of variables, which can be determined by various parameters and measurements.

According to [2] the main problem is to find the clusters of states in order to reflect the measurements and determine the general behavior of the system. Fuzzy systems may play here an important role, if iterative, real time techniques for rule base creation and evolution can be found. The eFSLab, Evolving Fuzzy Systems Laboratory, is a computational framework with that aim, briefly presented in this work.

The Visbreaker process of the Sines Refinery (Galp) is used to obtain real data for testing the performance of eFSLab to classify the state of the process into abnormal-normal condition. This classification is done through fuzzy systems produced by an evolving algorithm which is implemented in the platform.

An analysis is made of the performance of the fuzzy systems as classifier for this problem concluding that this algorithm has the potential to support maintenance activities as a failure detector.

The paper is organized as follows. Section 2 reviews briefly the eFS algorithm and in section 3 the eFSLab environment will be described. In section 4 the results obtained with eFSLab for data from Sines Refinery are presented and in section 5 these results are discussed and some conclusions and directions for future work are proposed.

2 The eFS algorithm

The algorithm to create Takagi-Sugeno fuzzy models is based on the approach proposed by Angelov et al. for online learning [3] and improved by [4] [5]. This algorithm allows creating zero or first orders Takagi-Sugeno fuzzy models with global or local recursive estimation parameters.

Fuzzy models are created with a data-driven approach, composed of an iterative process with 7 main stages [3] .

The procedure begins with the initialization of the rule-base structure (antecedent part of the rules) with a single rule, based on the first data sample. The second stage is to read the next data sample. Then, the potential of each new data sample is recursively calculated and in stage 4 occurs a recursively update of the potentials of the focal points (centers) of the existing rules/clusters.

After this, the potential of the new data sample is compared with the updated potential of existing centers, in order to decide if the rule-base is modified or upgraded. In the next stage consequence's parameters are recursively updated by Recursive Least Squares (RLS) or weighted RLS for globally or locally optimal parameters, respectively. Finally, the last stage is to predict the output of the system. By reading the next data sample at the next time step, the procedure returns to stage 2.

The algorithm intends to approach the subtractive clustering technique applied in a batch onset. Since it is to be applied in real-time environments, the data appears successively and at each time the new available data is used to improve the fuzzy system, in such a way that the new informative content of data should be reflected in the fuzzy rules, whether by change some of them, whether by creating new rules if there is sufficiently novelty in the process behavior. There are here some scientific challenges. The main one is how to

ISBN: 978-989-95079-6-8

measure the informative novelty in the new data. The second is how to change/create the rules consequently. The proposed techniques by [3] and [4] are based basically on heuristics and experimentation and as a consequence are case dependent. In the present work the same approach is followed. More conditions for rule update/create have been investigated and implemented in the computational framework eFSLab.

eFSLab allows to develop a Mamdani fuzzy model from a zero order Takagi-Sugeno (TS) fuzzy model. The Mamdani system is made based on the same antecedents of TS system previously created and by centering the fuzzy sets of Mamdani consequents in the Takagi-Sugeno constant consequent (only zero order models can be considered).

3 The eFSLab interface

The eFSLab interface is presented in Fig. 1(at the end of the paper). It was intended to create an interface as complete as possible, where the user could set the great majority of parameters needed to create a Takagi-Sugeno fuzzy system and transform it into a Mamdani system (in case of zero order TS systems), in a simple way.
The first step to use eFSLab interface is to select the file to import data. This file can be an Excel or a text file.
Then the user can choose TS order (zero or first), the recursive parameter estimations method, and set some other parameters needed to the recursive process.
Conditions for creation or modification of rules can also be chosen or defined by the user. There is a set of predefined conditions or the user can build up his (her) own conditions, as shown in Fig. 2

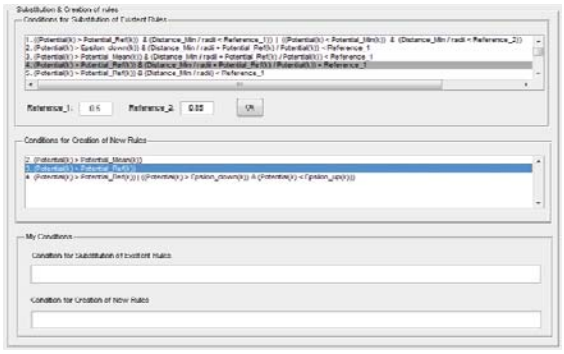
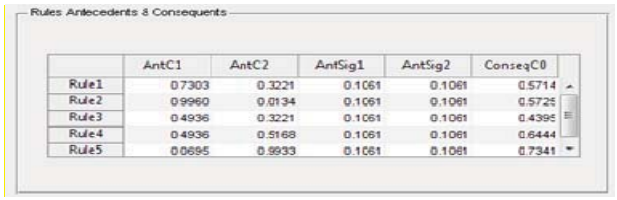


Figure 2. Creating or modifying rules in eFSLab.

The fuzzy system is created and its characteristics can be analysed using the Matlab Fuzzy Logic Toolbox, with all its functionalities (viewing the graphics, consulting antecedents and consequents fuzzy sets, etc). A diagnosis text file is created by eFSLab. The created rules are shown in a table in the interface, shown in Figure 3 for the case of two antecedents, Gaussian membership functions and five rules.



Rules Antecedents & Consequents

	AntC1	AntC2	AntSig1	AntSig2	ConseqC0
Rule1	0.7303	0.3221	0.1061	0.1061	0.5714
Rule2	0.9960	0.0134	0.1061	0.1061	0.5725
Rule3	0.4936	0.3221	0.1061	0.1061	0.4395
Rule4	0.4936	0.5168	0.1061	0.1061	0.6444
Rule5	0.0695	0.9933	0.1061	0.1061	0.7341

Figure 3. Table containing antecedents and consequents for each rule created by eFS procedure.

eFSLab produces a set of graphical information about the evolution of the TS systems, such as the number of rules, the instants of modification or creation of rules, the potential evolution of the data samples, the model output versus real output for validation data, the norm of covariance matrix and rules parameters evolution, the obtained cluster in input space. Some of them are shown in figures below.
The algorithm is based on recursive subtractive clustering [3]. At the end of the learning phase, the collection of points that has been used, together with the obtained clusters, can be observed as in Fig. 4.

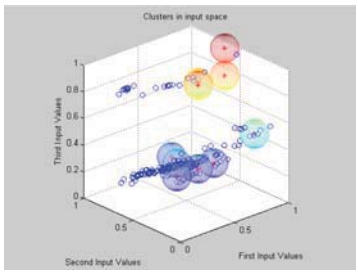


Figure 4. Clusters in input space after training. The dimensions of the spheres depend on the parameters of the clusters in a subtractive clustering environment.

In an on-line (iterative) implementation, the new rules can be created or the existent rules can be modified, and the instant when that happens can be shown as in Fig. 5

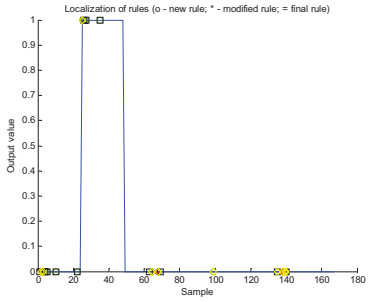


Figure 5. Localization of rules (new rules, rules modified and rules deleted).

The evolution of the number of rules during the process can be visualized as in Fig. 6.
To transform a zero-order TS fuzzy system into a Mamdani one, the *SugenoToMamdani* interface is opened as Fig. 7.
The user can choose the type of membership functions, the correspondent parameters and can preview their shape.

ISBN: 978-989-95079-6-8

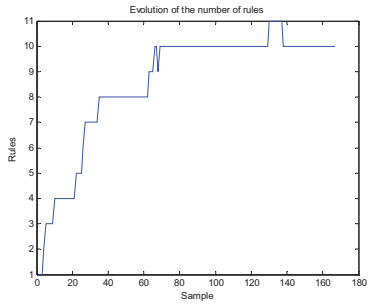


Figure 6. Evolution of the number of rules.

The Mamdani fuzzy system produced can be opened in the Matlab Fuzzy Logic Toolbox, benefiting from its functionalities for a deeper analysis .

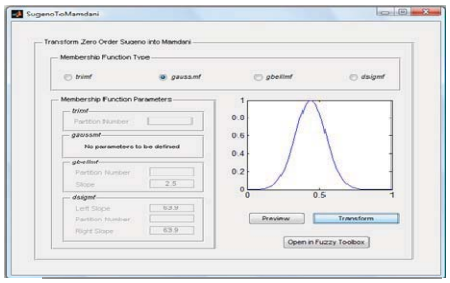


Figure 7.The *SugenoToMamdani* interface.

4 Applicatiom to the Visbreaker process

The Visbreaker Unit is intended to reduce the viscosity of the residual coming from the vacuum column. With this objective a thermal cracking process is used with a relatively low temperature, and a long residence time. As a result of the thermal process a low viscosity Visbreaker residual is obtained, as well as lighter products, such as hydrocarbonets (gas oil diesel, gasoline and gases).The great economical advantage of the Visbreaker process is in fact that it produces a residual with a lower viscosity that the load feed. By this way, it is possible to use a lower quantity of "cutterstocks" (some of them of high benefit) for the production of fuel oil.
Fig. 8 [9] presents its flow sheet. It is composed of several sub-processes, and its main part is a kiln operating at about 310 ºC.

Data averaged for each hour, between the 1st of January and the 5th of June of 2008 was collected. Actually, data from 160 tags is available. From these, after correlation analysis and process expertise, 59 were selected as sufficiently representative of the process. Multidimensional scaling is then applied to those 59 dimensions to obtain a three dimensional representation by using VisRed [6]. The three dimensional data is then fed to eFSLab that finds a zero-order TS fuzzy model.

Multidimensional scaling reduces an *n*-dimensional data set to *p*-dimensional data set, with $p<<n$, such that the distance (for example Euclidian) between each pair of points in the reduced space is similar to the distance in the high dimensional space. This means that the topology of the original data set is preserved, as well as the information embedded in the topology. The technique consists in the minimization of the Euclidian distance between two matrices: the dissimilarity matrix in the original space and the dissimilarity matrix in the reduced space. The (*i,j*) element of a dissimilarity matrix is the distance between points *i* and *j*.

Three simulated scenarios have been considered with increasing time horizon and number of faults: one week with one fault in one sensor (Case 1), two weeks with two faults in two sensors (Case 2), and one month with two sensors faults during one day (Case 3). In Case 1 a 20% reduction in data from one temperature sensor in the kiln, during one day, was introduced into the data between the 1st and the 6th of June 2008. In Case 2 a 20% reduction in data from two temperature sensors in the kiln, during one day, was introduced into the data between the 22nd of May and the 4th of June 2008. In Case 3, the process data between the 5th of May and the 5th of June 2008 was changed in the same way in the same two sensors. These scenarios allow to measure the diagnosis capability of the eFS, as a classifier, as the size of data set increase.

For each test, there are some parameters of eFSLab that had to be set in order to produce the fuzzy system. The first parameter to be set is the order of the TS fuzzy model (zero) and the parameter estimation (Global estimation) for all the three cases. Other parameters controlling the rule base evolutions were fixed at their default values in eFSLab.
Radii value (controlling the region of influence of each cluster in input space) changed from test to test in order to evaluate its influence on the process. The conditions for creation and modification of rules were chosen to have the capability to produce rules in regions with less data points, improving the classification performance.
Some results of the three cases are presented in the following. In Case 1 and Case 3was introduced perturbation both in test and train data sets. In Case 2 the perturbation was only introduced in test data set

Case 1. Data is from the 1st to the 6th of June 2008 (one week). In train data set it was added a reduction of 20% to data from the 2nd of June and in test data set the same perturbation was added to data from the 4th of June, in a temperature sensor.
Dimension reduction by multidimensional scaling was performed with cosine dissimilarity metric without normalization. Fig. 9 represents data after dimension reduction. Two different groups corresponding to normal and abnormal data can be observed.
The reduced data is then presented to eFSLab, obtaining a zero order TS system with 8 rules to classify the data into two classes 1 and 2. A Radii value of 0.3 was used. These rules can be visualized in Fig. 10 (using Matlab Fuzzy Logic Toolbox).
In Fig. 11 shows the real output and the fuzzy system output It can be seen that the fault was detected. This classifier shows here 95% of successful classification, although there

ISBN: 978-989-95079-6-8

are also some false faults detected, reducing the performance of the system.

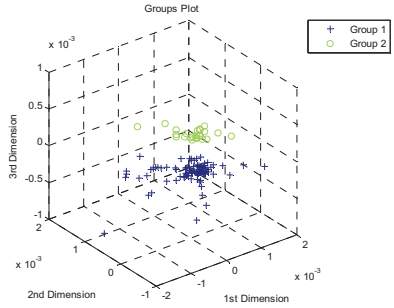


Figure 9. The three-dimensional state space for Case 1. Blue points (Group 1) are normal operating point. Group 2 are points of a simulated sensor fault.

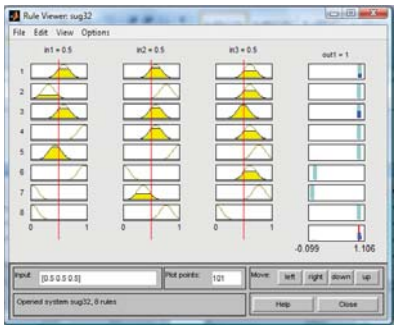


Figure 10. The 8 rules Takagi-Sugeno eFS obtained to classify the points in Case 1.

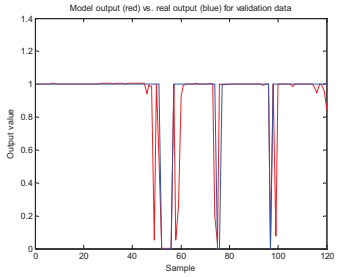


Figure 11. Real output (blue) and simulation output (red) obtained with zero-order TS system created for Case 1.

Case 2. Data from 22nd of May to 4th of June 2008 (two weeks) is considered, with a perturbation of 20% in data from two sensors in 23 of May and 4 of June.
Dimension reduction was made with Euclidean dissimilarity metric without normalization. Fig. 12 is a three dimensional representation of the reduced data.
Giving to eFSLab the reduced data, with Radii of 0.1, a 10 rules zero order TS system was obtained, shown in Fig. 13.

Case 3. Now all the data from 5th of May to the 5th of June (one month) is considered. Perturbation is the same as in the Case 2, but in 20th of May and 4th of June. Also the Radii value and the dissimilarity metric for dimension reduction are the same as in Case 2. In Fig. 15 is represented data set after dimension reduction.
Fig.14 compares the real output with the fuzzy system output. The classification is successful in 98.8% of the points.
The obtained zero-order TS system has now 11 rules, presented in Fig. 16.

Fig. 17 compares the output of the fuzzy system with the prepared data. The performance of the created fuzzy system is now substantially decreased: although the fault was detected, so have been many false positives. Overall 80.80% of the points have been correctly classified.

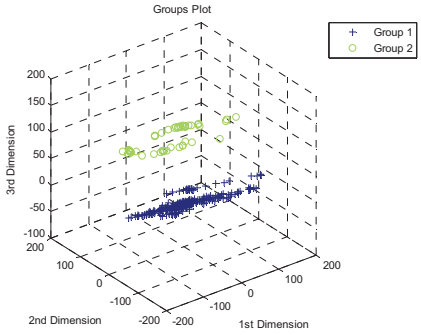


Figure 12: The three-dimensional state space representation of the Visbreaker process in the last week of May and the first week of June 2008. Blue points are normal operating point. Group 2 are points of a simulated sensor malfunction situation.

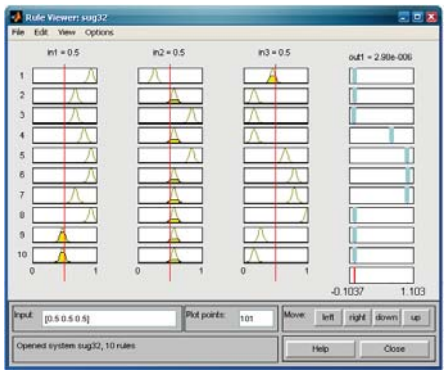


Figure 13: The 10 rules Takagi-Sugeno eFS obtained to classify the points in Case 2.

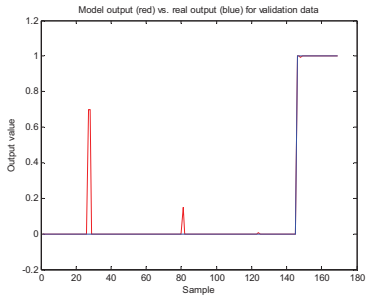


Figure 14. Real output (blue) and simulation output (red) obtained with fuzzy system created for Case 2.

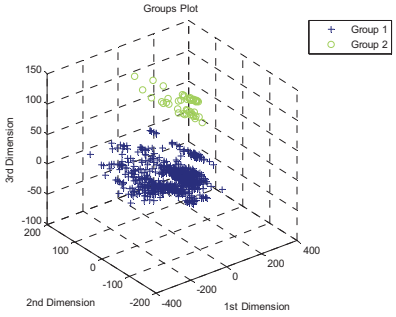


Figure 15: The three-dimensional state space representation of the Visbreaker process during the 5th of May and the 5th of June 2008. Blue points are normal operating point. Group 2 are points of a simulated sensor malfunction situation.

ISBN: 978-989-95079-6-8

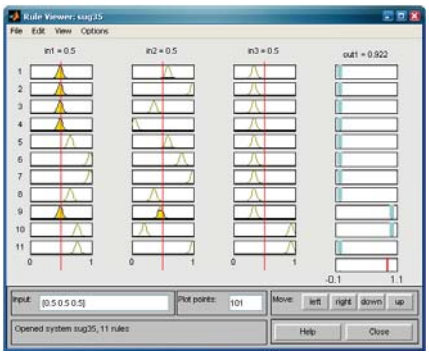


Figure 16: The 11 rules Takagi-Sugeno eFS for Case 3.

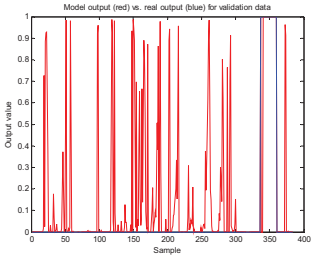


Figure 17. Case 3 :real output (blue) and fuzzy system output in Case 3.

Table 1 presents a summary of the results.

Table 1: Summary of results for the three cases.

Test	Data Set Size	MD Scaling Dissimilarity Metric	Radii	Output Matching (%)
1	1 week	Cosine	0.3	95%
2	2 weeks	Euclidean	0.1	98.8%
3	1 month	Euclidean	0.1	80.80%

When the time horizon increases (the data sets size increases proportionally), the performance of the fuzzy system as classifier seems to be worse. The way how the rules are created probably needs improvements. This is also supported by the distribution of the final clusters in the input space (see for example Fig. 4), that does not cover all the relevant space. However the results show a promising performance of the fuzzy system in terms of output (classification) matching. Furthermore, the low number of rules produced in each test contributes to a lower fuzzy sets superposition, increasing eventually interpretability and transparency, which is a main concern in fuzzy modelling [7] [8] [9]. Further work is needed to improve the iterative clustering technique, looking for better approximations to the subtractive clustering method.

5 Conclusions

Development of fuzzy systems from data is still an unattained aim of the fuzzy logic communities. Accuracy, interpretability and transparency are characteristics with an important practical relevance, allowing to give sense and credibility to the fuzzy systems. In the case of industrial monitoring, in big complexes of basic process industry, they may play an important role supporting the decision making in daily operation of the plants. *eFSLab* can be a useful computational framework to this goal. It is built with friendly user interfaces and allows an easy configuration and application to any data set built in proper form.

Evolving fuzzy systems, where rules are based on clusters, need more elaborated techniques for iterative clustering methods. This is not an easy task, since here neighbour points do not appear at the same time. Some kind of manageable memory must be developed to approach the iterative results to the non-iterative ones. This is one research direction of the authors.

New ways to measure novelty in data is needed. In a real time set the size of datasets tends to infinity. It is possible to maintain in memory only a finite time window of data Deleted points are forgotten history and new points may simply repeat pas behaviour but leading to new rules that are not only unnecessary but also prejudicial. This is another challenge to future work. The fuzzy community can have here a rich field of research to make fuzzy systems a credible and useful tool for data mining and knowledge discovery, whether in batch, whether in real-time.

Acknowledgments. This work was supported by Project CLASSE POSC/EIA/58162/2004 including EU FEDER Support.

References

[1] N. Palluat *et al*. A neuro-fuzzy monitoring system application to flexible production systems. *Computers in Industry*, 57:528-538, 2006.

[2] O. Simula *et al*, Monitoring and Modeling of Complex Processes using Hierarchical Self-Organization Maps. Helsinki University of Technology, Laboratory of Computer and Information Science. Espoo, Finland. 1996.

[3] P. Angelov and D. Filev. An Approach to Online Identification of Takagi-Sugeno Fuzzy Models. *IEEE Transactions on Systems, Man, and Cybernetics - Part B*, 34(1), 2004.

[4] J. Victor and A. Dourado, "Evolving Takagi-Sugeno fuzzy models," Technical Report, CISUC, (2003, September). Available: http://cisuc.dei.uc.pt/acg/view_pub.php?id_p=760.

[5] J. V. Ramos and A. Dourado, "On line interpretability by rule base simplification and reduction," Eunite Symposium, Aachen, 2004.

[6] Dourado, A. and Ferreira, E. and Barbeiro, P. , "VISRED - Numerical Data Mining with Linear and Nonlinear Techniques", in Advances in Data Mining. Theoretical Aspects and Applications,, pp. 93-106, LNAI, Vol. 4597, Springer, Industrial Conference on Data Mining 2007, Leipzig, Germany. Available for download from

[7] Mikut, R., J. Jäkel, L. Gröll "Interpretability issues in data-based learning of fuzzy systems", Fuzzy Sets and Systems , vol. 150, 179–197, 2005.

[8] C. Mencar and A.M. Fanelli. Interpretability constraints for fuzzy information granulation. *Information Sciences*, 178:4585–4618, 2008.

[9] González, J., I.Rojas, H. Pomares, L. J. Herrera, A. Guillén, J. M. Palomares, F. Rojas, "Improving the accuracy while preserving the interpretability of fuzzy function approximators by means of multi-objective evolutionary algorithms", International Journal of Approximate Reasoning ,Volume 44 , Is.1, , pp32-44, 2007.

[10] Galp (2006), Sines Refinery Flow Sheets.

[11] Filev D. P. and F. Tseng, Novelty Detection Based Machine Health Prognostics, Proc. of the 2006 International Symposium on Evolving Fuzzy Systems, Lake District, UK, pp. 193--199, 2006

[12] Lughofer, E. and C. Guardiola, On-Line Fault Detection with Data-Driven Evolving Fuzzy Models. Control and Intelligent Systems, Vol. 36 (4), pp. 307-317, 2008.

ISBN: 978-989-95079-6-8

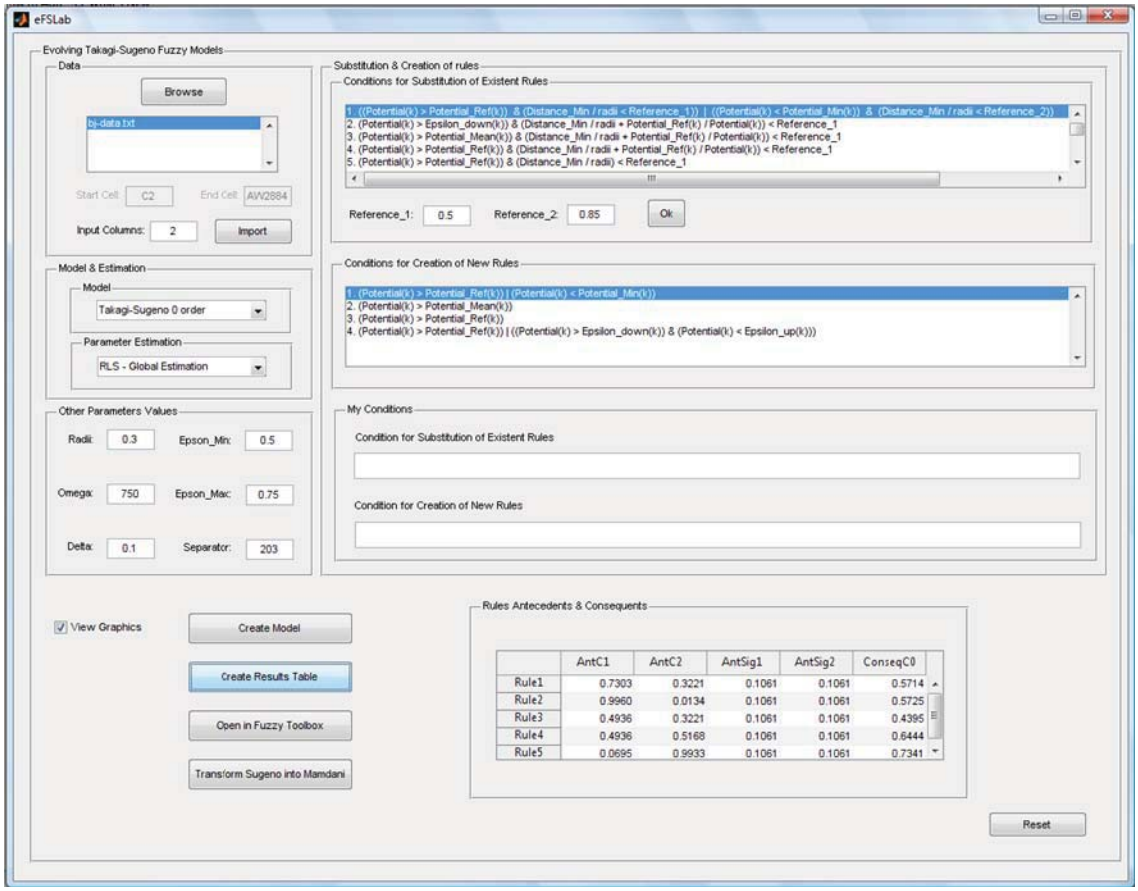


Figure 1: *eFSLab* interface.

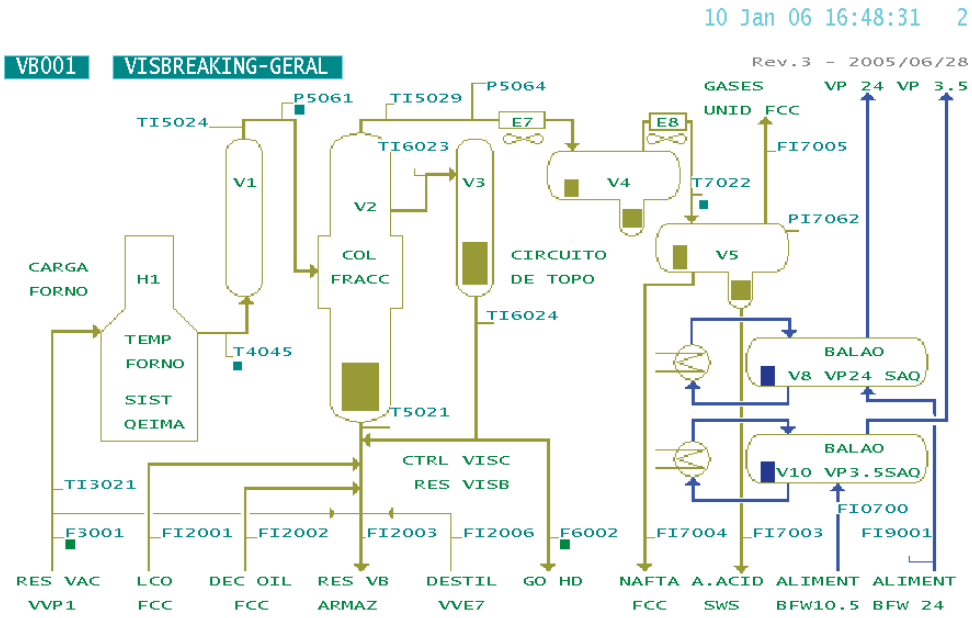


Figure 8: The Visbreaker Process [10].

ISBN: 978-989-95079-6-8

Probabilistic Decision Making with the OWA Operator and its Application in Investment Management

José M. Merigó

Department of Business Administration, University of Barcelona
Av. Diagonal 690, 08034 Barcelona, Spain
Email: jmerigo@ub.edu

Abstract—*We develop a new model for decision making under risk environment and under uncertainty. We introduce a new aggregation operator that unifies the probabilities and the ordered weighted averaging (OWA) operator in the same formulation. We call it the probabilistic ordered weighted averaging (POWA) operator. This aggregation operator provides a more complete representation of the decision problem because it is able to consider probabilistic information and the attitudinal character of the decision maker. We study different properties and families of the POWA operator. We also develop an illustrative example of the new approach in a decision making problem about selection of investments.*

Keywords— Decision making; OWA operator; Probabilities; Investment selection.

1 Introduction

Decision making problems [2-8,14,16,18] are very common in the literature. There are different ways and methods for solving the decision process. Usually, the method used for solving the decision problem depends on the available information. For example, when the decision maker has probabilistic information he will solve the problem calculating the expected value for each alternative. Thus, he will be in a problem of decision making under risk environment. However, in other problems the decision maker may not have probabilistic information. Therefore, he must use another approach for solving the problem such as the use of the ordered weighted averaging (OWA) operator [13] that reflects the attitudinal character (degree of optimism) of the decision maker. In this case, we are in a situation of decision making under uncertainty.

The OWA operator is a very useful technique for aggregating the information. It provides a parameterized family of aggregation operators between the maximum and the minimum. In decision making it is very useful for representing the attitudinal character of the decision maker. Since its appearance, the OWA operator has been studied and applied in a wide range of problems [1-4,6-7,9-18].

In [4], they presented a new model that tried to unify the concept of probability with the OWA operator. They introduced the immediate probabilities. This method formulates an aggregation operator that uses probabilities and OWA operators at the same time. Therefore, this approach permits to consider decision making problems under risk environment and under uncertainty at the same time. The concept of immediate probabilities has been further studied in [6-7,16]. It is very useful in some particular situations. However, further analysis (as those shown in the paper) show that it can only unify probabilities and OWAs giving a neutral degree of importance to each case. But the real situation is that sometimes the decision maker believes more on the probabilistic information or on the OWAs. Therefore, the real unification must consider the degree of importance of these to concepts in the aggregation process. The aim of this paper is to present a new formulation that unifies the probabilities and the OWA operators considering the degree of importance of each case in the aggregation process. We call this new approach as the probabilistic ordered weighted averaging (POWA) operator. The main advantage of this approach is that it provides more complete information of the decision process because we can unify the probabilistic information and the attitudinal character according to the available information in the decision process. We study some of its main properties and we see that the OWA operator and the usual probabilistic decision making are particular cases of this approach. We study other particular cases of the POWA operator such as the Hurwicz-POWA, the olympic-POWA and the S-POWA.

We study the applicability of the POWA operator and we see that it is incredibly broad because it can be used in most of the problems where it appears the probability. In this paper, we focus on a decision making problem about selection of investments. We see the usefulness of the POWA operator because we can use situations of risk and uncertainty at the same time.

This paper is organized as follows. In Section 2 we briefly review the OWA operator and the immediate probability. Section 3 presents the POWA operator and Section 4 different particular cases. In Section 5 we present a decision making problem with the POWA operator in investment selection and in Section 6 we give the main conclusions.

2 Preliminaries

2.1 The OWA operator

The OWA operator [13] is an aggregation operator that provides a parameterized family of aggregation operators between the minimum and the maximum. It can be defined as follows.

ISBN: 978-989-95079-6-8

Definition 1. An OWA operator of dimension n is a mapping OWA: $R^n \to R$ that has an associated weighting vector W of dimension n such that $w_j \in [0, 1]$ and $\sum_{j=1}^{n} w_j = 1$, then:

$$\text{OWA}(a_1, \ldots, a_n) = \sum_{j=1}^{n} w_j b_j \quad (1)$$

where b_j is the jth largest of the a_i.
Note that different properties can be studied such as the distinction between descending and ascending orders, different measures for characterizing the weighting vector and different families of OWA operators. Note that it is commutative, monotonic, bounded and idempotent. For further reading, refer, e.g., to [1-4,6-7,9-18].

2.2 Immediate probabilities

An immediate probability is a type of probability that includes the attitudinal character of the decision maker by using the OWA operator in the aggregation process. Therefore, it is an attemp to unify the probabilities with the OWA operator. It can be defined as follows.

Definition 2. An IPOWA operator of dimension n is a mapping IPOWA: $R^n \to R$ that has an associated weighting vector W of dimension n such that $w_j \in [0, 1]$ and $\sum_{j=1}^{n} w_j = 1$, according to the following formula:

$$\text{IPOWA}(a_1, \ldots, a_n) = \sum_{j=1}^{n} \hat{v}_j b_j \quad (2)$$

where b_j is the jth largest of the a_i, each argument has associated a probability v_i with $\sum_{i=1}^{n} v_i = 1$ and $v_i \in [0, 1]$, $\hat{v}_j = (w_j v_j / \sum_{j=1}^{n} w_j v_j)$ and v_j is the probability v_i ordered according to b_j, that is, according to the jth largest of the a_i.
Note that the IPOWA operator is a good approach for unifying probabilities and OWAs in some particular situations. But it is not always useful, especially in situations where we want to give more importance to the OWA operators or to the probabilities. One way to see why this unification does not seem to be a final model is considering other ways of representing $\hat{v}_j$. For example, we could also use $\hat{v}_j = [w_j + v_j / \sum_{j=1}^{n} (w_j + v_j)]$ or other similar approaches.

3 The probabilistic OWA operator

The probabilistic ordered weighted averaging (POWA) operator is an extension of the OWA operator for situations where we find probabilistic information. It can also be seen as a unification between decision making problems under uncertainty (with OWA operators) and under risk (with probabilities). This approach seems to be complete, at least as an initial real unification between OWA operators and probabilities.

However, note that some previous models already considered the possibility of using OWA operators and probabilities in the same formulation. The main model is the concept of immediate probability explained in Section 2 [4,6-7,16]. Although it seems to be a good approach it is not so complete than the POWA because it can unify OWAs and probabilities in the same model but it can not take in consideration the degree of importance of each case in the aggregation process. Other methods that could be considered are the hybrid averaging (HA) operator and the weighted OWA (WOWA) operator. These methods are focused on the weighted average (WA) but it is easy to extend them to probabilities because sometimes the WA is used as a subjective probability. As said before, these an other approaches are useful for some particular situations but they does not seem to be so complete than the POWA because they can unify OWAs with probabilities (or with WAs) but they can not unify them giving different degrees of importance to each case. Note that in future research we will also prove that these models can be seen as a special case of a general POWA operator (or its respective model with WAs) that uses quasi-arithmetic means. Obviously, it is possible to develop more complex models of the IP-OWA, the HA and the WOWA that takes into account the degree of importance of the OWAs and the probabilities (or WAs) in the model but they seem to be artificial and not a natural unification as it will be shown below.
In the following, we are going to analyze the POWA operator. It can be defined as follows.

Definition 3. A POWA operator of dimension n is a mapping POWA: $R^n \to R$ that has an associated weighting vector W of dimension n such that $w_j \in [0, 1]$ and $\sum_{j=1}^{n} w_j = 1$, according to the following formula:

$$\text{POWA}(a_1, \ldots, a_n) = \sum_{j=1}^{n} \hat{v}_j b_j \quad (3)$$

where b_j is the jth largest of the a_i, each argument a_i has an associated probability v_i with $\sum_{i=1}^{n} v_i = 1$ and $v_i \in [0, 1]$, $\hat{v}_j = \beta w_j + (1 - \beta) v_j$ with $\beta \in [0, 1]$ and v_j is the probability v_i ordered according to b_j, that is, according to the jth largest of the a_i.
Note that it is also possible to formulate the POWA operator separating the part that strictly affects the OWA operator and the part that affects the probabilities. This representation is useful to see both models in the same formulation but it does not seem to be as a unique equation that unifies both models.

Definition 4. A POWA operator is a mapping POWA: $R^n \to R$ of dimension n, if it has an associated weighting vector W, with $\sum_{j=1}^{n} w_j = 1$ and $w_j \in [0, 1]$ and a probabilistic vector V, with $\sum_{i=1}^{n} v_i = 1$ and $v_i \in [0, 1]$, such that:

ISBN: 978-989-95079-6-8

$$\text{POWA}(a_1, \ldots, a_n) = \beta \sum_{j=1}^{n} w_j b_j + (1-\beta) \sum_{i=1}^{n} v_i a_i \qquad (4)$$

where b_j is the jth largest of the arguments a_i and $\beta \in [0, 1]$.
In the following, we are going to give a simple example of how to aggregate with the POWA operator. We consider the aggregation with both definitions.

Example 1. Assume the following arguments in an aggregation process: (20, 40, 50, 30). Assume the following weighting vector W = (0.2, 0.2, 0.2, 0.4) and the following probabilistic weighting vector P = (0.4, 0.2, 0.3, 0.1). Note that the probabilistic information has a degree of importance of 60% while the weighting vector W a degree of 40%. If we want to aggregate this information by using the POWA operator, we will get the following. The aggregation can be solved either with (3) or (4). With (3) we calculate the new weighting vector as:

$$\hat{v}_1 = 0.4 \times 0.2 + 0.6 \times 0.3 = 0.26$$
$$\hat{v}_2 = 0.4 \times 0.2 + 0.6 \times 0.2 = 0.2$$
$$\hat{v}_3 = 0.4 \times 0.2 + 0.6 \times 0.1 = 0.14$$
$$\hat{v}_4 = 0.4 \times 0.4 + 0.6 \times 0.4 = 0.4$$

And then, we calculate the aggregation process as follows:

POWA = 0.26 × 50 + 0.2 × 40 + 0.14 × 30 + 0.4 × 20 = 33.2.

With (4), we aggregate as follows:

POWA = 0.4 × (0.2 × 50 + 0.2 × 40 + 0.2 × 30 + 0.4 × 20) + 0.6 × (0.4 × 20 + 0.2 × 40 + 0.3 × 50 + 0.1 × 30) = 33.2.

Obviously, we get the same results with both methods.
From a generalized perspective of the reordering step, it is possible to distinguish between the descending POWA (DPOWA) and the ascending POWA (APOWA) operator by using $w_j = w^*_{n-j+1}$, where w_j is the jth weight of the DPOWA and w^*_{n-j+1} the jth weight of the APOWA operator.
If B is a vector corresponding to the ordered arguments b_j, we shall call this the ordered argument vector and W^T is the transpose of the weighting vector, then, the POWA operator can be expressed as:

$$\text{POWA}(a_1, \ldots, a_n) = W^T B \qquad (5)$$

Note that if the weighting vector is not normalized, i.e., $W = \sum_{j=1}^{n} w_j \neq 1$, then, the POWA operator can be expressed as:

$$\text{POWA}(a_1, \ldots, a_n) = \frac{1}{W} \sum_{j=1}^{n} \hat{v}_j b_j \qquad (6)$$

The POWA is monotonic, commutative, bounded and idempotent. It is monotonic because if $a_i \geq u_i$, for all a_i, then, POWA(a_1, a_2,…, a_n) ≥ POWA(u_1, u_2…, u_n). It is commutative because any permutation of the arguments has the same evaluation. That is, POWA(a_1, a_2,…, a_n) = POWA(u_1, u_2,…, u_n), where (u_1, u_2,…, u_n) is any permutation of the arguments (a_1, a_2,…, a_n). It is bounded because the POWA aggregation is delimitated by the minimum and the maximum. That is, Min$\{a_i\}$ ≤ POWA(a_1, a_2,…, a_n) ≤ Max$\{a_i\}$. It is idempotent because if $a_i = a$, for all a_i, then, POWA(a_1, a_2,…, a_n) = a.
Another interesting issue to analyze are the measures for characterizing the weighting vector W. Following a similar methodology as it has been developed for the OWA operator [7,13] we can formulate the attitudinal character, the entropy of dispersion, the divergence of W and the balance operator. Note that these measures affect the weighting vector W but not the probabilities because they are given as some kind of objective information.

4 Families of POWA operators

First of all we are going to consider the two main cases of the POWA operator that are found by analyzing the coefficient β. Basically, if $\beta = 0$, then, we get the probabilistic approach and if $\beta = 1$, the OWA operator.
By choosing a different manifestation of the weighting vector in the POWA operator, we are able to obtain different types of aggregation operators. For example, we can obtain the probabilistic maximum, the probabilistic minimum, the probabilistic average and the probabilistic weighted average.

Remark 1. The probabilistic maximum is found when $w_1 = 1$ and $w_j = 0$ for all $j \neq 1$. The probabilistic minimum is formed when $w_n = 1$ and $w_j = 0$ for all $j \neq n$.

Remark 2. More generally, the step-POWA is formed when $w_k = 1$ and $w_j = 0$ for all $j \neq k$. Note that if $k = 1$, the step-POWA is transformed to the probabilistic maximum, and if $k = n$, the step-POWA becomes the probabilistic minimum operator.

Remark 3. The probabilistic average is obtained when $w_j = 1/n$ for all j, and the probabilistic weighted average is obtained when the ordered position of i is the same as the ordered position of j.

Remark 4. For the median-POWA, if n is odd we assign $w_{(n+1)/2} = 1$ and $w_{j^*} = 0$ for all others. If n is even we assign for example, $w_{n/2} = w_{(n/2)+1} = 0.5$ and $w_{j^*} = 0$ for all others.

Remark 5. For the weighted median-POWA, we select the argument b_k that has the kth largest argument such that the sum of the weights from 1 to k is equal or higher than 0.5 and the sum of the weights from 1 to $k - 1$ is less than 0.5.

Remark 6. The olympic-POWA is generated when $w_1 = w_n = 0$, and for all others $w_{j^*} = 1/(n - 2)$. Note that it is possible

ISBN: 978-989-95079-6-8

to develop a general form of the olympic-POWA by considering that $w_j = 0$ for $j = 1, 2, \ldots, k, n, n-1, \ldots, n-k+1$, and for all others $w_{j^*} = 1/(n-2k)$, where $k < n/2$. Note that if $k = 1$, then this general form becomes the usual olympic-POWA. If $k = (n-1)/2$, then this general form becomes the median-POWA aggregation. That is, if n is odd, we assign $w_{(n+1)/2} = 1$, and $w_{j^*} = 0$ for all other values. If n is even, we assign, for example, $w_{n/2} = w_{(n/2)+1} = 0.5$ and $w_{j^*} = 0$ for all other values.

Remark 7. Note that it is also possible to develop the contrary case, that is, the general olympic-POWA operator. In this case, $w_j = (1/2k)$ for $j = 1, 2, \ldots, k, n, n-1, \ldots, n-k+1$, and $w_j = 0$, for all other values, where $k < n/2$. Note that if $k = 1$, then we obtain the contrary case for the median-POWA.

Remark 8. Another interesting family is the S-POWA operator. It can be subdivided into three classes: the "or-like," the "and-like" and the generalized S-POWA operators. The generalized S-POWA operator is obtained if $w_1 = (1/n)(1-(\alpha+\beta)) + \alpha$, $w_n = (1/n)(1-(\alpha+\beta)) + \beta$, and $w_j = (1/n)(1-(\alpha+\beta))$ for $j = 2$ to $n-1$, where $\alpha, \beta \in [0, 1]$ and $\alpha + \beta \leq 1$. Note that if $\alpha = 0$, the generalized S-POWA operator becomes the "and-like" S-POWA operator, and if $\beta = 0$, it becomes the "or-like" S-POWA operator.

Remark 9. Another family of aggregation operator that could be used is the centered-POWA operator. We can define a POWA operator as a centered aggregation operator if it is symmetric, strongly decaying and inclusive. Note that these properties have to be accomplished for the weighting vector W of the OWA operator but not necessarily for the weighting vector P of the probabilities. It is symmetric if $w_j = w_{j+n-1}$. It is strongly decaying when $i < j \leq (n+1)/2$ then $w_i < w_j$ and when $i > j \geq (n+1)/2$ then $w_i < w_j$. It is inclusive if $w_j > 0$. Note that it is possible to consider a softening of the second condition by using $w_i \leq w_j$ instead of $w_i < w_j$. We shall refer to this as softly decaying centered-POWA operator. Another particular situation of the centered-POWA operator appears if we remove the third condition. We shall refer to it as a non-inclusive centered-POWA operator.

Remark 10. Another type of aggregation that could be used is the E-Z POWA weights. In this case, we should distinguish between two classes. In the first class, we assign $w_{j^*} = (1/q)$ for $j^* = 1$ to q and $w_{j^*} = 0$ for $j^* > q$, and in the second class, we assign $w_{j^*} = 0$ for $j^* = 1$ to $n-q$ and $w_{j^*} = (1/q)$ for $j^* = n-q+1$ to n.

Remark 11. A further interesting type is the non-monotonic-POWA operator. It is obtained when at least one of the weights w_j is lower than 0 and $\sum_{j=1}^{n} w_j = 1$. Note that a key aspect of this operator is that it does not always achieve monotonicity. Therefore, strictly speaking, this particular case is not a POWA operator. However, we can see it as a particular family of operators that is not monotonic but nevertheless resembles a POWA operator.

Remark 12. Note that other families of POWA operators could be used following the recent literature about different methods for obtaining OWA weights. Some of these methods are explained in [1,7,9,15,18].

5 Decision making with the POWA operator in investment selection

The POWA operator is applicable in a wide range of situations where it is possible to use probabilistic information and OWA operators. Therefore, we see that the applicability is incredibly broad because all the previous models, theories, etc., that uses the probability can be extended by using the POWA operator. The reason is that all the problems with probabilities deal with uncertainty. Usually, in most of the problems it is assumed a neutral attitudinal character against the probabilistic information but we are still under uncertainty. Thus, sometimes we may prefer to be more or less optimistic against the probabilistic information. Note that this problem can be proved by looking to utility theory. In this theory, people prefer more safety results independently that the expected results (obtained with probabilities) give better results to a more risky alternative. For example, people prefer 100.000 euros with probability 1 instead of 90% probability of obtaining 1.000.000 euros and 10% probability of obtaining 0 euros. If we calculate the expected result (with probabilities) we get that the first alternative gives 100.000 euros and the second one 900.000 euros. However, most of the people will prefer the first alternative because there is no risk involved. This problem has been solved by the economists by using utility theory but as we can see, the use of the POWA operator is also useful for solving this type of problems.

Summarizing some of the main fields where it is possible to apply the POWA operator, we can mention:

- Statistics (especially in probability theory).
- Mathematics
- Economics
- Decision theory
- Engineering
- Physics
- Etc.

In this paper, we focus on an application in decision making about selection of investments. The main reason for using the POWA operator is that we are able to assess the decision making problem considering probabilities and the attitudinal character of the decision maker. Thus, we get a more complete representation of the decision problem.

In the following, we present a numerical example of the new approach in investment selection. We analyze a company that operates in Europe and North America that wants to invest some money in a new market. They consider five alternatives.

ISBN: 978-989-95079-6-8

- A_1 = Invest in the Asian market.
- A_2 = Invest in the South American market.
- A_3 = Invest in the African market.
- A_4 = Invest in all three markets.
- A_5 = Do not invest money in any market.

In order to evaluate these investments, the investor has brought together a group of experts. This group considers that the key factor is the economic situation of the world economy for the next period. They consider 5 possible states of nature that could happen in the future: S_1 = Very bad economic situation, S_2 = Bad economic situation, S_3 = Regular economic situation, S_4 = Good economic situation, S_5 = Very good economic situation.
The results of the available investments, depending on the state of nature S_i and the alternative A_k that the decision maker chooses, are shown in Table 1.

Table 1: Available investment alternatives

	S_1	S_2	S_3	S_4	S_5
A_1	20	40	60	70	80
A_2	50	70	90	40	20
A_3	30	40	50	80	60
A_4	40	40	50	60	70
A_5	80	60	50	40	30

In this problem, the experts assume the following weighting vector: W = (0.3, 0.2, 0.2, 0.2, 0.1). They assume that the probability that each state of nature will happen is: P = (0.1, 0.3, 0.3, 0.2, 0.1). Note that the OWA operator has an importance of 40% and the probabilistic information an importance of 60%. With this information, we can calculate how the attitudinal character of the decision maker may affect the probabilistic information. For doing so, we will use (3) to calculate the "attitudinal probabilities". The results are shown in Table 2.

Table 2: Attitudinal probabilities

	$\hat{v}_1$	$\hat{v}_2$	$\hat{v}_3$	$\hat{v}_4$	$\hat{v}_5$
A_1	0.18	0.2	0.26	0.26	0.1
A_2	0.3	0.26	0.14	0.2	0.1
A_3	0.24	0.14	0.26	0.26	0.1
A_4	0.18	0.2	0.26	0.26	0.1
A_5	0.18	0.26	0.26	0.2	0.1

With this information, we can aggregate the expected results for each state of nature in order to make a decision. In Table 3, we present different results obtained by using different types of IOWAAC operators.

Table 3: Aggregated results

	Prob.	OWA	PAM	POWA
A_1	54	60	54	56.4
A_2	63	61	59.4	62.2
A_3	52	57	52	54
A_4	50	55	50.8	52
A_5	52	57	52	54

Note that we can also obtain these results by using (4). Then, we will calculate separately the OWA and the probabilistic approach as shown in Table 4.

Table 4: First aggregation process

	Prob.	AM	OWA
A_1	54	54	60
A_2	63	54	61
A_3	52	52	57
A_4	50	52	55
A_5	52	52	57

After that, we will aggregate both models in the same process considering that the OWA model has a degree of importance of 40% and the probabilistic information 60% as shown in Table 5.

Table 5: Final aggregated results

	Prob.	OWA	PAM	POWA
A_1	54	60	54	56.4
A_2	63	61	59.4	62.2
A_3	52	57	52	54
A_4	50	55	50.8	52
A_5	52	57	52	54

Obviously, we get the same results with both methods. If we establish an ordering of the alternatives, a typical situation if we want to consider more than one alternative, then, we get the results shown in Table 6. Note that the first alternative in each ordering is the optimal choice.

Table 6: Ordering of the investments

	Ordering
Probabilistic	$A_2 \succ A_1 \succ A_3 = A_5 \succ A_4$
OWA	$A_2 \succ A_1 \succ A_3 = A_5 \succ A_4$
PAM	$A_2 \succ A_1 \succ A_3 = A_5 \succ A_4$
POWA	$A_2 \succ A_1 \succ A_3 = A_5 \succ A_4$

As we can see, depending on the aggregation operator used, the ordering of the investments may be different. Therefore, the decision about which investment select may be also different. Note that in this example, we get the same results for all the cases but in other problems or other types of aggregation operators we may find different results and decisions.

6 Conclusions

We have presented an approach for decision making with probabilities and OWA operators. We have developed a new aggregation operator that unifies the probabilities with the OWA operators. We have called it the POWA operator. The main advantage is that it gives a more complete representation of the decision problem because we are taking into account the probabilistic information and the attitudinal character of the decision maker. We have studied this new formulation and we have seen that it includes decision making problems under risk environment and under uncertainty as particular cases. We have studied some of its main properties and different families of POWA operators.

ISBN: 978-989-95079-6-8

We have also studied the applicability of the POWA operator and we have seen that there are a lot of potential applications that can be developed because in almost all the studies where it appears the probabilities, it is possible to extend the analysis by using the POWA operator. The reason is that the probabilistic information is always affected by the uncertainty. Therefore, it is always possible to consider the attitudinal character of the decision maker by using the OWA operator. We have focused on a decision making problem about selection of investments. We have seen the usefulness of using the POWA operator because we are able to consider probabilities and OWAs at the same time.

In future research, we expect to develop further extensions to this approach by adding new characteristics in the problem such as the use of order inducing variables, uncertain information (interval numbers, fuzzy numbers, linguistic variables, etc.), generalized and quasi-arithmetic means and distance measures. We will also extend this approach to situations where we use the WA instead of probabilities and further developments that have been initially developed in [7]. We will also consider different applications giving special attention to business decision making problems such as strategic and product management.

Acknowledgment

This paper is partly supported by the Spanish *Ministerio de Asuntos Exteriores y de Cooperación*, *Agencia Española de Cooperación Internacional para el Desarrollo* (AECID) (project A/016239/08).

References

[1] G. Beliakov, A. Pradera and T. Calvo, *Aggregation Functions: A Guide for Practitioners*. Berlin: Springer-Verlag, 2007.

[2] H. Bustince, F. Herrera and J. Montero, *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models*. Berlin: Springer-Verlag, 2008.

[3] T. Calvo, G. Mayor and R. Mesiar, *Aggregation Operators: New Trends and Applications*. New York: Physica-Verlag, 2002.

[4] K.J. Engemann, D.P. Filev and R.R. Yager, Modelling decision making using immediate probabilities. International Journal of General Systems, 24:281-294, 1996.

[5] J. Gil-Aluja, *The interactive management of human resources in uncertainty*. Dordrecht: Kluwer Academic Publishers, 1998.

[6] J.M. Merigó, Using immediate probabilities in fuzzy decision making. In: *Proceedings of the 22nd ASEPELT Conference*, Barcelona, pp. 1650-1664, 2008.

[7] J.M. Merigó, *New extensions to the OWA operators and its application in decision making* (In Spanish). PhD Thesis, Department of Business Administration, University of Barcelona, 2008.

[8] J.M. Merigó and A.M. Gil-Lafuente, Unification point in methods for the selection of financial products. *Fuzzy Economic Review*, 12:35-50, 2007.

[9] J.M. Merigó and A.M. Gil-Lafuente, The induced generalized OWA operator. *Information Sciences*, 179:729-741, 2009.

[10] V. Torra, The weighted OWA operator. *International Journal of Intelligent Systems*, 12:153-166, 1997.

[11] V. Torra and Y. Narukawa, *Modeling Decisions: Information Fusion and Aggregation Operators*. Berlin: Springer-Verlag, 2007.

[12] Z.S. Xu and Q.L. Da, An overview of operators for aggregating information. *International Journal of Intelligent Systems*, 18:953-968, 2003

[13] R.R. Yager, On ordered weighted averaging aggregation operators in multi-criteria decision making. *IEEE Transactions on Systems, Man and Cybernetics B*, 18:183-190, 1988.

[14] R.R. Yager, Decision making under Dempster-Shafer uncertainties. *International Journal of General Systems*, 20:233-245, 1992.

[15] R.R. Yager, Families of OWA operators. *Fuzzy Sets and Systems*, 59:125-148, 1993.

[16] R.R. Yager, Including decision attitude in probabilistic decision making. *International Journal of Approximate Reasoning*, 21:1-21, 1999.

[17] R.R. Yager and D.P. Filev, Induced ordered weighted averaging operators. *IEEE Transactions on Systems, Man and Cybernetics B*, 29:141-150, 1999.

[18] R.R. Yager and J. Kacprzyk, *The Ordered Weighted Averaging Operators: Theory and Applications*. Norwell: Kluwer Academic Publishers, 1997.

ISBN: 978-989-95079-6-8

Some results on Lipschitz quasi-arithmetic means

Gleb Beliakov[1], Tomasa Calvo[2], Simon James[3]

[1,3]School of Information Technology, Deakin University
221 Burwood Hwy, Burwood, 3125, Australia.
[2]Departamento de Ciencias de la Computación, Universidad de Alcalá
Crta. Barcelona Km 33,6. 28871-Alcalá de Henares (Madrid), Spain.
Email: {[1]gleb, [3]sgj}@deakin.edu.au, tomasa.calvo@uah.es

Abstract— *We present in this paper some properties of k-Lipschitz quasi-arithmetic means. The Lipschitz aggregation operations are stable with respect to input inaccuracies, what is a very important property for applications. Moreover, we provide sufficient conditions to determine when a quasi–arithemetic mean holds the k-Lipschitz property and allow us to calculate the Lipschitz constant k.*

Keywords— k-Lipschitz aggregation functions, quasi-arithmetic means, stability, triangular norms.

1 Introduction

Aggregation of several input values into a single output value is an indispensable tool in many disciplines and applications such as decision making [1], pattern recognition, expert and decision support systems, information retrieval, etc [2]. There is a wide range of aggregation functions which provide flexibility to the modeling process, including different types of aggregation functions. There are several recent books that provide details of many aggregation methods [3, 4, 5, 6, 7].

For applications it is important to design aggregation functions that are stable with respect to small perturbations of inputs (e.g., due to input inaccuracies). Such aggregation functions need not only be continuous, but Lipschitz continuous [8]. Kernel and 1-Lipschitz aggregation functions have been studied in [9, 10, 11]. It is known, for instance, that 1-Lipschitz triangular norms are copulas [12, 3, 7]. More recently, k-Lipschitz t-norms and t-conorms were studied [13, 14, 15]. k-Lipschitz t-norms do not increase the perturbation of inputs by more than a factor of k, which is suitable for many applications.

There are many other generated functions constructed similarly to the Archimedean triangular norms with the help of additive generators. In this article we examine quasi-arithmetic means and establish conditions under which these functions are Lipschitz or not Lipschitz.

Firstly, in Section 2 we recall some basic notions to develop the rest of the work. Section 3 contains the main results involving quasi–arithmetic means. At the end we provide some conclusions.

2 Preliminaries and related works

We restrict ourselves to aggregation functions defined on $[0,1]^n$.

Definition 1 *A function $f : [0,1]^n \to [0,1]$ is called an aggregation function if it is monotone non-decreasing in each variable and satisfies $f(0,\ldots,0) = 0$, $f(1,\ldots,1) = 1$.*

Now, we will pay attention to a special class of aggregation function, -the class of weighted quasi–arithmetic means-, for this we need to consider a continuous strictly monotone function $g : [0,1] \to [-\infty,\infty]$, which we call a *generating function* or generator. Of course, g is invertible, but it is not necessarily a bijection (i.e., its range may be $Ran(g) \subset [-\infty,\infty]$). Other two examples of generated functions are Archimedean t-norms and t-conorms. Further there exists a class of uninorms, known as *representable uninorms* or *generated uninorms*, that can also be built by means of additive generators. Further, a vector $\vec{w} = (w_1,\ldots,w_n)$ is called a weighting vector if $w_i \in [0,1]$ and $\sum_{i=1}^{n} w_i = 1$.

Definition 2 *For a given generating function g, and a weighting vector $\vec{w}$, the weighted quasi-arithmetic mean is the function*

$$M_{\vec{w},g}(\vec{x}) = g^{-1}\left(\sum_{i=1}^{n} w_i g(x_i)\right). \qquad (1)$$

From this definition, we have the following particular quasi–arithmetic means:

Arithmetic mean	$M(\vec{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i$
Geometric mean	$G(\vec{x}) = (\prod_{i=1}^{n} x_i)^{\frac{1}{n}}$
Harmonic mean	$H(\vec{x}) = n(\sum_{i=1}^{n} \frac{1}{x_i})^{-1}$
Power mean	$M_r(\vec{x}) = \frac{1}{n}(\sum_{i=1}^{n} x_i^r)^{\frac{1}{r}}$, if $r \neq 0$ and $M_0(\vec{x}) = G(\vec{x})$.

Another class of aggregation operators is the following

Definition 3 *Let $g : [0,1] \to [-\infty,\infty]$ be a continuous strictly monotone function and let $\vec{w}$ be a weighting vector. The function*

$$GenOWA_{\vec{w},g}(\vec{x}) = g^{-1}\left(\sum_{i=1}^{n} w_i g(x_{(i)})\right) \qquad (2)$$

is called a generalized OWA (also known as ordered weighted quasi-arithmetic mean [4]). As for OWA, $x_{(i)}$ denotes the i-th largest value of $\vec{x}$.

Another aggregation operator that include the previous one for the case of a symmetric fuzzy measure is the generalized discrete Choquet integral which is defined as follows

ISBN: 978-989-95079-6-8

Definition 4 *Let $g : [0,1] \to [-\infty,\infty]$ be a continuous strictly monotone function. The generalized Choquet integral with respect to a fuzzy measure v is the function*

$$C_{v,g}(\vec{x}) = g^{-1}\left(C_v(g(\vec{x}))\right),$$

where C_v is the discrete Choquet integral with respect to v and $g(\vec{x}) = (g(x_1),\ldots,g(x_n))$.

Now, we consider the crucial concept of this work

Definition 5 *An aggregation function f is called Lipschitz continuous if there is a positive number k, such that for any two vectors $\vec{x},\vec{y}$ in the domain of definition of f:*

$$|f(\vec{x}) - f(\vec{y})| \leq kd(\vec{x},\vec{y}), \tag{3}$$

where $d(\vec{x},\vec{y})$ is a distance between $\vec{x}$ and $\vec{y}$. The smallest such number k is called the Lipschitz constant of f (in the distance d). We shall call such functions k-Lipschitz.

Typically the distance is chosen as a p-norm $d(\vec{x},\vec{y}) = ||\vec{x}-\vec{y}||_p$, with $||\vec{x}||_p = \left(\sum_{i=1}^{n}|x_i|^p\right)^{1/p}$, for finite p, and $||\vec{x}||_\infty = \max_{i=1,\ldots,n}|x_i|$. In this work we concentrate on the 1–norm.

Definition 6 *A function f is called locally Lipschitz continuous on Ω if for every $x \in \Omega$ there exists a neighbourhood $D(x)$ such that f restricted to $D(x)$ is Lipschitz.*

Of course, duality w.r.t. standard negation preserves Lipschitz property (and the Lipschitz constant). It is easy to see that if an aggregation function A is k–Lipschitz, it is also m–Lipschitz for any $m \geq k$. Also any convex combination of $k-$Lipschitz aggregation functions $f = \alpha f_1 + \beta f_2$, $\alpha+\beta = 1, \alpha,\beta \geq 0$, is $k-$Lipschitz.

The class of k–Lipschitz t–norms, whenever $k > 1$, has been already characterized (see [13]). Note that 1–Lipschitz t–norms are copulas, see, e.g., [3, 5]. A strictly decreasing continuous function $g : [0,1] \to [0,1]$ with $g(1) = 0$ is an additive generator of a 1–Lipschitz Archimedean t–norm if and only if g is convex.

The k–Lipschitz property implies continuity of the t–norm. Recall that a continuous t–norm can be represented by means of an ordinal sum of continuous Archimedean t–norms, and that a continuous Archimedean t–norm can be represented by means of a continuous additive generator [3, 16]. Characterization of all k–Lipschitz t–norms can be reduced to the problem of characterization of all Archimedean k–Lipschitz t–norms.

Definition 7 *Let $g : [0,1] \to [0,+\infty]$ be a strictly monotone function and let $k \in]0,+\infty[$ be a real constant. Then g will be called k–convex if*

$$g(x+k\varepsilon) - g(x) \leq g(y+\varepsilon) - g(y)$$

holds for all $x \in [0,1[, y \in]0,1[$, with $x \leq y$ and $\varepsilon \in]0, \min(1-y, \frac{1-x}{k})]$.

Obviously, if $k = 1$ the function g is convex.

Observe that, a k–convex monotone function is also continuous in $]0,1[$, as was earlier. A decreasing function g can be k–convex only for $k \geq 1$. Moreover, when a decreasing function g is k–convex, it is also m–convex for all $m \geq k$. In the case of a strictly increasing function $g^\star$, it can be k-convex only for $k \leq 1$. Moreover, when $g^\star$ is k–convex, it is m–convex for all $m \leq k$.

Considering $k \geq 1$ and a strictly decreasing function g, we provide the following characterization given in [13].

Proposition 1 *Let $T : [0,1]^2 \to [0,1]$ be an Archimedean t–norm and let $g : [0,1] \to [0,+\infty], g(1) = 0$ be an additive generator of T. Then T is k–Lipschitz if and only if g is k–convex.*

Another useful characterizations are the following

Corollary 1 *(Y.-H. Shyu) [17] Let $g : [0,1] \to [0,\infty]$ be an additive generator of a t–norm T which is differentiable on $]0,1[$ and let $g'(x) < 0$ for $0 < x < 1$. Then T is k–Lipschitz if and only if $g'(y) \geq kg'(x)$ whenever $0 < x < y < 1$.*

Corollary 2 *Let $T : [0,1]^2 \to [0,1]$ be an Archimedean t–norm and let $g : [0,1] \to [0,\infty]$ be an additive generator of T such that g is differentiable on $]0,1[\backslash S$, where $S \subset [0,1]$ is a discrete set. Then T is k–Lipschitz if and only if $kg'(x) \leq g'(y)$ for all $x,y \in [0,1], x \leq y$ such that $g'(x)$ and $g'(y)$ exist.*

The following useful results follow from Corollary 1, with it we can determine whether a given piecewise differentiable t–norm is k–Lipschitz.

Corollary 3 *Let $T : [0,1]^2 \to [0,1]$ be an Archimedean t–norm and let $g : [0,1] \to [0,\infty]$ be its additive generator differentiable on $]0,1[$, and $g'(t) < 0$ on $]0,1[$. If*

$$\inf_{t\in]x,1[} g'(t) \geq k \sup_{t\in]0,x[} g'(t)$$

holds for every $x \in]0,1[$ then T is k–Lipschitz.

Corollary 4 *Let $g : [0,1] \to [0,\infty]$ be a strictly decreasing function, differentiable on $]0,a[\cup]a,1[$. If g is k-convex on $[0,a[$ and on $]a,1]$, and if*

$$\inf_{t\in]a,1[} g'(t) \geq k \sup_{t\in]0,a[} g'(t),$$

then g is k–convex on $[0,1]$.

Remark 1 *Generated uninorms are not k–Lipschitz since that they are not continuous at (0,1) and (1,0) (in the binary case). Nullnorms are aggregation functions related to t–norms and t–conorms. In this case, it is clear that a nullnorm V is k–Lipschitz if and only if the underlying t-norm and t–conorm are k–Lipschitz, and the k–Lipschitz constant of V is the maximum of Lipschitz constants of the underlying t–norm and t–conorm.*

ISBN: 978-989-95079-6-8

3 Quasi-arithmetic means

Consider a univariate continuous strictly monotone function $g : [0,1] \to [-\infty, \infty]$, called generator.

For a given g, the quasi-arithmetic means are defined in Definition 2, and are denoted by M_g. We start with bivariate quasi-arithmetic means of such operations. Quasi-arithmetic means are continuous if and only if $Ran(g) \neq [-\infty, \infty]$ [18]. Moreover, its generator is not defined uniquely, i.e., if $g(t)$ is a generating function of some weighted quasi-arithmetic mean, then $ag(t) + b$, $a, b \in \Re$, $a \neq 0$ is also a generating function of the same mean provided $Ran(g) \neq [-\infty, \infty]$. For this reason, one can assume that g is monotone increasing, as otherwise we can simply take $-g$.

We shall consider two cases: I) $g(0) = 0$, $g(1) = 1$, and II) $g(0) = -\infty, g(1) = 0$. Of course, by duality we also cover the case $g(1) = \infty, g(0) = 0$, and by using appropriate linear transformations, all generators can be reduced to the mentioned cases.

Let us make some preliminary remarks on convexity. As opposed to the case of convex additive generators of t-norms, where the resulting t-norms are 1-Lipschitz, convexity of the generator g does not play any role by itself for quasi-arithmetic means. Since both g and $-g$ are generators of the same mean, and obviously when g is convex $-g$ is concave, convexity of g by itself does not lead to the Lipschitz condition. Also note that $g(x) = -\ln(x)$ is a convex generator of the geometric mean $G(x, y) = \sqrt{xy}$, which is not Lipschitz. Further, even if g is convex and increasing or convex and decreasing, this does not imply Lipschitz condition either: note that $g_d(x) = 1 - g(1 - x)$ is a generator of a quasi-arithmetic mean dual to the one generated by g, and Lipschitz condition is preserved under duality. If g is convex increasing, then g_d is convex decreasing and vice versa. Thus we will look for a different condition.

3.1 Case of finite generators

We start with the case I) of g finite. First, let us show that g must be Lipschitz on [0,1]. For short, we will denote $M_{\vec{w},g}$ by M_g.

Lemma 1 *Let g be finite locally Lipschitz and continuously differentiable except at a point $a \in [0;1]$. Then M_g is not k–Lipschitz for any k.*

Proof. Suppose that M_g is k-Lipschitz, which means it is differentiable almost everywhere in its domain (Rademacher's theorem), and we must have

$$\frac{\partial M_g}{\partial x}(x, y) \leq k, \ x, y \neq a$$

whenever such a derivative exists, and similarly for the other partial derivative. Since M_g is symmetric, only the derivative with respect to x is needed.

$$\frac{\partial M_g}{\partial x} = \frac{1}{g'\left(M_g(x, y)\right)} \cdot \frac{1}{2} g'(x) \leq k.$$

Since g is strictly increasing we must have

$\frac{1}{2} g'(x) \leq k g'(M_g(x, y))$ for all $x, y \in [0, 1]$ such that $x \neq a$ and $M_g(x, y) \neq a$ or

$$\frac{1}{2} g'(x) \leq k \cdot \inf_{y \in [0,1]} g'(M_g(x, y)). \tag{4}$$

Since g is finite, M_g does not have an absorbing element. Let $\lim_{x \to a} g'(x) = \infty$.

$\exists y \neq a : z = M_g(a, y) \neq a$ such that $g'(z) \leq M < \infty$ (because g is locally Lipschitz). Then inequality (4) fails, because we can always choose such $x \neq a$ that $g'(x) > 2kM$, which would give us

$$kM < \frac{1}{2} g'(x) \leq k g'(z) \leq kM,$$

which is false. Then $\frac{\partial M_g}{\partial x} > k$, hence M_g is not Lipschitz. □

Now, since g is Lipschitz on $[0, 1]$, it is differentiable almost everywhere, which means that the left- and right-derivatives exist in $[0, 1]$. We start with the case of g differentiable on $[0, 1]$, and then adapt it to g differentiable almost everywhere by using left- and right-derivatives. Let M_g be k-Lipschitz. Then we must have

$$\frac{\partial M_g}{\partial x}(x, y) \leq k.$$

Following the same procedure as in Lemma 1, we get the condition

$$\frac{1}{2} g'(x) \leq k \cdot \inf_{y \in [0,1]} g'(M_g(x, y)) \tag{5}$$

for all $x \in [0, 1]$. Finally, by using left- and right- derivatives g'_-, g'_+ we obtain a general condition for non-smooth increasing generators

$$\begin{aligned} \frac{1}{2} g'_-(x) &\leq k \inf_{z \in [M(x,0), M(x,1)]} g'_-(z) \\ \frac{1}{2} g'_+(x) &\leq k \inf_{z \in [M(x,0), M(x,1)]} g'_+(z) \end{aligned} \tag{6}$$

for all $x \in]0, 1[$, and only one of the above inequalities for $x = 0$ and $x = 1$.

Remark 2 *If g is finite and concave increasing, then it is sufficient to check*

$$\frac{1}{2} g'(x) \leq k g'(M(x, 1)) = k(g' \circ g^{-1}) \left(\frac{g(x)}{2} + \frac{1}{2} \right),$$

(and similarly for left- and right-derivatives if g is not smooth). If g is finite and convex increasing, it is sufficient to check

$$\frac{1}{2} g'(x) \leq k g'(M(x, 0)) = k(g' \circ g^{-1}) \left(\frac{g(x)}{2} \right).$$

Let us provide some examples of Lipschitz and non-Lipschitz quasi-arithmetic means.

Example 1 *If g is linear (M_g is the arithmetic mean), $g'(x) = const$, and M_g is k-Lipschitz for $k = \frac{1}{2}$.*

Example 2 *If $g(x) = x^p, p > 1$ (M_g is a power mean $M_{[p]}$), $g'(x) = px^{p-1}$, and M_g is k-Lipschitz for $k = \left(\frac{1}{2}\right)^{\frac{1}{p}}$. It follows from*

$$\frac{1}{2} p x^{p-1} \leq kp \left(\frac{x^p}{2} \right)^{\frac{p-1}{p}} = kp x^{p-1} \left(\frac{1}{2} \right)^{\frac{p-1}{p}}.$$

Example 3 *If $g(x) = x^p, 0 < p < 1$ (M_g is a power mean $M_{[p]}$), M_g is not Lipschitz by Lemma 1.*

ISBN: 978-989-95079-6-8

3.2 Case of generators infinite at 0

Now we turn to the case II), g increasing with $g(0) = -\infty$, which entails that 0 is the absorbing element of M_g. We have an analogue of Lemma 1. The proof is similar, except that it fails for $a = 0$, hence the modification.

Let g be finite locally Lipschitz and continuously differentiable except at a point a 2 [0; 1]. Then Mg is not kLipschitz for any k

Lemma 2 *Let g be locally Lipschitz and continuously differentiable except at a point $a \in]0,1]$. Then M_g is not k-Lipschitz for any k.*

Let g be finite locally Lipschitz and continuously differentiable except at a point $a \in [0;1]$.

For $x \in]0,1]$ we have condition (6), to which we add condition

$$\frac{1}{2} \lim_{x \to 0^+} \frac{g'_+(x)}{g'_+(M(x,y))} \leq k \tag{7}$$

for any fixed $y \in]0,1]$. This condition may or may not be satisfied depending on the rate at which $M(x,y) \to 0$ as $x \to 0$. The choice of $y > 0$ is irrelevant as $g(y)$ is finite and disappears under the limit.

Example 4 *If $g(x) = -x^p, p < -1$ (M_g is a power mean $M_{[p]}$), $g'(x) = -px^{p-1}$, and M_g is k-Lipschitz for $k = \left(\frac{1}{2}\right)^{\frac{1}{p}}$. Differentiating M_g*

$$\begin{aligned}
\tfrac{\partial M_g}{\partial x} &= \tfrac{1}{p}\left(\tfrac{x^p+y^p}{2}\right)^{\frac{1}{p}-1} \tfrac{p}{2}x^{p-1} \\
&= \left(\tfrac{1}{2}\right)^{\frac{1}{p}} x^{-p(\frac{1}{p}-1)}(x^p+y^p)^{\frac{1}{p}-1} \\
&= \left(\tfrac{1}{2}\right)^{\frac{1}{p}} (1+x^{-p}y^p)^{\frac{1}{p}-1}.
\end{aligned}$$

Given $p < -1$,

$$k = \left(\tfrac{1}{2}\right)^{\frac{1}{p}} \lim_{x\to 0} (1+x^{-p}y^p)^{\frac{1}{p}-1} = \left(\tfrac{1}{2}\right)^{\frac{1}{p}}.$$

Example 5 *Let $M_{[p]}$, $-1 < p < 0$ be a power mean with a generator given by $g(x) = -x^p = -x^{-\frac{1}{q}}$, $q > 1$). The Lipschitz constant will be $k = \sup \frac{\partial M}{\partial x} = 2^q$. To see this*

$$\begin{aligned}
\tfrac{\partial M_g}{\partial x} &= \tfrac{-q}{2}\left(\tfrac{x^{-\frac{1}{q}}+y^{-\frac{1}{q}}}{2}\right)^{-q-1} \cdot \left(-\tfrac{1}{q}x^{-\frac{1}{q}-1}\right) \\
&= 2^q x^{\frac{1}{q}(-q-1)}(x^{-\frac{1}{q}}+y^{-\frac{1}{q}})^{-q-1} \\
&= 2^q\left(1+\left(\tfrac{x}{y}\right)^{\frac{1}{q}}\right)^{-q-1}. \\
k &= \sup\left\{2^q\left(1+\left(\tfrac{x}{y}\right)^{\frac{1}{q}}\right)^{-q-1}\right\} \\
&= 2^q = (\tfrac{1}{2})^{\frac{1}{p}}.
\end{aligned}$$

Condition 7 deals with the asymptotic behavior of the additive generators near 0. Its direct verification for a given g may be difficult. In the remainder of this section we will establish two sufficient conditions that guarantee that a quasi-arithmetic mean is not Lipschitz (although it is continuous). These conditions are easier to verify, and they provide a tool for a quick screening of additive generators with respect to their suitability for applications. One sufficient condition involves an inequality on the derivatives of the inverse of an additive generator. The other condition is that a decreasing additive generator cannot decrease slower than a certain rate (1/polynomial) when $x \to 0$. We will express this rate through the growth of an auxiliary function $1/g^{-1}$, for which the growth is expressed in traditional terms (e.g., polynomial) when $x \to \infty$. First, two simple auxiliary results.

Lemma 3 *If two functions f, g are continuous and differentiable at $x = 0$ and, $f(0) = g(0)$ and $f(x) \geq g(x)$ for $x > 0$, then $f'(0) \geq g'(0)$.*

Proof: Follows directly from the definition of the derivative.

The next result is a well-known condition for comparability of quasi-arithmetic means, see, e.g., [19].

Theorem 1 *Let g_1, g_2 be the generators of quasi-arithmetic means M_{g_1} and M_{g_2},and g_1 decreasing. Then $M_{g_1} \leq M_{g_2}$ if and only if $g_1 \circ g_2^{-1}$ is convex.*

Theorem 2 *Let g be an increasing (decreasing) twice continuously differentiable on $]0,1]$ generator of a quasi-arithmetic mean M_g where $g^{-1} = h$, and $\lim_{x\to 0} g(x) = -\infty$ ($\lim_{x\to 0} g(x) = +\infty$). If $h'^2 - hh'' \geq 0$ then M_g is not Lipschitz.*

Proof. We will show that $M_{[p]} \leq M_g$ for any $-1 < p < 0$ decreasing, and hence by Lemma 3 is not Lipschitz. If $x^p \circ g^{-1}$ is convex, for $-1 < p < 0$ by Theorem 1, with $g_1(x) = x^p$, $M_{[p]} \leq M_g$. Let us show that $(x^p \circ h)'' \geq 0$.

$$\begin{aligned}
(x^p \circ h)' &= ph^{p-1}h' \\
(x^p \circ h)'' &= p(p-1)h^{p-2}h'^2 + ph^{p-1}h'' \\
&= ph^{p-2}\left((p-1)h'^2 + hh''\right) \geq 0.
\end{aligned}$$

Given $ph^{p-2} < 0$ for $p < 0, h > 0$, convexity will hold if for all $p < 0$

$$(1-p)h'^2 - hh'' \geq 0. \tag{8}$$

Therefore $h'^2 - hh'' \geq 0$ implies $(x^p \circ h)'' \geq 0$ and $M_{[p]} \leq M_g$, and by Lemma 3 the Lipschitz constant of M_g is greater than that of $M_{[p]}$, which is $2^{-\frac{1}{p}}$, and $p \to 0^-$. □

Remark 3 *The generator g can be either increasing or decreasing. Clearly when changing g to $-g$, we change $h(x)$ to $h(-x)$. Then $h^{'}$ changes the sign but $h^{''}$ does not, hence the inequality in Theorem 2is the same for either increasing or decreasing generators.*

Example 6 *Using the geometric mean M_g, take $g(x) = \ln x$ with $h(x) = h^{'}(x) = h^{''}(x) = e^x$. Then $(h'^2 - hh'')(x) = e^{2x} - e^{2x} = 0$. Therefore M_g is not Lipschitz.*

For the sake of convenience, we will formulate our next result for decreasing additive generators satisfying $g(0) = \infty$. To obtain the respective condition on the increasing generators, we simply invert the sign of g.

Theorem 3 *Let $h = g^{-1}$ be the inverse of a decreasing generator g of a quasi-arithmetic mean M_g. If the function $\tilde{h} = \frac{1}{h}$ grows faster than any power $x^q, q > 0$, then M_g is not Lipschitz.*

ISBN: 978-989-95079-6-8

Proof. Fix y so that $g(y) = h^{-1}(y) = 0$, which is always possible (we remind that g is defined up to an arbitrary linear transformation).

$$\begin{aligned}\lim_{x\to 0}\frac{\partial M_g(x,y)}{\partial x} &= \lim_{x\to 0}\frac{dh\left(\frac{h^{-1}(x)}{2}\right)}{dx},\\ &= \lim_{x\to 0}\tfrac{1}{2}h'\left(\frac{h^{-1}(x)}{2}\right)(h^{-1})'(x)\\ &= \tfrac{1}{2}\lim_{x\to 0}h'\left(\frac{h^{-1}(x)}{2}\right)\tfrac{1}{h'(h^{-1}(x))}.\end{aligned}$$

Let $z = \frac{h^{-1}(x)}{2}$. Then

$$\lim_{x\to 0}\frac{\partial M_g(x,y)}{\partial x} = \frac{1}{2}\lim_{z\to\infty}\frac{h'(z)}{h'(2z)}.$$

Since h decreases faster than the power function $p(z) = Cz^r$, l'Hôpital's rule gives

$$0 = \lim_{z\to\infty}\frac{h(z)}{p(z)} = \lim_{z\to\infty}\frac{h'(z)}{p'(z)} = \lim_{z\to\infty}\frac{h'(2z)}{p'(2z)}.$$

For convenience of notation take p such that $p'(z) = \frac{1}{z^q}$. Then $p'(z) = p'(2z)2^q$.

$$\begin{aligned}\lim_{x\to 0}\frac{\partial M_g(x,y)}{\partial x} &= \tfrac{1}{2}\lim_{z\to\infty}\tfrac{h'(z)}{h'(2z)} = \tfrac{1}{2}\lim_{z\to\infty}\tfrac{h'(z)}{h'(2z)}\tfrac{p'(2z)2^q}{p'(z)}\\ &= 2^{q-1}.\end{aligned}$$

Since q can be arbitrarily large, the derivative is unbounded and M_g is not Lipschitz.

□

Example 7 *Let the generator be $g(x) = -\ln x$ as in Example 6. Clearly its inverse is $\exp(-x)$, and the auxiliary function $\frac{1}{h(x)} = \exp(x)$, which grows faster than any polynomial, hence the corresponding geometric mean is not Lipschitz.*

Further take any power of the logarithm $g(x) = (-\ln x)^r, r > 1$. The auxiliary function $\frac{1}{h(x)} = \exp(x^{\frac{1}{r}})$, it grows faster than a polynomial, hence the resulting mean is not Lipschitz either. Note that this quasi-arithmetic mean is related to the Aczél-Alsina family of t-norms [3, 5] by the equation

$$M_g = (T_r^{AA})^{\frac{1}{\sqrt{2}}},$$

which shows directly that M_g is not Lipschitz ($f(x) = M_g(x,1) = T_r^{AA}(x,1)^{\frac{1}{\sqrt{2}}} = x^{\frac{1}{\sqrt{2}}}$ is not Lipschitz).

Example 8
Consider the generator $g(x) = (-\ln x)^2, \tilde{h}(x) = e^{\sqrt{x}}$. From the previous example, $r = 2$ and we know the resulting mean is not Lipschitz, however this would not have been apparent from the application of Theorem 2, as

$$\begin{aligned}h'^2 - hh'' &= \tfrac{1}{4x}e^{-2\sqrt{x}} - \tfrac{1}{4x}e^{-2\sqrt{x}} - \tfrac{1}{4\sqrt{x^3}}e^{-2\sqrt{x}}\\ &= -\tfrac{1}{4\sqrt{x^3}}e^{-2\sqrt{x}} < 0.\end{aligned}$$

3.3 Weighted quasi-arithmetic means

We adapt conditions (6) and (7) for the case of unequal weights. For this we take partial derivatives with respect to all arguments. The Lipschitz constant is the largest, hence we have conditions

$$\begin{aligned}g'_-(x) &\le \frac{k}{\max w_i}\min_z g'_-(z) \qquad (9)\\ g'_+(x) &\le \frac{k}{\max w_i}\min_z g'_+(z)\end{aligned}$$

where
the minimum for z is over $[M(x,0,\dots,0), M(x,1,\dots,1)]$, and

$$\lim_{x\to 0^+}\frac{g'_+(x)}{g'_+(M(x,c,\dots,c))} \le \frac{k}{\max w_i} \qquad (10)$$

with $c \in]0,1]$.

Conditions (9) and (10) can also be used for symmetric means in the multivariate case, where $\max w_i = \frac{1}{n}$. It is clearly seen that the higher the number of variables, the smaller is the Lipschitz constant, if it exists.

Remark 4 *Similar, results can be obtained for generalized OWAs and generalized Choquet integrals.*

4 Conclusions

k-Lipschitz aggregation functions are important for applications because they can control the changes in the outputs due to input inaccuracies, to a fixed factor of k. k-Lipschitz triangular norms and conorms have been already characterized by k-convex additive generators, however no analogous results were available for quasi-arithmetic means. We have found verifiable conditions which guarantee that an aggregation function is $k-$Lipschitz for a given k, or alternatively, not Lipschitz. We also presented various examples of both Lipschitz and non-Lipschitz aggregation functions. Our results will benefit those who design aggregation functions for practical applications, as they allow one to make an informed choice on suitability of specific functions for these applications.

Acknowledgements

This work was supported by the Spanish project MTM2006-08322, PR2007-0193 and the European project 143423-2008-LLP-ES-KA3-KA3MP.

References

[1] H.-J. Zimmermann and P. Zysno. Latent connectives in human decision making. *Fuzzy Sets and Systems*, 4:37–51, 1980.

[2] D. Dubois and H. Prade. On the use of aggregation operations in information fusion processes. *Fuzzy Sets and Systems*, 142:143–161, 2004.

[3] E.P. Klement, R. Mesiar, and E. Pap. *Triangular Norms*. Kluwer, Dordrecht, 2000.

[4] T. Calvo, A. Kolesárová, M. Komorníková, and R. Mesiar. Aggregation operators: properties, classes and construction methods. In T. Calvo, G. Mayor, and R. Mesiar, editors, *Aggregation Operators. New Trends and Applications*, pages 3–104. Physica-Verlag, Heidelberg, New York, 2002.

[5] G. Beliakov, A. Pradera, and T. Calvo. *Aggregation Functions: A Guide for Practitioners*. Springer, Heidelberg, Berlin, New York, 2007.

[6] V. Torra and Y. Narukawa. *Modeling Decisions. Information Fusion and Aggregation Operators*. Springer, Berlin, Heidelberg, 2007.

[7] C. Alsina, M.J. Frank, and B. Schweizer. *Associative Functions: Triangular Norms And Copulas*. World Scientific, Singapore, 2006.

[8] T. Calvo and R. Mesiar. Stability of aggregation operators. In *2nd Conf. of the Europ. Soc. for Fuzzy Logic and Technology*, pages 475–478, Leicester, 2001.

[9] E.P. Klement and A. Kolesárová. Extension to copulas and quasi-copulas as special 1-Lipschitz aggregation operators. *Kybernetika*, 41:329–348, 2005.

[10] A. Kolesárová. 1-Lipschitz aggregation operators and quasi-copulas. *Kybernetika*, 39:615–629, 2003.

[11] A. Kolesárová, E. Muel, and J. Mordelová. Construction of kernel aggregation operators from marginal values. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10 Suppl.:37–49, 2002.

[12] B. Schweizer and A. Sklar. *Probabilistic Metric Spaces*. North-Holland, New York, 1983.

[13] A. Mesiarová. *Special Classes of Triangular Norms*. Phd thesis, Slovak Technical University, 2005.

[14] A. Mesiarová. Extremal k-Lipschitz t-conorms. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 14:247–257, 2006.

[15] A. Mesiarová. Approximation of k-Lipschitz t-norms by strict and nilpotent k-Lipschitz t-norms. *Int. J. General Syst*, 36:205 – 218, 2007.

[16] C.M. Ling. Representation of associative functions. *Publ. Math. Debrecen*, 12:189–212, 1965.

[17] Y.-H. Shyu. *Absolute continuity in the τ-operations*. Phd thesis, Illinois Institute of Technology, 1984.

[18] M. Komorníková. Aggregation operators and additive generators. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9:205–215, 2001.

[19] P.S. Bullen. *Handbook of Means and Their Inequalities*. Kluwer, Dordrecht, 2003.

ISBN: 978-989-95079-6-8

Modeling Position Specificity in Sequence Kernels by Fuzzy Equivalence Relations

Ulrich Bodenhofer, Karin Schwarzbauer, Mihaela Ionescu, and Sepp Hochreiter

Institute of Bioinformatics, Johannes Kepler University Linz, 4040 Linz, Austria
Email: {bodenhofer,schwarzbauer,ionescu,hochreit}@bioinf.jku.at

***Abstract*— *This paper demonstrates that several known sequence kernels can be expressed in a unified framework in which the position specificity is modeled by fuzzy equivalence relations. In addition to this interpretation, we address the practical issues of positive semi-definiteness, computational complexity, and the extraction of interpretable features from the final support vector machine classifier.*

***Keywords*—** fuzzy equivalence relation, kernel, sequence classification, support vector machines.

1 Introduction

The classification of biological sequences — in particular, nucleotide sequences (DNA and RNA) and amino acid sequences (proteins) — is one of the fundamental tasks in computational biology [1]. In the early days of computational biology, sequence statistics were employed. These approaches were later improved by Hidden Markov Models (HMM) and Artificial Neural Networks (ANN). In the last decade, *Support Vector Machines* (SVM) have become increasingly popular for sequence classification [2]. In many tasks — in computational biology, but also in many other domains — SVMs have outperformed all competing methods. It is justified to state that SVMs are nowadays widely considered the most powerful class of classification methods.[1] SVMs have been applied successfully in a wide spectrum of sequence classification tasks, ranging from promoter and splice site detection (both on DNA data) [3, 4, 5] to protein fold and secondary structure prediction (both on amino acid sequences) [6, 7, 8].

SVM classifiers, in their simplest form, are nothing else but linear classifiers. What distinguishes SVMs from other linear classifiers like perceptrons or logistic regressors is the fact that SVMs determine the classification function by maximizing the margin between the classes — a principle that is known to be optimal in terms of bounds on the generalization error [9, 10]. Since linear SVMs use input vectors only to compute scalar products, SVMs facilitate the so-called *kernel trick*, i.e. the replacement of the scalar products by a *kernel*, i.e. a nonlinear two-place function that is positive semi-definite. The use of kernels enables support vector machines to be applicable to almost any kind of data, including raw sequence data, signals, images, or graphs — only an appropriate kernel is needed. It is not surprising, therefore, that the design of appropriate sequence kernels is one of the central research topics in sequence classification with SVMs. Most of these sequence kernels are based on comparing sub-sequences and do not take the positions of these occurrences into account. In many applications, however, the occurrence of a specific sub-sequence is only indicative if it is at a specific position or region.

[1] although one has to admit that there are applications and circumstances/requirements under which other methods may be preferable

This paper repeats established sequence kernels that are based on occurrences of specific patterns along with some position-specific variants. We demonstrate that these kernels can be expressed in a unified framework in which the position specificity is modeled by fuzzy equivalence relations. This framework, on the one hand, allows for an interpretation of existing position-specific sequence kernels from the viewpoint of fuzzy equivalence relations. On the other hand, and more importantly from the practical viewpoint, this framework gives rise to new position-specific sequence kernels by allowing to use fuzzy equivalence relations that have not been considered for modeling position specificity previously.

2 Support vector classification

Suppose that we have to do binary classification of samples from a given arbitrary non-empty set $\mathcal{X}$. The two classes are denoted with $+1$ (positive class) and -1 (negative class). For a given training set

$$\{(x_i, y_i) \mid 1 \leq i \leq l\}$$

with $x_i \in \mathcal{X}$ and $y_i \in \{-1,+1\}$ for all $i = 1,\ldots,l$, the support vector machine classifier is represented as the following discriminant function:

$$f(x) = b + \sum_{i=1}^{l} \alpha_i \cdot y_i \cdot k(x, x_i) \tag{1}$$

In this formula, b is a real value, α_i are non-negative factors, and $k(.,.)$ is the so-called *kernel*, that is, a symmetric $\mathcal{X} \times \mathcal{X} \to \mathbb{R}$ mapping fulfilling *positive semi-definitenes*, i.e.

$$\sum_{i=1}^{n} \sum_{j=1}^{n} z_i \cdot z_j \cdot k(x_i, x_j) \geq 0 \tag{2}$$

for all $n \in \mathbb{N}$, all $(z_1,\ldots,z_n) \in \mathbb{R}^n$, and all $(x_1,\ldots x_n) \in \mathcal{X}^n$. With the notations $\mathbf{z} = (z_1,\ldots,z_n)^T$ and $\mathbf{K} = (k(x_i,x_j))_{i=1,\ldots,n}^{j=1,\ldots,n}$, Eq. (2) can be written as $\mathbf{z}^T\mathbf{K}\mathbf{z} \geq 0$. Hence, the positive semi-definiteness of k corresponds to the positive semi-definiteness of any kernel matrix $\mathbf{K}$.

The factors $\alpha_1,\ldots,\alpha_n$ are the Lagrange multipliers of a convex quadratic optimization problem arising from margin maximization. For details, the reader is referred to introductory tutorials [11, 12] and standard SVM literature [9, 10, 13, 14]. We only note shortly that samples contributing to the final classifier, obviously those x_i for which $\alpha_i > 0$, are called *support vectors*.

The positive semi-definiteness of the kernel k serves for two purposes. First, it ensures that the SVM optimization problem is a convex quadratic one, which ensures the existence

ISBN: 978-989-95079-6-8

of a global solution that can be determined efficiently. More importantly, by the famous *Mercer theorem* [15], it guarantees that there exists a Hilbert space $(\mathcal{H}, \langle .,. \rangle)$ and a mapping $\varphi : \mathcal{X} \to \mathcal{H}$ such that the representation

$$k(x,y) = \langle \varphi(x), \varphi(y) \rangle \tag{3}$$

holds for all $x, y \in \mathcal{X}$. This means that the kernel k can be understood as a scalar product in a feature space $\mathcal{H}$.

3 An overview of sequence kernels

In this paper, we restrict ourselves to sequence kernels that are based on the occurrence of specific patterns. Assume in the following that we have a certain finite alphabet $\mathcal{A}$ and that $\mathcal{X}$ is a set of finite strings over the alphabet $\mathcal{A}$. Further assume that $\mathcal{M}$ is a finite set of patterns. A pattern can either be a string over $\mathcal{A}$ itself (i.e. an *exact pattern*) or contain *wildcards*, i.e. positions that match to an arbitrary symbol or a subset of symbols from the alphabet $\mathcal{A}$. We consider sequence kernels that can be expressed as

$$k(x,y) = \sum_{m \in \mathcal{M}} N(m,x) \cdot N(m,y), \tag{4}$$

where $N(m,x)$ denotes the number of occurrences/matches of pattern m in string x. It is obvious that, according to the representation (3), such a kernel is nothing else but a scalar product in an $|\mathcal{M}|$-dimensional real feature space, where the feature mapping φ is explicitly given as

$$\varphi(x) = (N(m,x))_{m \in \mathcal{M}}. \tag{5}$$

In order to accommodate different matching criteria used in existing sequence kernels, let us express $N(m,x)$ as

$$N(m,x) = \sum_{p=1}^{\text{length}(x)} \mathbf{1}(m,x,p), \tag{6}$$

where $\mathbf{1}(m,x,p)$ is a kernel-specific indicator function whose value is 1 if pattern m matches string x at position p and 0 if not. So, in the framework presented here, a sequence kernels is basically given by its pattern set $\mathcal{M}$ and its match indicator function $\mathbf{1}$. This representation accommodates the following well-known sequence kernels:

Spectrum kernel [6]: $\mathcal{M} = \mathcal{A}^K$, i.e. the set of patterns is the set of all K-length strings; the indicator function is given as[2]

$$\mathbf{1}(m,x,p) = \begin{cases} 1 & \text{if } p+K-1 \leq \text{length}(x) \text{ and} \\ & m = x[p \ldots p+K-1], \\ 0 & \text{otherwise.} \end{cases}$$

In other words, the spectrum kernel maps each of the two sequences to the numbers of occurrences of all strings of length K and computes the scalar product of these two feature vectors.

[2]For a given string x, we define $x[r]$ to be the r-th character in x and $x[r \ldots q]$ to be the substring that starts with the r-th and ends with the q-th position.

Mismatch kernel [7]: this kernel is analogous to the spectrum kernel with the exception that matches are considered up to a maximal number of M mismatches, i.e. $\mathcal{M} = \mathcal{A}^K$ and

$$\mathbf{1}(m,x,p) = \begin{cases} 1 & \text{if } p+K-1 \leq \text{length}(x) \text{ and} \\ & |\{1 \leq r \leq K \mid m[r] \neq x[p+r-1]\}| \leq M, \\ 0 & \text{otherwise.} \end{cases}$$

Motif kernel [16]: $\mathcal{M}$ is a set of predefined patterns known or assumed to be related to the given classification task; the indicator mapping $\mathbf{1}$ is given as follows:[3]

$$\mathbf{1}(m,x,p) = \begin{cases} 1 & \text{if } p+\text{length}(m)-1 \leq \text{length}(x) \text{ and} \\ & m[r] = x[p+r-1] \text{ for all } 1 \leq r \leq \text{length}(m) \\ & \text{such that } m[r] \neq \text{“}*\text{”} \\ 0 & \text{otherwise} \end{cases}$$

Beside these three prominent kernels, also some others comply with the representation (4), e.g. spatial sample kernels [8].

The class of kernels (4) facilitates easy feature extraction, which is not surprising given the fact that the feature mapping φ is known explicitly. Suppose we are given a support vector machine with Lagrange multipliers $\alpha_1, \ldots, \alpha_l$ and an offset b. Then we can make the following rearrangements:

$$\begin{aligned} f(x) &= b + \sum_{i=1}^{l} \alpha_i \cdot y_i \cdot k(x,x_i) \\ &= b + \sum_{i=1}^{l} \alpha_i \cdot y_i \cdot \sum_{m \in \mathcal{M}} N(m,x) \cdot N(m,x_i) \\ &= b + \sum_{m \in \mathcal{M}} N(m,x) \cdot \underbrace{\sum_{i=1}^{l} \alpha_i \cdot y_i \cdot N(m,x_i)}_{=w(m)} \end{aligned} \tag{7}$$

So we obtain $w(m)$ as the linear scaling factor with which every occurance/match of m in the sequence x is scored. Obviously, the higher the absolute value $w(m)$, the more important the pattern m is for the final classification. If $w(m)$ is positive, the pattern m is indicative for the positive class and, if $w(m)$ is negative, m is indicative for the negative class.

Note that the representation (7) is not only beneficial for feature extraction, i.e. for determining which patterns are indicative for the classification tasks. If the total number of patterns actually occurring in the training set is not excessively large, the explicit representation (7) also has strong advantages in terms of computational complexity: the weights $w(m)$ can be stored in a simple hash table indexed by the patterns. For a new sample x, it is then only necessary to sum up the weights for all occurrences of all patterns in x. For the spectrum kernel, for instance, this means adding up $\text{length}(x) - K + 1$ numbers, whereas the direct implementation of the general SVM (1) requires the computation of all $k(x,x_i)$ for which $\alpha_i > 0$ and, therefore, much more computational effort.

[3]We restrict to fixed-length motifs with only regular characters and the wildcard “$*$” that matches all single characters, although the generalization to more sophisticated patterns, e.g. by regular expressions, is straightforward.

ISBN: 978-989-95079-6-8

4 Fuzzy position preference

The position-independent kernels presented in Section 3 have been used successfully in protein classification. Many protein classification tasks are concerned with inferring the structure and/or function of proteins from their amino acid sequences. Structure and function of proteins are largely determined by shorter sub-sequences (patterns/motifs), mostly irrespective of their exact position in the sequence, so position specificity is not an important issue.

When analyzing DNA or RNA nucleotide sequences, the situation is quite different. There is a rather small alphabet (four nucleic acids instead of twenty amino acids), so some patterns (in particular, shorter ones) are less indicative as they may occur by chance. The position where a pattern occurs, for example, relative to the start or end of a gene, plays a much bigger role. Therefore, there is a strong need for position-specific sequence classification.

We now want to illustrate how position specificity can be integrated into the class of kernels introduced in the previous section. So let us assume that all sequences we consider have the same length L, i.e. $\mathcal{X} = \mathcal{A}^L$, and are *aligned in a way meaningful for the given classification task*.

The simplest way to include position dependence is to consider each pattern at each position completely independent of the other occurrences at other positions (similar to the philosophy behind the *weighted degree kernel* [3]). This means that, instead of mapping a sequence to the total numbers of occurrences of all patterns (cf. (5)), each pattern induces L features, one for each position. The explicit feature mapping can be formulated as follows:

$$\varphi(x) = \Big(\underbrace{(\mathbf{1}(m,x,1),\ldots,\mathbf{1}(m,x,L))}_{\text{binary list of occurrences of pattern } m} \Big)_{m\in\mathcal{M}} \tag{8}$$

Then the corresponding position-specific kernel is given as follows:

$$\bar{k}(x,y) = \sum_{m\in\mathcal{M}} \sum_{p=1}^{L} \mathbf{1}(m,x,p)\cdot\mathbf{1}(m,y,p) \tag{9}$$

So this kernel basically counts the number of positions where the same pattern matches both sequences x and y. It is clear that the number of features (i.e. the dimension of the feature space $\mathcal{H}$) grows by a factor of L compared to the position-independent kernel (4) — potentially increasing the risk of overfitting. Moreover, this approach is not ideally suited for tasks in which patterns are indicative if they appear in certain regions (but not necessarily at exact positions).

The *oligo kernel* [5] solves this challenge by replacing the binary indicators in the feature mapping (8) by the sum of proximity functions around the occurrences of the pattern, with Gaussian bell functions being the standard choice. We leave this variant aside for a moment and concentrate on a different approach in line with the *shifted weighted degree (SWD) kernel* [4]. As in the kernel $\bar{k}$, a pattern m occurring at the same position in both sequences contributes 1 to the sum, but also an occurrence of a pattern m at position p in the sequence x and an occurrence of the same pattern m at position q in the sequence y may contribute a non-zero value to the sum. This is solved by a two-place weighting function $E : \{1,\ldots,L\}^2 \to [0,1]$ which corresponds to the closeness of positions p and q. The farer p and q are apart, the lower the value $E(p,q)$ should be. If $p = q$, i.e. if the positions coincide exactly, $E(p,q) = 1$ holds. So we may generalize the kernel $\bar{k}$ in the following way:

$$\tilde{k}(x,y) = \sum_{m\in\mathcal{M}} \sum_{p=1}^{L} \sum_{q=1}^{L} \mathbf{1}(m,x,p)\cdot E(p,q)\cdot\mathbf{1}(m,y,q) \tag{10}$$

If we leave some constant factors and the SWD kernel's option to define importance factors for each position aside, the position preference weights used by the SWD kernel essentially come down to

$$E_{\mathrm{SWD}}(p,q) = 2^{-|p-q|}.$$

It is easy to see that the kernel $\bar{k}$ given in (9) is a special case of $\tilde{k}$ with the position weighting function

$$E_{=}(p,q) = \begin{cases} 1 & \text{if } p = q, \\ 0 & \text{otherwise,} \end{cases}$$

i.e. the ordinary equality relation, and the position-independent variant defined in (4) is also a special case of (10) with the trivial position-independent weighting function $E_1(p,q) = 1$ (for all $p,q \in \{1,\ldots,L\}$).

Since the range of the weighting function E is the unit interval and since it is intended to be a model of the closeness of the two positions, it is immediate to ask the question whether fuzzy equivalence relations [17, 18] would be reasonable choices for E. As usual, a mapping $E : \mathcal{X}^2 \to [0,1]$ is called *fuzzy equivalence relation* with respect to a given triangular norm T if it has the following properties (for all $x,y,z \in \mathcal{X}$):

(i) Reflexivity: $E(x,x) = 1$

(ii) Symmetry: $E(x,y) = E(y,x)$

(iii) T-transitivity: $T(E(x,y),E(y,z)) \le E(x,z)$

It is trivial that the two special cases $E_=$ and E_1 highlighted above are fuzzy equivalence relations, no matter which triangular norm (t-norm) we consider.

It is also straightforward to prove that E_{SWD} is a fuzzy equivalence relation with respect to the product t-norm $T_{\mathbf{P}}(x,y) = x\cdot y$. This can be proved directly, but it is essentially a consequence of the well-known correspondence between pseudo-metrics and fuzzy equivalence relations with respect to continuous Archimedean t-norms [19]. As further consequences of this result, we can deduce some more examples (generally for $x,y \in \mathbb{R}$, containing the special case of natural numbers $1,\ldots,L$):

- $E_{\mathrm{lin},\sigma}(x,y) = \max(1 - \frac{1}{\sigma}|x-y|, 0)$ is a fuzzy equivalence relation with respect to the Łukasiewicz t-norm $T_{\mathbf{L}}(x,y) = \max(x+y-1,0)$.
- $E_{\exp,\sigma}(x,y) = \exp(-\frac{|x-y|}{\sigma})$ is a fuzzy equivalence relation with respect to the product t-norm $T_{\mathbf{P}}$.
- $E_{\mathrm{Gauss},\sigma}(x,y) = \exp(-\frac{(x-y)^2}{\sigma^2})$ is a fuzzy equivalence relation with respect to the t-norm

  $$T(x,y) = \exp\big(-(\sqrt{-\ln x} + \sqrt{-\ln y})^2\big),$$

  which is nothing else but the Aczél-Alsina t-norm with parameter $\lambda = \frac{1}{2}$ [20].

ISBN: 978-989-95079-6-8

The question arises which choices of E are feasible. In principle, every choice for which $E(p,q)$ is a non-increasing transformation of $|p-q|$ seems reasonable, and all examples mentioned above are of that kind.[4] However, there is also a technical requirement that should not be forgotten: the resulting function $\tilde{k}$ must be positive semi-definite. In order to deduce meaningful requirements on the fuzzy equivalence relation E, let us make a simple reformulation of $\tilde{k}$. It is trivial that the two inner sums are just a vector-matrix-vector product

$$\sum_{p=1}^{L}\sum_{q=1}^{L}\mathbf{1}(m,x,p)\cdot E(p,q)\cdot\mathbf{1}(m,y,q)=\varphi_m(x)^T\cdot\mathbf{E}\cdot\varphi_m(y),$$

with

$$\varphi_m(x)=\big(\mathbf{1}(m,x,1),\ldots,\mathbf{1}(m,x,L)\big)^T,$$
$$\varphi_m(y)=\big(\mathbf{1}(m,y,1),\ldots,\mathbf{1}(m,y,L)\big)^T,$$
$$\mathbf{E}=\begin{pmatrix}E(1,1)&\cdots&E(1,L)\\ \vdots&\ddots&\vdots\\ E(L,1)&\cdots&E(L,L)\end{pmatrix}.$$

Hence, we can rewrite $\tilde{k}$ as

$$\tilde{k}(x,y)=\sum_{m\in\mathcal{M}}\underbrace{\varphi_m(x)^T\cdot\mathbf{E}\cdot\varphi_m(y)}_{=\tilde{k}_m(x,y)}.$$

The sum of kernels is again a kernel [21]. So if we can guarantee that each $\tilde{k}_m$ is a kernel, $\tilde{k}$ is guaranteed to be a kernel. It is clear that, if the matrix $\mathbf{E}$ is positive semi-definite, $\tilde{k}_m$ is a scalar product, hence a kernel. Then the positive semi-definiteness of E, i.e. that E can be understood as a kernel itself, is a sufficient criterion for $\mathbf{E}$ to be positive semi-definite — and finally for $\tilde{k}$ to be a kernel.

For the first two examples above, the situation is clear: $E_=$ induces the identity matrices, E_1 induces matrices containing only 1's; both classes of matrices are trivially positive semi-definite. For non-trivial situations, fortunately, several results are known as well:

- Every fuzzy equivalence relation with respect to the minimum t-norm $T_{\mathbf{M}}(x,y)=\min(x,y)$ is positive semi-definite [22].
- $E_{\exp,\sigma}$ and $E_{\text{Gauss},\sigma}$ are long known to be positive semi-definite [23], where the latter is one of the most used kernels — the RBF kernel.
- $E_{\text{lin},\sigma}$ has recently been proved to be positive semi-definite [24].
- E_{SWD} is also positive semi-definite, as obviously every matrix $\mathbf{E}$ induced by E_{SWD} is diagonally dominant.

So we can conclude that the three choices $E_{\exp,\sigma}$, $E_{\text{Gauss},\sigma}$, $E_{\text{lin},\sigma}$ and E_{SWD} are reasonable and technically feasible. Further note that the former three choices have in common that the resulting kernels tend to the position-independent variant (4) as σ goes to infinity, and the resulting kernels tend to the most position-specific variant (9) as $\sigma\to 0$. So the parameter σ allows for adjusting the degree of position specificity in a continuous manner.

We further note that also the oligo kernel [5] can be accommodated in the framework introduced above. This kernel uses an explicit feature representation for each pattern that convolves the occurrences with Gaussian neighborhood (for some a priori chosen σ). So we can integrate the oligo kernel into the above framework with

$$\mathbf{E}_{\text{oligo},\sigma}=\mathbf{E}_{\text{Gauss},\sigma}\cdot\mathbf{E}_{\text{Gauss},\sigma}.$$

So, from the computational point of view, everything said here is also valid for the oligo kernel. However, since entries of the matrix $\mathbf{E}_{\text{oligo},\sigma}$ can exceed the unit interval, this distance weighting matrix cannot be interpreted as a fuzzy equivalence relation.

5 Feature extraction and computational issues

There is no doubt that the position-specific variant introduced in the previous section is more complicated than the position-independent variant presented in Section 3. The question arises whether the position-specific framework also allows for simple feature extraction and computational efficiency like the position-independent variant. Fortunately, this is the case. The basis of this result is the following rearrangement of the SVM discriminant function (analogous to Section 3):

$$\begin{aligned}f(x)&=b+\sum_{i=1}^{l}\alpha_i\cdot y_i\cdot\tilde{k}(x,x_i)\\&=b+\sum_{i=1}^{l}\alpha_i\cdot y_i\cdot\sum_{m\in\mathcal{M}}\sum_{p=1}^{L}\sum_{q=1}^{L}\mathbf{1}(m,x,p)\cdot E(p,q)\cdot\mathbf{1}(m,x_i,q)\\&=b+\sum_{m\in\mathcal{M}}\sum_{p=1}^{L}\mathbf{1}(m,x,p)\cdot\underbrace{\sum_{i=1}^{l}\sum_{q=1}^{L}\mathbf{1}(m,x_i,q)\cdot\alpha_i\cdot y_i\cdot E(p,q)}_{=\tilde{w}(m,p)}\end{aligned}$$

A value $\tilde{w}(m,p)$ can be interpreted as the weight to which an occurrence of pattern m at position p contributes to the final result. Analogously to Section 3, $\tilde{w}(m,p)>0$ means that an occurrence of pattern m at position p is indicative for the positive class, whereas $\tilde{w}(m,p)<0$ tells us that an occurrence of pattern m at position p is indicative for the negative class. The absolute value of $\tilde{w}(m,p)$ corresponds to the strength of this influence.

It is worth to point out that the way the weights $\tilde{w}(m,p)$ are computed has a very intuitive interpretation too. Suppose that pattern m occurs at position q in a support vector x_i. Obviously, if we consider p a free variable, $E(p,q)$ is the proximity (equivalence class) of position q. In case $E=E_{\text{lin},\sigma}$ is used, it is nothing else but a triangle with width σ; in the case $E=E_{\exp,\sigma}$, it is a Gaussian bell. So the term $\alpha_i\cdot y_i\cdot E(p,q)$ is nothing else but this proximity scaled by the Lagrange multiplier α_i and the sign y_i. Thus, the list $(\tilde{w}(m,p))_{p=1,\ldots,L}$ is the superimposition of proximities of all occurrences of pattern m in the training set scaled by corresponding Lagrange multipliers and signs/classes.

The list $(\tilde{w}(m,p))_{p=1,\ldots,L}$ can be plotted as a graph over the sequence showing the influence of pattern m at each position, indicating in which regions of the sequence the pattern is indicative for either class or not indicative at all. Thus,

[4] More generally, we could also apply a transformation to the positions first.

ISBN: 978-989-95079-6-8

the framework presented here allows for the same simple feature extraction as the position-independent variant. For other methods to extract features from SVM-based sequence classifiers, see [25, 26].

Moreover, it is obvious from the reformulation above that the discriminant function $f(x)$ can be evaluated in the same way as the position-independent variant — by adding up weights over all patterns occurring in the sequence x. So there is no difference in computational complexity between the position-specific and the position-independent variant. However, we have to store L times as many weights. This may be an obstacle if the number of patterns occurring in the training set and the sequence length L are both high.

For the position-specific spectrum kernel, for instance, the situation is very convenient. A sequence contains exactly $L-K+1$ sub-strings. Thus, the computation of $f(x)$ can be done by summing up $L-K+2$ values (again assuming that the lists of weights are stored in a convenient hash table). Obviously, the choice of K is of little influence on this complexity, which is is not true if we consider memory requirements. For low K's, the memory needed to store the weights grows exponentially (as $|\mathcal{M}| = \mathcal{A}^K$), but the total number of patterns that need to be considered is bounded above by the total number of patterns occurring in all support vectors.

All said above is valid for the spectrum kernel only. The mismatch kernel, for instance, is a much more difficult matter, as every sub-string of length K can match several patterns (where this number grows exponentially with M). Thus, the total number of patterns to be considered can be much higher and, more importantly, the number of patterns that need to be taken into account for computing $f(x)$ according to the above principle is much higher. Moreover, the extraction of features in the above manner is possible in principle, but impeded by the fact that the sets of sub-strings matching two different patterns may have large overlaps.

6 A DNA classification example

In this section, we demonstrate the concept introduced above by means of a case study. We consider the task of characterizing long nucleosome-free DNA segments.

Nucleosomes are a DNA packing mechanism in eukaryotic genomes. A nucleosome basically consists of a protein complex (a histone octamer) around which approximately 147 DNA base pairs (bp) are wrapped in $1\frac{2}{3}$ turns. Nonchalantly speaking, the histone complex acts as a reel around which DNA is wrapped. The DNA segment around the histone is mostly inaccessible for interactions with other molecules, most importantly, RNA polymerase. That is why the positions of nucleosomes and possible change of those positions play an essential role in transcription regulation. Therefore, it is of essential interest for the systems biology of eukaryotic cells which mechanisms determine the positioning and repositioning of nucleosomes. An interesting sub-topic is whether there are specific feature of the DNA that favor or hamper the positioning of nucleosomes.

For the yeast sub-species *Saccharomyces cerevisiae*, a commonly used model organism, data are available about where nucleosomes are located [27]. Instead of trying to find out which DNA features favor nucleosomes [28, 29], we tried to elicit DNA features that hamper the positioning of nucleosomes. To this end, we selected a total number of 10226 nucleosome-free sub-sequences of 100 bp length which occur before (on either strand) a well-positioned nucleosome (with the end of the sequence aligned to the start of the nucleosome coming thereafter). This was done using the data published along with [27]. These sequences were labeled as $+1$. The data set was complemented by the same number of random 100 bp long DNA fragments occurring in nucleosomes or in between two nucleosomes that are less than 100 bp apart. These sequences were labeled as -1.

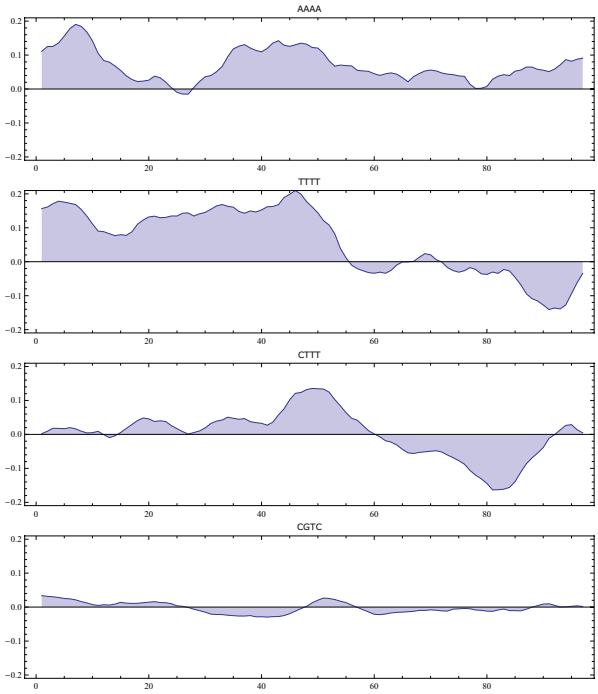


Figure 1: Weight profiles for the four patterns "AAAA", "TTTT", "CTTT" and "CGTC" (from top to bottom).

Our investigations were carried out as follows. The choices $\mathcal{A} = \{\text{"A"}, \text{"G"}, \text{"C"}, \text{"T"}\}$ and $L = 100$ were given by the data set under consideration. We used the spectrum kernel, both position-independent and position-specific, with different values for K ranging from 2 to 6. For the position-specific variant, we used $E = E_{\text{lin},\sigma}$ with σ's of 5, 10, and 20. We employed libSVM [30] to train standard soft margin SVMs. The best cross validation accuracy of 79.2% was obtained for the position-specific variant with $K = 4$ and $\sigma = 10$.

Let us have a closer look at the best setting $K = 4$. Obviously, we have $|\mathcal{M}| = 4^4 = 256$ different patterns. Hence, in order to realize the principle described in the previous section we have to store at most $256 \cdot 100 = 25,600$ numbers, which is easily manageable. Computing the discriminant function $f(x)$ for a new sequence x requires only to slide a window of length $K = 4$ over the sequence and to sum up the corresponding weights, hence only 98 additions are necessary, which is negligible compared to directly applying Eq. (1).

Figure 1 shows four examples of such weight profiles obtained for the given classification task. It is obvious that "AAAA" is indicative for the positive class (longer nucleosome-free positions), more or less regardless of the position. This is in line with the result of Peckham et al. [29]

ISBN: 978-989-95079-6-8

who have found out that poly-A occurrences are indicative for non-nucleosome sequence regions. One would expect the same situation for the complementing pattern "TTTT", but the profile in Figure 1 shows that this pattern is indicative for the positive class only in the first half of the sequence, whereas it is not indicative for the last 45 bases before a nucleosome. The pattern "CTTT" is not indicative for either class in the first half of the sequence, it seems to occur frequently around 50 bases before a nucleosome, and it occurs with less-than-random probability in the last 40 bases before a nucleosome (note that the negative class is randomly sampled from nucleosomes and shorter nucleosome-free fragments). The last graph shows that pattern "CGTC" seems to be of no significance at all. Although the biological interpretation of these results is a more advanced topic, the examples demonstrate how easily claims about the significance of specific patterns can be deduced from the weight profiles.

7 Concluding remarks

In this paper, we have formulated a generalization of position-specific sequence kernels by using fuzzy equivalence relations for modeling position specificity. This is not only an interesting link, but also gives rise to new kernels, e.g. the one based on $E_{\mathrm{lin},\sigma}$. We have obtained that these kernels facilitate an explicit representation which allows for computational efficiency and easy feature extraction.

References

[1] R. Durbin, S. R. Eddy, A. Krogh, and G. Mitchison. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press, 1998.

[2] B. Schölkopf, K. Tsuda, and J.-P. Vert, editors. *Kernel Methods in Computational Biology*. MIT Press, Cambridge, MA, 2004.

[3] G. Rätsch and S. Sonnenburg. Accurate splice site detection for *Caenorhabditis elegans*. In B. Schölkopf, K. Tsuda, and J.-P. Vert, editors, *Kernel Methods in Computational Biology*, pages 277–298. MIT Press, Cambridge, MA, 2004.

[4] G. Rätsch, S. Sonnenburg, and B. Schölkopf. RASE: recognition of alternatively spliced exons in *C.elegans*. *Bioinformatics*, 21(Suppl. 1):i369–i377, 2005.

[5] P. Meinicke, M. Tech, B. Morgenstern, and R. Merkl. Oligo kernels for datamining on biological sequences: a case study on prokaryotic translation initiation sites. *BMC Bioinformatics*, 5:169, 2004.

[6] C. Leslie, E. Eskin, and W. S. Noble. The spectrum kernel: a string kernel for SVM protein classification. In R. B. Altman, A. K. Dunker, L. Hunter, K. Lauderdale, and T. E. D. Klein, editors, *Pacific Symposium on Biocomputing 2002*, pages 566–575. World Scientific, 2002.

[7] C. Leslie, E. Eskin, A. Cohen, J. Weston, and W. S. Noble. Mismatch string kernels for discriminative protein classification. *Bioinformatics*, 1(1):1–10, 2003.

[8] P. Kuksa, P.-H. Huang, and V. Pavlovic. A fast, large-scale learning method for protein sequence classification. In *8th Int. Workshop on Data Mining in Bioinformatics*, pages 29–37, Las Vegas, NV, 2008.

[9] C. Cortes and V. N. Vapnik. Support vector networks. *Machine Learning*, 20:273–297, 1986.

[10] V. N. Vapnik. *Statistical Learning Theory*. Adaptive and Learning Systems. Wiley Interscience, 1998.

[11] C. J. M. Burges. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.*, 2:121–167, 1998.

[12] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf. An introduction to kernel-based learning algorithms. *IEEE Trans. Neural Networks*, 12(2):181–201, 2001.

[13] N. Christianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2000.

[14] B. Schölkopf and A. J. Smola. *Learning with Kernels*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, 2002.

[15] J. Mercer. Functions of positive and negative type and their connection with the theory of integral equations. *Philos. Trans. Roy. Soc. London*, A 209:415–446, 1909.

[16] A. Ben-Hur and D. Brutlag. Remote homology detection: a motif based approach. *Bioinformatics*, 19:i26–i33, 2003.

[17] L. A. Zadeh. Similarity relations and fuzzy orderings. *Inform. Sci.*, 3:177–200, 1971.

[18] L. Valverde. On the structure of F-indistinguishability operators. *Fuzzy Sets and Systems*, 17(3):313–328, 1985.

[19] B. De Baets and R. Mesiar. Pseudo-metrics and T-equivalences. *J. Fuzzy Math.*, 5(2):471–481, 1997.

[20] E. P. Klement, R. Mesiar, and E. Pap. *Triangular Norms*, volume 8 of *Trends in Logic*. Kluwer Academic Publishers, Dordrecht, 2000.

[21] C. H. FitzGerald, C. A. Micchelli, and A. Pinkus. Functions that preserve families of positive semidefinite matrices. *Linear Alg. Appl.*, 221:83–102, 1995.

[22] B. Moser. On the T-transitivity of kernels. *Fuzzy Sets and Systems*, 157(1):1787–1796, 2006.

[23] C. A. Micchelli. Interpolation of scattered data: Distance matrices and conditionally positive definite functions. *Constr. Approx.*, 2:11–22, 1986.

[24] L. Belanche, J. L. Vázquez, and M. Vázquez. Distance-based kernels for real-valued data. In C. Preisach, H. Burkhardt, L. Schmidt-Thieme, and R. Decker, editors, *Data Analysis, Machine Learning and Applications*, Studies in Classification, Data Analysis, and Knowledge Organization, pages 3–10. Springer, Berlin, 2008.

[25] T. Lingner and P. Meinicke. Remote homology detection based on oligomer distances. *Bioinformatics*, 22(18):2224–2231, 2006.

[26] S. Sonnenburg, A. Zien, P. Philips, and G. Rätsch. POIMs: positional oligomer importance matrices—understanding support vector machine-based signal detectors. *Bioinformatics*, 24:i6–i14, 2008.

[27] W. Lee, D. Tillo, N. Bray, R. H. Morse, R. W. Davis, T. R. Hughes, and C. Nislow. A high-resolution atlas of nucleosome occupancy in yeast. *Nature Genetics*, 39(10):1235–1244, September 2007.

[28] E. Segal, Y. Fondufe-Mittendorf, L. Chen, A. C. Thåström, Y. Field, I. K. Moore, J.-P. Z. Wang, and J. Widom. A genomic code for nucleosome positioning. *Nature*, 442(17):1235–1244, August 2006.

[29] H. E. Peckham, R. E. Thurman, Y. Fu, J. A. Stamatoyannopoulos, W. S. Noble, K. Struhl, and Z. Weng. Nucleosome positioning signals in genomic DNA. *Genome Res.*, 17:1170–1177, 2007.

[30] C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

ISBN: 978-989-95079-6-8

Universal Approximation of a Class of Interval Type-2 Fuzzy Neural Networks Illustrated with the Case of Non-linear Identification

Juan R. Castro[1], Oscar Castillo[2], Patricia Melin[2], Antonio Rodriguez-Diaz[1], Olivia Mendoza[1]

[1]Baja California Autonomous University, UABC, Tijuana, B. C., Mexico.
2 Division of Graduate Studies and Research, Tijuana Institute of Technology, Tijuana, B. C., Mexico.
Email: jrcastror@uabc.mx, ocastillo@tectijuana.mx

Abstract—*Neural Networks (NN), Type-1 Fuzzy Logic Systems (T1FLS) and Interval Type-2 Fuzzy Logic Systems (IT2FLS) are universal approximators, they can approximate any non-linear function. Recent research shows that embedding T1FLS on an NN or embedding IT2FLS on an NN can be very effective for a wide number of non-linear complex systems, especially when handling imperfect information. In this paper we show that an Interval Type-2 Fuzzy Neural Network (IT2FNN) is a universal approximator with some precision using a set of rules and Interval Type-2 membership functions (IT2MF) and the Stone-Weierstrass Theorem. Also, simulation results of non-linear function identification using the IT2FNN for one and three variables with 10-fold cross-validation are presented.*

Keywords— Interval Type-2 Fuzzy Logic Systems, Interval Type-2 Fuzzy Neural Networks, Neural Networks, Universal Approximation.

1 Introduction

It has been shown that a three layer NN can approximate any real continuous function [1]. The same has been shown for a T1FLS [2, 3] using Stone-Weierstrass Theorem [6]. A similar analysis was made by Kosko [4, 5] using the concept of fuzzy regions. In [6, 7] Buckley shows that, with a Sugeno model [9], a T1FLS can be built with the ability to approximate any non-linear continuous function. Also, combining neural and fuzzy logic paradigms [10, 11], an effective tool can be created for approximating any non-linear function [8]. In this sense, an expert can use a Type-1 Fuzzy Neural Network (T1FNN) [13, 14] or IT2FNN systems and find interpretable solutions [12, 17-20]. In general, Takagi-Sugeno-Kang (TSK) T1FLS's approximate well by the use of polynomial consequent rules [9, 23].

This paper uses the Levenberg-Marquardt backpropagation learning algorithm for adapting antecedent and consequent parameters for an adaptive IT2FNN, since its efficiency and soundness characteristics made them fit for these optimizing problems.

An Adaptive IT2FNN is used as a universal approximator of any non-linear functions. A set Ψ of IT2FNN is universal if and only if (iff), given any process Ω, there is a system $\Phi \in \Psi$ such that the difference between the output from IT2FNN and output from Ω is less than a given ε.

2 Interval Type-2 Fuzzy Neural Networks

An IT2FNN [12] combines a TKS interval type-2 fuzzy inference system (TSKIT2FIS) [15,16] with an adaptive NN in order to take advantage of both models best characteristics. Even though Mamdami and Tsukamoto are much complex models [13, 14, 16], TSK is preferred due to the lower computational cost in type reduction.

2.1 IT2FNN-2:A2C0 Architecture

An IT2FNN-2:A2C0 (IT2FNN-0) is a seven layer IT2FNN, which integrates a first order TSKIT2FIS (interval type-2 fuzzy antecedents and real consequents), with an adaptive NN. The IT2FNN-2:A2C0 (Figure 1) layers are described as follows:

Layer 0: Inputs

$$o_i^0 = x_i \quad ; i = 1,...,n$$

Layer 1: Adaptive type-1 fuzzy neuron (T1FN)

for i=1,…,n
  for k=s_i+1,…,s_{i+1}
    $net_k^1 = w_{k,i}^{1,0}\, x_i + b_k^1$
    $o_k^1 = \mu(net_k^1)$
  end
end

Layer 2: Non-adaptive T1FN

$o_{2k-1}^2 = o_{2k-1}^1 \cdot o_{2k}^1 \quad \forall\, k = 1,...,\vartheta$
$o_{2k}^2 = o_{2k-1}^1 + o_{2k}^1 - o_{2k-1}^2$

for k=1,…,r
  for i=1,…,n
    $\tau = 1_{k,i} \quad ; \quad \pi = \sum_{j=1}^{i-1} v_j + |\tau|$
    if $\tau > 0$
      $\underline{\mu}_{k,i} = o_{i_l(\pi)}^2 \,;\, \overline{\mu}_{k,i} = o_{i_u(\pi)}^2$
    else
      $\underline{\mu}_{k,i} = null \,;\, \overline{\mu}_{k,i} = null$
    end
  end
end

Layer 3: Lower-upper firing strength rule normalization. Nodes in this layer are non-adaptive and the output is defined as a the ratio between the k^{th} lower-upper firing strength rule and the sum of lower-upper firing strength of all rules:

ISBN: 978-989-95079-6-8

$$o_{2k-1}^3 = \underline{w}^k = \frac{\underline{f}^k}{\sum_{l=1}^{r} \underline{f}^l} \quad ; \quad k = 1,...,r$$

$$o_{2k}^3 = \overline{w}^k = \frac{\overline{f}^k}{\sum_{l=1}^{r} \overline{f}^l}$$

Layer 4: Rule consequents. Each node is adaptive and its parameters are $\{c_i^k, c_0^k\}$. The node's output corresponds to partial output of k[th] rule y^k.

$$y^k = \sum_{i=1}^{n} c_i^k x_i + c_0^k \quad ; \quad k = 1,...,r$$

$$o_{2k-1}^4 = y^k$$

$$o_{2k}^4 = y^k$$

Layer 5: Type reduction. Estimates left-right interval values $[\hat{y}_l, \hat{y}_r]$, nodes are non-adaptive with outputs $\hat{y}_l, \hat{y}_r$. Layer 5 output is defined by:

$$o_1^5 = \hat{y}_l = \sum_{k=1}^{r} \underline{w}^k \cdot y^k$$

$$o_2^5 = \hat{y}_r = \sum_{k=1}^{r} \overline{w}^k \cdot y^k$$

Layer 6: Defuzzification. This layer's node is adaptive, where the output $\hat{y}$ is defined as weighted average of left-right values and parameter γ. Parameter γ (default value 0.5) adjusts the uncertainty interval defined by left-right values $[\hat{y}_l, \hat{y}_r]$.

$$o_1^6 = \hat{y} = \gamma \hat{y}_l + (1-\gamma)\hat{y}_r$$

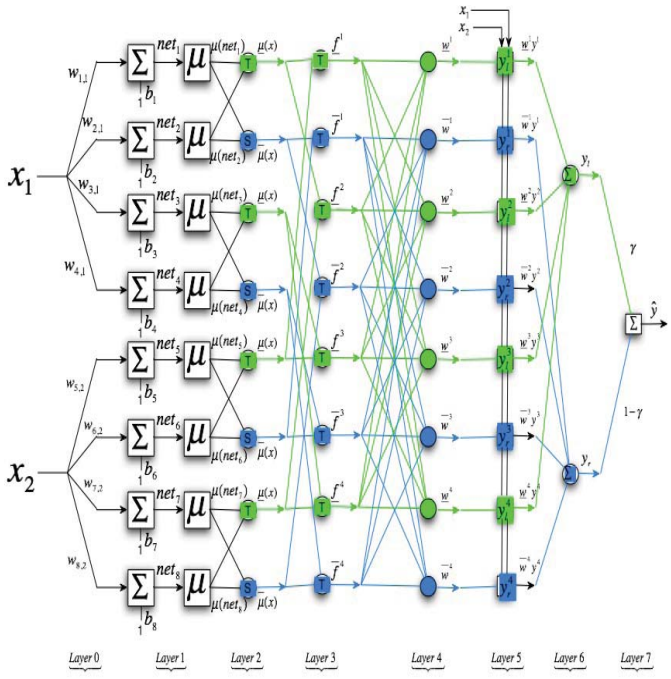


Figure 1: IT2FNN-2:A2C0 (IT2FNN-0) Architecture

2.2 IT2FNN-2:A2C1 Architecture

An IT2FNN-2:A2C1 (IT2FNN-2) [12] is a six layer IT2FNN, which integrates a first order TSKIT2FIS (interval type-2 fuzzy antecedents and interval type-1 fuzzy consequents), with an adaptive NN. The IT2FNN-2:A2C1 (Figure 2) layers are described as follows:

Layers 0-2 of IT2FNN-2:A2C1 architecture are the same as layers 0-2 of IT2FNN-2:A2C0 architecture.

Layer 3: Rule consequents. Each node is adaptive with parameters $\{c_i^k, c_0^k\}$ and $\{s_i^k, s_0^k\}$. Node's output corresponds to k[th] rule's partial output, $\tilde{y}^k \in [y_l^k, y_r^k]$.

$$o_{2k-1}^3 = y_l^k = \sum_{i=1}^{n} c_i^k x_i + c_0^k - \sum_{i=1}^{n} s_i^k |x_i| - s_0^k \quad ; \quad k = 1,...,r$$

$$o_{2k}^3 = y_r^k = \sum_{i=1}^{n} c_i^k x_i + c_0^k + \sum_{i=1}^{n} s_i^k |x_i| + s_0^k$$

Layer 4: Type reduction. Estimates left-right values of interval $[\hat{y}_l, \hat{y}_r]$ and left(f_l^k)-right(f_r^k) firing strength that are used for computing $[\hat{y}_l, \hat{y}_r]$ using KM algorithm [15, 16].

$$f_l^k = leftpoint(\underline{f}^k, \overline{f}^k, y_l^k) \quad ; \quad f_r^k = rightpoint(\underline{f}^k, \overline{f}^k, y_r^k)$$

$$o_1^4 = \hat{y}_l = \frac{\sum_{k=1}^{r} f_l^k y_l^k}{\sum_{k=1}^{r} f_l^k} \quad ; \quad o_2^4 = \hat{y}_r = \frac{\sum_{k=1}^{r} f_r^k y_r^k}{\sum_{k=1}^{r} f_r^k}$$

Layer 5: Defuzzification. This layer's node is adaptive with output $\hat{y}$ defined by:

$$o_1^5 = \hat{y} = \gamma \hat{y}_l + (1-\gamma)\hat{y}_r$$

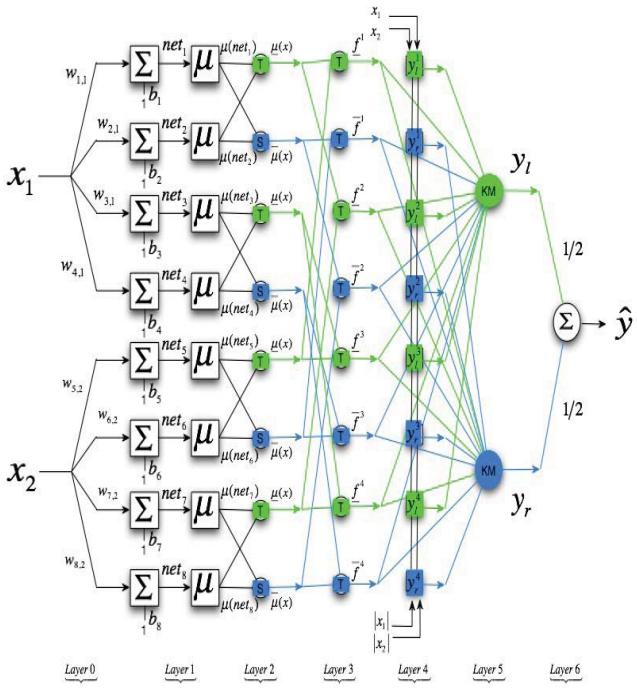


Figure 2: IT2FNN-2:A2C1 (IT2FNN-2) Architecture

ISBN: 978-989-95079-6-8

3 IT2FNN as a Universal Approximator

With this simplified interval type-2 fuzzy if-then rule, it is possible to prove that under certain circumstances, the resulting IT2FIS has unlimited approximation power to match any non-linear functions on a compact [24, 25] set using the Stone-Weierstrass theorem [13, 14, 21, 22] as follows:

3.1 Stone-Weierstrass Theorem.

Let the domain **D** be a compact space of **N** dimensions, and let $\mathfrak{I}$ be a set of continuous real-valued function on **D**, satisfying the following criteria [24, 25]:

1. **Identity function:** The constant $f(x) = 1$ is in $\mathfrak{I}$.
2. **Separability:** For any two points $x_1 \neq x_2$ in ***D***, there is an ***f*** in $\mathfrak{I}$ such that $f(x_1) \neq f(x_2)$.
3. **Algebraic closure:** If ***f*** and ***g*** are any two functions in $\mathfrak{I}$, then ***fg*** and ***af + bg*** are in $\mathfrak{I}$ for any two real numbers ***a*** and ***b***.

Then $\mathfrak{I}$ is dense in ***C***(***D***), the set of continuous real-value functions on **D**. In other words, for any $\varepsilon > 0$, and any function ***g*** in ***C***(***D***), there is a function ***f*** in $\mathfrak{I}$ such that $|g(x) - f(x)| < \varepsilon$ for all $\mathbf{x} \in \mathbf{D}$.

3.2 Applying Stone-Weierstrass Theorem used on IT2FNN-2:A2C0 Architecture.

In the IT2FNN, the domain in which we operate is almost always compact. It is a standard result in real analysis that every closed and bounded set in $\Re^n$ is compact. Now we shall apply the Stone-Weierstrass theorem to show the representational power of IT2FNN with simplified fuzzy if-then rules.

3.2.1 Identity Function

The first requirement of Stone-Weierstrass theorem needs our IT2FNN to be able to compute the identity function f(x)=1. An obvious way to compute the function is to set the consequence part of each rule equal to [1-δ, 1+δ], where δ is the uncertainty spread. In fact, an IT2FNN with only one rule suffices to satisfy this requirement.

3.2.2 Separability

The second requirement of the Stone-Weierstrass theorem needs the IT2FNN to be able to compute functions that have different values for different points. Without this requirement, the trivial set of functions f: f(x)= [c-δ, c+δ] c $\in \Re$ would satisfy Stone-Weierstrass theorem. Separability is satisfied whenever IT2FNN can compute strictly monotonic functions of each input variable. This can easily be achieved by adjusting the membership functions of the premise part.

3.2.3 Algebraic Closure-Additive

The third requirement of Stone-Weierstrass theorem needs the IT2FNN to be able to approximate sums and products of functions. Suppose we have two IT2FNNs S and $\overline{S}$ each of which has two rules. The output of each system can be expressed as

$$S: z = \gamma \cdot y_l + (1-\gamma) y_r = \gamma \frac{\underline{f}^1 y^1 + \underline{f}^2 y^2}{\underline{f}^1 + \underline{f}^2} + (1-\gamma) \frac{\overline{f}^1 y^1 + \overline{f}^2 y^2}{\overline{f}^1 + \overline{f}^2}$$

$$\overline{S}: \overline{z} = \overline{\gamma} \cdot \overline{y}_l + (1-\overline{\gamma}) \overline{y}_r = \overline{\gamma} \frac{\underline{g}^1 \overline{y}^1 + \underline{g}^2 \overline{y}^2}{\underline{g}^1 + \underline{g}^2} + (1-\overline{\gamma}) \frac{\overline{g}^1 \overline{y}^1 + \overline{g}^2 \overline{y}^2}{\overline{g}^1 + \overline{g}^2}$$

then the sum of z and $\overline{z}$ is equal to

$$az + b\overline{z} = \frac{\underline{f}^1 \overline{f}^1 \underline{g}^1 \overline{g}^1}{fg}\left[a y^1 + b\overline{y}^1\right] + \frac{\underline{f}^1 \overline{f}^1 \underline{g}^1 \overline{g}^2}{fg}\left[a y^1 + b\left(\overline{\gamma}\overline{y}^1 + (1-\overline{\gamma})\overline{y}^2\right)\right] +$$

$$\frac{\underline{f}^1 \overline{f}^1 \overline{g}^1 \underline{g}^2}{fg}\left[a y^1 + b\left(\overline{\gamma}\overline{y}^2 + (1-\overline{\gamma})\overline{y}^1\right)\right] + \frac{\underline{f}^1 \overline{f}^1 \underline{g}^2 \overline{g}^2}{fg}\left[a y^1 + b\overline{y}^2\right] +$$

$$\frac{\underline{f}^1 \overline{f}^2 \underline{g}^1 \overline{g}^1}{fg}\left[a\left(\gamma y^1 + (1-\gamma)y^2\right) + b\overline{y}^1\right] + \frac{\underline{f}^1 \overline{f}^2 \underline{g}^1 \overline{g}^2}{fg}\left[a\left(\gamma y^1 + (1-\gamma)y^2\right) + b\left(\overline{\gamma}\overline{y}^1 + (1-\overline{\gamma})\overline{y}^2\right)\right] +$$

$$\frac{\underline{f}^1 \overline{f}^2 \overline{g}^1 \underline{g}^2}{fg}\left[a\left(\gamma y^1 + (1-\gamma)y^2\right) + b\left(\overline{\gamma}\overline{y}^2 + (1-\overline{\gamma})\overline{y}^1\right)\right] + \frac{\underline{f}^1 \overline{f}^2 \underline{g}^2 \overline{g}^2}{fg}\left[a\left(\gamma y^1 + (1-\gamma)y^2\right) + b\overline{y}^2\right] +$$

$$\frac{\overline{f}^1 \underline{f}^2 \underline{g}^1 \overline{g}^1}{fg}\left[a\left(\gamma y^2 + (1-\gamma)y^1\right) + b\overline{y}^1\right] + \frac{\overline{f}^1 \underline{f}^2 \underline{g}^1 \overline{g}^2}{fg}\left[a\left(\gamma y^2 + (1-\gamma)y^1\right) + b\left(\overline{\gamma}\overline{y}^1 + (1-\overline{\gamma})\overline{y}^2\right)\right] +$$

$$\frac{\overline{f}^1 \underline{f}^2 \overline{g}^1 \underline{g}^2}{fg}\left[a\left(\gamma y^2 + (1-\gamma)y^1\right) + b\left(\overline{\gamma}\overline{y}^2 + (1-\overline{\gamma})\overline{y}^1\right)\right] + \frac{\overline{f}^1 \underline{f}^2 \underline{g}^2 \overline{g}^2}{fg}\left[a\left(\gamma y^2 + (1-\gamma)y^1\right) + b\overline{y}^2\right] +$$

$$\frac{\underline{f}^2 \overline{f}^2 \underline{g}^1 \overline{g}^1}{fg}\left[a y^2 + b\overline{y}^1\right] + \frac{\underline{f}^2 \overline{f}^2 \underline{g}^1 \overline{g}^2}{fg}\left[a y^2 + b\left(\overline{\gamma}\overline{y}^1 + (1-\overline{\gamma})\overline{y}^2\right)\right] +$$

$$\frac{\underline{f}^2 \overline{f}^2 \overline{g}^1 \underline{g}^2}{fg}\left[a y^2 + b\left(\overline{\gamma}\overline{y}^2 + (1-\overline{\gamma})\overline{y}^1\right)\right] + \frac{\underline{f}^2 \overline{f}^2 \underline{g}^2 \overline{g}^2}{fg}\left[a y^2 + b\overline{y}^2\right]$$

therefore we can construct an IT2FNN that computes $az + b\overline{z}$.

3.2.4 Algebraic Closure-Multiplicative

Modeling the product of $z\overline{z}$ of two IT2FNNs is the last capability we must demonstrate before we can conclude that the Stone-Weierstrass theorem can be applied to the proposed reasoning mechanism. The product $z\overline{z}$ can be expressed as

$$z\overline{z} = \frac{\underline{f}^1 \underline{g}^1 \left(\overline{f}^1 + \overline{f}^2\right)\left(\overline{g}^1 + \overline{g}^2\right)}{fg} \gamma y^1 \overline{\gamma}\overline{y}^1 + \frac{\underline{f}^1 \underline{g}^2 \left(\overline{f}^1 + \overline{f}^2\right)\left(\overline{g}^1 + \overline{g}^2\right)}{fg} \gamma y^1 \overline{\gamma}\overline{y}^2 +$$

$$\frac{\underline{f}^2 \underline{g}^1 \left(\overline{f}^1 + \overline{f}^2\right)\left(\overline{g}^1 + \overline{g}^2\right)}{fg} \gamma y^2 \overline{\gamma}\overline{y}^1 + \frac{\underline{f}^2 \underline{g}^2 \left(\overline{f}^1 + \overline{f}^2\right)\left(\overline{g}^1 + \overline{g}^2\right)}{fg} \gamma y^2 \overline{\gamma}\overline{y}^2 +$$

$$\frac{\underline{f}^1 \overline{g}^1 \left(\overline{f}^1 + \overline{f}^2\right)\left(\underline{g}^1 + \underline{g}^2\right)}{fg} \gamma y^1 (1-\overline{\gamma})\overline{y}^1 + \frac{\underline{f}^1 \overline{g}^2 \left(\overline{f}^1 + \overline{f}^2\right)\left(\underline{g}^1 + \underline{g}^2\right)}{fg} \gamma y^1 (1-\overline{\gamma})\overline{y}^2 +$$

ISBN: 978-989-95079-6-8

$$\frac{\underline{f}^2\underline{g}^1\left(\overline{f}^1+\overline{f}^2\right)\left(\underline{g}^1+\underline{g}^2\right)}{fg}\gamma y^2(1-\bar{\gamma})\bar{y}^1+\frac{\underline{f}^2\underline{g}^2\left(\overline{f}^1+\overline{f}^2\right)\left(\underline{g}^1+\underline{g}^2\right)}{fg}\gamma y^2(1-\bar{\gamma})\bar{y}^2+$$

$$\frac{\overline{f}^1\underline{g}^1\left(\underline{f}^1+\underline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}{fg}(1-\gamma)y^1\bar{\gamma}\bar{y}^1+\frac{\overline{f}^1\underline{g}^2\left(\underline{f}^1+\underline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}{fg}(1-\gamma)y^1\bar{\gamma}\bar{y}^2+$$

$$\frac{\overline{f}^2\underline{g}^1\left(\underline{f}^1+\underline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}{fg}(1-\gamma)y^2\bar{\gamma}\bar{y}^1+\frac{\overline{f}^2\underline{g}^2\left(\underline{f}^1+\underline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}{fg}(1-\gamma)y^2\bar{\gamma}\bar{y}^2+$$

$$\frac{\overline{f}^1\overline{g}^1\left(\underline{f}^1+\underline{f}^2\right)\left(\underline{g}^1+\underline{g}^2\right)}{fg}(1-\gamma)y^1(1-\bar{\gamma})\bar{y}^1+\frac{\overline{f}^1\overline{g}^2\left(\underline{f}^1+\underline{f}^2\right)\left(\underline{g}^1+\underline{g}^2\right)}{fg}(1-\gamma)y^1(1-\bar{\gamma})\bar{y}^2+$$

$$\frac{\overline{f}^2\overline{g}^1\left(\underline{f}^1+\underline{f}^2\right)\left(\underline{g}^1+\underline{g}^2\right)}{fg}(1-\gamma)y^2(1-\bar{\gamma})\bar{y}^1+\frac{\overline{f}^2\overline{g}^2\left(\underline{f}^1+\underline{f}^2\right)\left(\underline{g}^1+\underline{g}^2\right)}{fg}(1-\gamma)y^2(1-\bar{\gamma})\bar{y}^2$$

therefore we can construct an IT2FNN that computes $z\bar{z}$.

3.3 Applying Stone-Weierstrass theorem to the IT2FNN-2:A2C1 architecture

As in the case of IT2FNN-2:A2C0, let there be two distinct IT2FNN-2:A2C1, S and $\overline{S}$, each with two rules; we can proceed in this case as follows:

3.3.1 Algebraic Closure-Additive

$$S: z = \frac{1}{2}\left[\frac{\overline{f}^1 y_l^1+\underline{f}^2 y_l^2}{\overline{f}^1+\underline{f}^2}+\frac{\underline{f}^1 y_r^1+\overline{f}^2 y_r^2}{\underline{f}^1+\overline{f}^2}\right]$$

$$\overline{S}: \bar{z} = \frac{1}{2}\left[\frac{\overline{g}^1 \bar{y}_l^1+\underline{g}^2 \bar{y}_l^2}{\overline{g}^1+\underline{g}^2}+\frac{\underline{g}^1 \bar{y}_r^1+\overline{g}^2 \bar{y}_r^2}{\underline{g}^1+\overline{g}^2}\right]$$

then the sum of z and $\bar{z}$ is equal to

$$az+b\bar{z}=\frac{1}{2}\frac{\overline{f}^1\overline{g}^1}{\left(\overline{f}^1+\underline{f}^2\right)\left(\overline{g}^1+\underline{g}^2\right)}\left(ay_l^1+b\bar{y}_l^1\right)+\frac{1}{2}\frac{\overline{f}^1\underline{g}^2}{\left(\overline{f}^1+\underline{f}^2\right)\left(\overline{g}^1+\underline{g}^2\right)}\left(ay_l^1+b\bar{y}_l^2\right)+$$

$$\frac{1}{2}\frac{\underline{f}^2\overline{g}^1}{\left(\overline{f}^1+\underline{f}^2\right)\left(\overline{g}^1+\underline{g}^2\right)}\left(ay_l^2+b\bar{y}_l^1\right)+\frac{1}{2}\frac{\underline{f}^2\underline{g}^2}{\left(\overline{f}^1+\underline{f}^2\right)\left(\overline{g}^1+\underline{g}^2\right)}\left(ay_l^2+b\bar{y}_l^2\right)+$$

$$\frac{1}{2}\frac{\underline{f}^1\underline{g}^1}{\left(\underline{f}^1+\overline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}\left(ay_r^1+b\bar{y}_r^1\right)+\frac{1}{2}\frac{\underline{f}^1\overline{g}^2}{\left(\underline{f}^1+\overline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}\left(ay_r^1+b\bar{y}_r^2\right)+$$

$$\frac{1}{2}\frac{\overline{f}^2\underline{g}^1}{\left(\underline{f}^1+\overline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}\left(ay_r^2+b\bar{y}_r^1\right)+\frac{1}{2}\frac{\overline{f}^2\overline{g}^2}{\left(\underline{f}^1+\overline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}\left(ay_r^2+b\bar{y}_r^2\right)$$

therefore we can construct an IT2FNN that computes $az+b\bar{z}$.

3.3.2 Algebraic Closure-Multiplicative

The product $z\bar{z}$ can be expressed as

$$z\bar{z}=\frac{1}{4}\frac{\overline{f}^1\overline{g}^1}{\left(\overline{f}^1+\underline{f}^2\right)\left(\overline{g}^1+\underline{g}^2\right)}y_l^1\bar{y}_l^1+\frac{1}{4}\frac{\overline{f}^1\underline{g}^2}{\left(\overline{f}^1+\underline{f}^2\right)\left(\overline{g}^1+\underline{g}^2\right)}y_l^1\bar{y}_l^2+$$

$$\frac{1}{4}\frac{\underline{f}^2\overline{g}^1}{\left(\overline{f}^1+\underline{f}^2\right)\left(\overline{g}^1+\underline{g}^2\right)}y_l^2\bar{y}_l^1+\frac{1}{4}\frac{\underline{f}^2\underline{g}^2}{\left(\overline{f}^1+\underline{f}^2\right)\left(\overline{g}^1+\underline{g}^2\right)}y_l^2\bar{y}_l^2+$$

$$\frac{1}{4}\frac{\overline{f}^1\underline{g}^1}{\left(\overline{f}^1+\underline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}y_l^1\bar{y}_r^1+\frac{1}{4}\frac{\overline{f}^1\overline{g}^2}{\left(\overline{f}^1+\underline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}y_l^1\bar{y}_r^2+$$

$$\frac{1}{4}\frac{\underline{f}^2\underline{g}^1}{\left(\overline{f}^1+\underline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}y_l^2\bar{y}_r^1+\frac{1}{4}\frac{\underline{f}^2\overline{g}^2}{\left(\overline{f}^1+\underline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}y_l^2\bar{y}_r^2+$$

$$\frac{1}{4}\frac{\underline{f}^1\overline{g}^1}{\left(\underline{f}^1+\overline{f}^2\right)\left(\underline{g}^1+\underline{g}^2\right)}y_r^1\bar{y}_l^1+\frac{1}{4}\frac{\underline{f}^1\underline{g}^2}{\left(\underline{f}^1+\overline{f}^2\right)\left(\underline{g}^1+\underline{g}^2\right)}y_r^1\bar{y}_l^2+$$

$$\frac{1}{4}\frac{\overline{f}^2\overline{g}^1}{\left(\underline{f}^1+\overline{f}^2\right)\left(\underline{g}^1+\underline{g}^2\right)}y_r^2\bar{y}_l^1+\frac{1}{4}\frac{\overline{f}^2\underline{g}^2}{\left(\underline{f}^1+\overline{f}^2\right)\left(\underline{g}^1+\underline{g}^2\right)}y_r^2\bar{y}_l^2+$$

$$\frac{1}{4}\frac{\underline{f}^1\underline{g}^1}{\left(\underline{f}^1+\overline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}y_r^1\bar{y}_r^1+\frac{1}{4}\frac{\underline{f}^1\overline{g}^2}{\left(\underline{f}^1+\overline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}y_r^1\bar{y}_r^2+$$

$$\frac{1}{4}\frac{\overline{f}^2\underline{g}^1}{\left(\underline{f}^1+\overline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}y_r^2\bar{y}_r^1+\frac{1}{4}\frac{\overline{f}^2\overline{g}^2}{\left(\underline{f}^1+\overline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}y_r^2\bar{y}_r^2$$

therefore we can construct an IT2FNN that computes $z\bar{z}$.

The IT2FNN architectures that compute $az+b\bar{z}$ and $z\bar{z}$ have the same S and $\overline{S}$ if and only if the class of membership functions is invariant under multiplication. This is loosely true if the class of interval type-2 membership functions is the set of all bell-shaped functions and Gaussian-shaped (igbellmtype2, igbellstype2 and igbelltype2 for bell-shaped and igaussmtype2, igaussstype2 and igausstype2 for Gaussian-shaped, all of which can be found in the Matlab's Interval Type-2 Fuzzy Logic Toolbox [26]).

Therefore by choosing an appropriate class of membership functions, we can conclude that the IT2FNN with simplified fuzzy if-then rules satisfy the four criteria of Stone-Weierstrass theorem. Consequently, for any given $\varepsilon > 0$ and any real-valued function g, there is an IT2FNN S such that $|g(x)-S(x)|<\varepsilon$ for all x in underlying compact set. Moreover since the simplified IT2FNN is a proper subset of two types of architecture, we can conclude that the IT2FNN architecture has unlimited approximation power to match any given data set.

4 Application Examples

Two application examples are used to illustrate the proofs of universality, as follows.

Experiment 1. Identification of a one variable non-linear function.

In this experiment we approximate a non-linear function $f:\Re\to\Re$:

ISBN: 978-989-95079-6-8

$$f(u)=0.6\sin(\pi u)+0.3\sin(3\pi u)+0.1\sin(5\pi u)+\eta$$

(where η is a uniform noise component) using a one input-one output IT2FNN, 50 training data sets with 10 fold cross-validation with uniform noise levels, 6 IT2MF type igaussmtype2, 6 rules and 50 epochs. Once ANFIS and IT2FNN models are identified a comparison was made, taking into account RMSE statistic values with 10 fold cross-validation. Table 1 and Figure 3 show resulting RMSE values for ANFIS and IT2FNN, it can be seen that IT2FNN architectures [12] perform better than ANFIS.

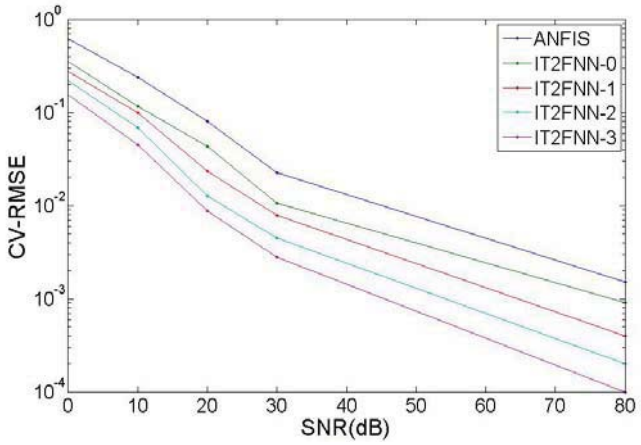


Figure 3: RMSE values of ANFIS and IT2FNN using 10-fold cross-validation for identifying non-linearity in Experiment 1.

Table 1: RMSE values of ANFIS and IT2FNN with 10-fold cross-validation for identifying non-linearity of Experiment 1.

SNR(dB)	ANFIS	IT2FNN-0	IT2FNN-1	IT2FNN-2	IT2FNN-3
0	0.6156	0.3532	0.2764	0.2197	0.1535
10	0.2375	0.1153	0.0986	0.0683	0.0453
20	0.0806	0.0435	0.0234	0.0127	0.0087
30	0.0225	0.0106	0.0079	0.0045	0.0028
free	0.0015	0.0009	0.0004	0.0002	0.0001

Experiment 2. Identification of a three variable non-linear function.

A three input – one output IT2FNN is used to approximate non-linear Sugeno [9] function $f:\Re^3\rightarrow\Re$:

$$f(x_1,x_2,x_3)=\left(1+\sqrt{x_1}+\frac{1}{x_2}+\frac{1}{\sqrt{x_3^3}}\right)^2+\eta$$

216 training data sets are generated with 10-fold cross-validation and 125 for tests; 2 igaussmtype2 IT2MF for each input, 8 rules and 50 epochs. Once the ANFIS and IT2FNN models are identified, a comparison is made with RMSE statistic values and 10-fold cross-validation. Table 2 and Figure 4 show the resultant RMSE values for ANFIS and IT2FNN. It can be seen that IT2FNN architectures [12] perform better than ANFIS.

Table 2: Resulting RMSE values in ANFIS and IT2FNN for non-linearity identification in Experiment 2 with 10-fold cross-validation.

SNR(dB)	ANFIS	IT2FNN-0	IT2FNN-1	IT2FNN-2	IT2FNN-3
0	1.0432	0.7203	0.6523	0.5512	0.5267
10	0.3096	0.2798	0.2583	0.2464	0.2344
20	0.1703	0.1637	0.1592	0.1465	0.1387
30	0.1526	0.1408	0.1368	0.1323	0.1312
free	0.1503	0.1390	0.1323	0.1304	0.1276

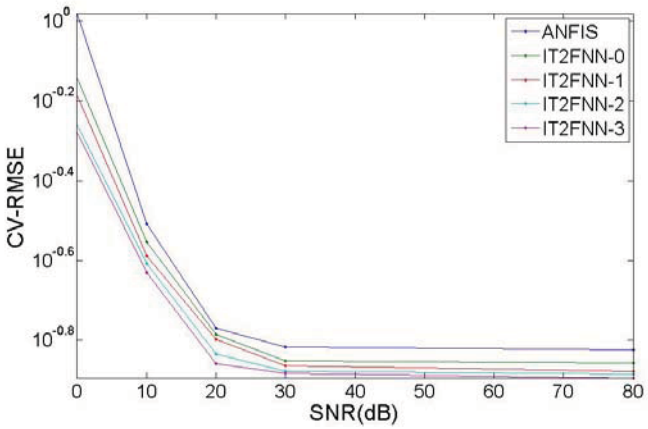


Figure 4: Resulting RMSE values obtained by ANFIS and IT2FNN for non-linearity identification in Experiment 2 with 10-fold cross-validation.

5 Conclusions

In these experiments, the estimated RMSE values for non-linear function identification with 10-fold cross-validation for the hybrid architectures IT2FNN-2:A2C0 and IT2FNN-2:A2C1 illustrate the proof based on Stone-Weierstrass theorem that they are universal approximators for efficient identification of non-linear functions. Also, it can be seen that while increasing the Signal Noise Ratio (SNR), IT2FNN architectures handle uncertainty more efficiently.

Acknowledgment

The authors would like to thank CONACYT and DGEST for the financial support given for this research project. The student (Juan R. Castro) is supported by a scholarship from UABC-CONACYT.

References

[1] K. Hornick, M., Stinchcombe and H. White, Multilayer feedforward networks are universal approximators. Neural Networks, 2(5):359-366, 1989.

ISBN: 978-989-95079-6-8

[2] L. Wang, and J. Mendel, Fuzzy basis functions, universal approximation, and orthogonal least squares learning. IEEE Trans. Neural Networks, 3:807-813, 1992.

[3] L. Wang and J. Mendel, Generating fuzzy rules by learning from examples. IEEE Trans. Syst. Man and Cyber., 22(6):1414-1427, 1992.

[4] B. Kosko, Neural Networks and Fuzzy Systems. Englewood Cliffs, NJ, Prentice Hall, 1992.

[5] B. Kosko, Fuzzy systems as universal approximators. IEEE Trans. Computers, 43(1):1329-1333, 1994.

[6] J. Buckley, Universal fuzzy controllers. Automatica, 28(6):1245-1248, 1992.

[7] J. Buckley, Sugeno type controllers are universal controllers. Automatica, 53:299-304, 1993.

[8] J. Buckley and Y. Hayashi, Can fuzzy neural networks approximate continuos fuzzy functions?, Fuzzy Sets and Systems, 61:43-52, 1994.

[9] T. Takagi and M. Sugeno, Fuzzy Identification of Systems and its applications to modelling and Control, *IEEE Transactions on Systems, Man, and Cybernetics*, 15(1): 116-132, Jan-Feb 1985.

[10] Lotfi A. Zadeh. Fuzzy sets. *Information and Control*, 8:338-353, 1965.

[11] K. Hirota and W. Pedrycz, OR/AND neuron in modeling fuzzy set connectives, IEEE Transactions on Fuzzy Systems, 2(994) 151-161.

[12] J.R. Castro, O. Castillo, P. Melin, A. Rodriguez-Diaz, A Hybrid Learning Algorithm for Interval Type-2 Fuzzy Neural Networks: The Case of Time Series Prediction; In Soft Computing for Hybrid Intelligent Systems. Book Chapter, Volume 154/2008: 363-386, Springer Berlin / Heidelberg. ISBN: 978-3-540-70811-7, ISSN: 1860-949X (Print) 1860-9503 (Online), September 2008.

[13] J.-S.R. Jang, "ANFIS: adaptive-network-based fuzzy inference system". IEEE Trans. Syst., Man, Cybern., vol. 23, no. 3, pp. 665-684, 1993.

[14] J.-S.R. Jang, C.T. Sun, and E. Mizutani, Neuro-fuzzy and Soft Computing. New York: Prentice-Hall, 1997.

[15] Q. Liang and J.M. Mendel, "Interval type-2 logic systems: theory and design," IEEE Transactions on Fuzzy Systems, vol. 8 no. 5, pp. 535-550, Oct. 2000.

[16] J.M. Mendel, Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions, Prentice Hall, NJ. 2001.

[17] Lee, C-H., Hu, T-W, Lee, C-T, Lee, Y-C, "A Recurrent Interval Type-2 Fuzzy Neural Network with Asymmetric Membership Functions for Nonlinear System Identification". In FUZZ-IEEE 2008, Hong Kong, June 2008.

[18] J.R. Castro, O. Castillo, P. Melin, A. Rodriguez-Diaz, Hybrid learning algorithm for interval type-2 fuzzy neural networks, In Proceedings of Granular Computing 2007, pages 157-162, Silicon Valley, CA, USA, ISBN: 0-7695-3032-X, November 2007.

[19] C.H. Lee, J.L. Hong, Y.C. Lin, and W.Y. Lai, "Type-2 Fuzzy Neural Network Systems and Learning," International Journal of Computational Cognition, Vol. 1, No. 4, pp. 79-90, 2003.

[20] C.H. Wang, C.S. Cheng, and Tsu-Tian Lee, "Dynamical optimal training for interval type-2 fuzzy neural network (T2FNN)," IEEE Trans. on Systems, Man, and Cybernetics Part B: Cybernetics, Vol. 34, No. 3, pp. 1462-1477, 2004.

[21] L.V. Kantorovich and G.P. Akilov. Functional analysis. Pergamon, Oxford, 2nd edition, 1982.

[22] H.L. Royden. Real analysis. Macmillan, New York, 2nd edition 1968.

[23] V. Kreinovich, G.C. Mouzouris and H.T. Nguyen, "Fuzzy Rule Based Modelling as a Universal Aproximation Tool," in Fuzzy Systems, Modeling and Control (H. T. Nguyen and M. Sugeno, Eds.), pp. 135-195, Kluwer Ac. Publ., Boston, 1998.

[24] C.T. Scarborough, A.H. Stone, Products of nearly compact spaces, Transactions of the American Mathematical Society 124: 131-147, 1966.

[25] R.E. Edwards, Functional Analysis: Theory and Applications, Dover Publications, 1995.

[26] J.R. Castro, O. Castillo, P. Melin, A. Rodriguez-Diaz, Building fuzzy inference systems with a new interval type-2 fuzzy logic toolbox, In Transactions on Computational Science. Book Chapter, Volume 4750:104-114, Springer-Verlag. ISBN: 978-3-540-79298-7, ISSN: 0302-9743 (Print) 1611-3349 (Online), June 2008.

ISBN: 978-989-95079-6-8

Spectral Learning with Type-2 Fuzzy Numbers for Question/Answering System

Asli Celikyilmaz[1] I. Burhan Turksen[2]

1. Computer Sciences Division, University in California, Berkeley, CA, USA
2. TOBB Economy and Technology University, Ankara, Turkey & University of Toronto, Canada

Abstract— *Graph-based semi-supervised learning has recently emerged as a promising approach to data-sparse learning problems in natural language processing. They rely on graphs that jointly represent each data point. The problem of how to best formulate the graph representation remains an open research topic. In this paper, we introduce a type-2 fuzzy arithmetic to characterize the edge weights of a formed graph as type-2 fuzzy numbers. The fuzzy numbers are identified by the changing parameters of the fuzzy kernel nearest neighbor algorithm, namely the degree of fuzziness and the hyper-parameter of the Gaussian kernel function, both of which have an effect on the uncertainty in forming the affinity matrix of the graph. We introduce a new graph-based semi-supervised learning with the type-2 arithmetic operations. We apply this technique in the framework of label propagation and evaluate on a question answering task. We demonstrate that the type-2 SSL can improve the prediction accuracy and can be considered to be the an alternative tool for text mining applications of computational linguistics.*

Keywords— Graph-based semi-supervised learning, kernel fuzzy k-nearest neighbor, type-2 fuzzy numbers.

1 Introduction and Motivation

In building reliable models for real systems, identification of exact values of variables of model equations are required. In real life practices, precise values of parameters may not be obtained due to imprecise, noisy, vague, or incomplete nature of information. Fuzzy logic provides explanatory tools for such tasks, mainly because of its capability to manage imprecise categories to represent imperfect information, by means of fuzzy sets, graduality, measures of resemblance or aggregation methods. Type-1 fuzzy sets may not be enough to explain the whole spectrum of possible results, mainly because the values used to characterize the membership functions of type-1 fuzzy numbers are usually overly precise. Usually the level of information is improperly set to define membership functions, thus it is rather necessary to use type-2 fuzzy numbers to represent uncertainties in model parameters.

In this work, we mainly focus on uncertainties in finding similarities in text mining. We consider one of the most commonly used learning methods, namely the semi-supervised learning (SSL) method [1]. It is often the case in the areas of machine learning for classification problems such as the problem of text classification on web pages, automatic translation or online question/ answering systems, etc. that one needs to deal with a very small portion of labeled data and vast amounts of unlabeled data. For such cases, graph-based SSL methods (spectral learning methods) have proved to outperform other learning methods. In graph- based methods the data is represented by the nodes of a graph (Fig. 1), the edges of which are labeled with the pairwise distance of the incident nodes. One problem with spectral learning methods is that the procedure is highly sensitive to the choice of the kernel, for example it is very sensitive to the choice of the spread (variance) of a Gaussian kernel, which naturally effects the similarity matrix defined for the given dataset.

As in the phrase of "words can mean different things to different people", an entailment relation between a candidate sentence and a question posed by the user may be evaluated differently by different people. For instance, a different degree of entailment may be assigned by different people for pairs of question "Who bought Overture?" and candidate sentences such as "Yahoo bought Overture", "Yahoo owns Overture", "Overture acquisition by Yahoo", using linguistic terms such $strict$, $loose$, or $direct$ entailment. Current methods can only use crisp values to define such relations, which cannot be explained to a full extent. Type-2 fuzzy logic is the best fit to define the entailment relations between each sentence. To our knowledge, characterization of edge weights of a graph as type-2 fuzzy numbers, as presented in this paper, is a new approach. The novel type-2 SSL defines such uncertain entailment relation between two sentences by characterizing soft linked graphs.

In this paper we concentrate on characterization of the uncertainties in similarity measure when discovering knowledge from unstructured text using graph-based SSL algorithm. A common way to construct the affinity matrix of a graph is by application of a nearest neighbor method. We use a fuzzy k-nearest neighbor (FKNN) to allow fuzzy decisions based on fuzzy labels. In addition we use its kernel extension [2] to enable solving possible non-linearly separable problems and get non-linear fuzzy boundaries instead of linear boundaries when necessary. In addition kernel methods have proven to prevent over-fitting in high dimensional feature spaces. For these reasons, we consider applying type-2 fuzzy arithmetic to situations where the similarity between two objects, i.e, two sentences, is imprecise.

Thus, the novel type-2 SSL method learns the edge link weights via kernel fuzzy k-nearest neighbor algorithm [2] (KFKNN). We use the arithmetic operations on type-2 fuzzy numbers defined in [3]. For ease of calculations, we graduate the interval valued degree of fuzziness and the kernel parameter and obtain bounded discrete valued weights (interval valued) with associated type-2 membership grades, enabling to represent each weight link with a type-2 fuzzy number. In a way each membership value is further stretched [1] based on fuzziness of the model to capture uncertainty interval of membership values (Fig. 2). Using the interval type-2 fuzzy

[1] Zadeh[4] defines the membership values as elastic constraints that has to be stretched to get their full meaning.

ISBN: 978-989-95079-6-8

weights, we construct fuzzy graph Laplacian. The novel type-2 SSL uses label propagation algorithm [1] to obtain interval valued output values. Defuzzification follows to measure the performance of the new classifier.

In this paper we mainly focus on the application of the novel spectral classification method on a specific area of natural language processing. Text-based question answering (QA) is the process of automatically finding the answers to arbitrary questions in plain English by searching collection of text files. Recently intensive research has been continuing to grow in this area fostered by evaluation-based conferences, one of which is the Text REtrieval Conference (TREC) [5]. Among different question types, most current research focuses on the $factoid$ type questions, e.g., "*How tall is Eiffel Tower*", where the answer is short string indicating a fact with a named entity. We construct a textual entailment model to find the degree of match between the question posed by the user and the candidate answers retrieved by the search engine and then rank the sentences likely to contain the correct answer at the top.

We will briefly review graph-based SSL methods, KFKNN, and type-2 fuzzy numbers. Then we will present the novel type-2 SSL method followed by the benchmark analysis on the TREC dataset. Finally conclusions are drawn.

2 Graph Based Semi-Supervised Learning

We begin with notation and a brief summary of SSL algorithms [1], [6]. Let $X = \{x_1, ..., x_n\}$ represent n data points in $\Re^d$ and $Y = \{y_1, ..., y_n\}$ be their output targets. The labeled part of X is represented with $X_l = \{x_1, ..., x_l\}$ with associated labels $Y_l = \{y_1, ..., y_l\}^T$. For ease of presentation, we concentrate on binary classification, where y_i can take on either of $\{-1, +1\}$. X has also unlabeled part, i.e., $X = X_u \cup X_l$. The aim is to predict labels for the remaining unlabeled points, $X_u = \{x_1, ..., x_u\}$.

In this section we review the most general graph-based semi-supervised learning method for binary classification problem. A graph is denoted with $g = (V, E)$, where $V = X_l \cup X_u$ is set of nodes. E represents the edges connecting two vertexes and is represented by the $n \times n$ symmetric affinity matrix W, where an edge $W(x_i, x_j) = e(i, j)$ represents the similarity between any vertexes x_i and x_j and 0 if there is no similarity between them. The most common similarity measure is the Gaussian kernel of width σ,

$$W_{ij} = K(x_i, x_j) = e^{-\|x_i - x_j\|^2/2\sigma^2} \tag{1}$$

Earlier research [1] suggests that the choice of σ can strongly influence the results (it is one of the imprecise parameters we use to characterize uncertainty interval of membership functions in our approach.) The diagonal degree matrix D is defined for g by D=$\sum_j W_{ij}$.

In graph-based SSL, a function over the graph is estimated such that it satisfies two conditions: 1) it should be close to the given labels (L), and 2) it should be smooth (S) on the whole graph. These two conditions are presented in a regularization form to minimize the following basic algorithm. Let **f**=$[f_1, ..., f_l, f_{l+1}, ..., f_n]$ denote the predicted labels of X where $f \in \Re^n$. According to the mathematical representation of graph based approach, the objective is to minimize

$$\arg min_f \; E_L(\mathbf{f}) + \lambda E_S(\mathbf{f}). \tag{2}$$

In (2) $E_L(\mathbf{f})$ is a loss function to penalize the deviation from the given labels,

$$E_L(\mathbf{f}) = \sum_{i \subset L} (f_i - y_i)^2 \tag{3}$$

and $E_S(\mathbf{f})$ is a regularizer to represent the label smoothness,

$$E_S(\mathbf{f}) = \frac{1}{2} \sum_{i,j \in L \cup U} W_{ij}(f_i - f_j)^2 = f^T \mathbf{L} f \tag{4}$$

In (4), $L = D - W$ is the graph Laplacian. To satisfy the local and global consistency [1], the normalized combinatorial Laplacian is used such that the $E_S(\mathbf{f})$ is replaced with normalized Laplacian, $\mathcal{L} = I - D^{-1/2} L D^{-1/2}$, as follows:

$$E_S(\mathbf{f}) = \sum_{i,j \in L \cup U} W_{ij}(\frac{f_i}{\sqrt{d_i}} - \frac{f_j}{\sqrt{d_j}})^2 = f^T \mathcal{L} f \tag{5}$$

The classification function **f** is learned through any of the $label\ propagation$(LP) algorithms, e.g., [1], [7] on the graph. From consistency approach [1] $f*$ can be found as :

$$f^* = (I - \lambda \left(D^{-1/2} W D^{-1/2} \right)^{-1} Y \tag{6}$$

Nearest neighbors are most commonly used method to identify the affinity matrix of the graph. In this work, we use the KFKNN method by capturing the uncertainty interval of similarities via perturbations on learning parameters.

3 Kernel Fuzzy k-Nearest Neighbor

The k-nearest neighbor approach (KNN) is commonly used in spectral learning to obtain the link weights of the graph Laplacian. It does not depend on the distribution of selected k objects, whereas fuzzy KNN [8] deals with the distribution of the selected k objects to determine the sum of similarities between labeled data points. They both use Euclidean distance function. To handle with more complex real systems, instead of Euclidean distance, we use kernel function to calculate distance between two objects. Through some non-linear mapping the input data is mapped onto a higher-dimensional feature space, i.e., $X \rightarrow \phi(X)$ to conduct the fuzzy KNN. With the kernel trick, instead of mapping the data point and calculating the distance in the feature space, we can use the kernel function and compute the dot product in some feature space, $K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$. The distances are calculated in this space as follows:

$$\|\phi(x) - \phi(x_j)\|^2 = 2 - 2K(x, x_j) \tag{7}$$

Gaussian kernel functions as in (1) is used. Let $C_i,\ i = 1, ...c$ represent each class c of distinct labels. The FKNN algorithm [8] assigns memberships as a function of vector distance from their nearest neighbors and those neighbor memberships in the possible classes. The membership values of x_i^u in each labeled class $c = 1, ..., k$ is calculated with

$$\mu_i(x^u) = \frac{\sum_{j=1}^k \mu_{ij} \left(1/\left(1 - K(x - x_j)^{1/m-1}\right) \right.}{\sum_{j=1}^k \left(1/\left(1 - K(x - x_j)^{1/m-1}\right) \right.} \tag{8}$$

ISBN: 978-989-95079-6-8

Algorithm 1 Kernel Fuzzy κ-Nearest Neighbor

procedure KFKNN(k, X^l, X^u, m, σ)
  for $j \leftarrow 1, n_u$, $x_j^u \in X^u$ **do**
    $w_{x_j^u, x_l^l} \leftarrow K(x_j, x_l^l)$ and σ using (1).
    Select k nearest neighbors among all labeled objects, $N_j = \left\{x_{1,j}^l, ...x_{k,j}^l\right\}$
    Calculate the membership values of x_j^u, $\mu_i(x_j^u)$, in each class $i = 1, ..., c$ using N_j data points via (8)
    If required, classify x_j^u into any of C_i using criteria in (9)
  end for
end procedure

The initial class memberships of labeled points are calculated using the function in [2]. Class separation criteria is based on;

$$\hat{y}^u = \arg max_{i=1...c} \ \mu_i(x) \tag{9}$$

The input parameters of KFKNN are the k-nearest neighbors, the labeled X^l data points, and unlabeled X^u data points, the degree of fuzziness parameter, m and kernel distance variance parameter σ in (1). The parameter m determines how heavily the distance is weighted when calculating each neighbor contribution to the membership value. It has been recently shown [9], [10], [11] that changing the m parameter in fuzzy clustering methods [12] such as fuzzy c-means and fuzzy c-regression have an effect on the outcome of the partition matrix, as well as the classification results. With some perturbations on the fuzziness parameter, we can identify uncertainty interval of membership values distributions. In this work, we identify the parameter m as an interval valued parameter, $m := [m^l, m^u]$, $l : lower$, $u : upper$. Next we graduate the interval manually into crisp values, $m^r = \left\{m^1, ..., m^{nr}\right\}$. We do the same discretization for the Gaussian kernel hyperparameter, $\sigma^s = \left\{\sigma^1, ..., \sigma^{ns}\right\}$. Each pair of parameter corresponds to one discrete affinity matrix obtained from KFKNN. We obtain the degree of membership values of relations, degree of similarity of an edge, via fuzzy $max - min$ rule, to be explained in section 5.

4 Type-2 Fuzzy Numbers and Operations

Before we present the novel type-2 SSL method, we give a brief review of type-2 fuzzy numbers in the next. Introduced by Zadeh [13] in the trilogy of papers, a type-2 fuzzy set $\tilde{A}$ is characterized in [14] with a type-2 membership function, $\mu_{\tilde{A}}(x, u)$, which maps elements $x \subset X$ to their membership values $u \subset J_x \subseteq [0, 1]$ as follows:

$$\tilde{A} = \left\{\left((x, u), \mu_{\tilde{A}}(x, u)\right) | \forall x \subset X, \forall u \subset J_x \subseteq [0, 1]\right\} \tag{10}$$

where J_x is a type-1 fuzzy set, i.e., primary membership value, and $0 \leq \mu_{\tilde{A}}(x, u) \leq 1$ is called secondary membership grade. Any type-2 fuzzy set can be represented as a collection of embedded type-2 fuzzy sets [14], as follows:

$$\tilde{A}_e = \sum_{i=1}^{N} \left[f_{x_i}(u_i)/u_i\right]/x_i; u_i \subset J_{x_i} \subseteq [0, 1] \tag{11}$$

where u_i is the primary and $f_{x_i}(u_i)$ is the secondary membership grade. The total number of type-2 sets embedded in $\tilde{A}$ is given by $n_e = \prod_{i=1}^{N} M_i$, where M_i is the cardinality of each primary membership grade on $x_i \in X$. When X has a numeric domain, $\tilde{A}$ can be characterized as fuzzy number. Thus for $\tilde{A}_e^j = \left\{\left(u_i^j, f_{x_i}(u_i^j)\right), j = 1..N\right\}$ and $u_i^j \in \{u_{ik}, k = 1, ..., M\}$, we can define a fuzzy number with

$$\tilde{A} = \sum_{j=1}^{n_e} A_e^j \tag{12}$$

Zadeh [13] introduced the union and intersection of type-2 fuzzy sets. Mendel and John [14] then defined the type-2 operations with embedded type-2 fuzzy sets. The union of discrete [1] type-2 fuzzy sets $\tilde{A}$ and $\tilde{B}$ is given as:

$$\tilde{A} \cup \tilde{B} = \sum_{j=1}^{n_A} \tilde{A}_e^j \sum_{i=1}^{n_B} \tilde{B}_e^i = \sum_{j=1}^{n_A} \sum_{i=1}^{n_B} \tilde{A}_e^j \cup \tilde{B}_e^i \tag{13}$$

The union of two type-2 embedded sets in (13) is defined as:

$$\tilde{A}_e^j \cup \tilde{B}_e^i = \bigcup_{k=1}^{N} \left[F_{x_k}\left(u_k^j, w_k^j\right) / \left(u_k^j \vee w_k^i\right)\right] / x_k \tag{14}$$

where $F_{x_k}\left(u_k^j, w_k^j\right)$ is the flag that is computed by the t-norm of the secondary membership grades of any kth embedded set, $\mu_{A_e^j}\left(x_k, f_{x_k}\left(u_k^j\right)\right)$ and $\mu_{B_e^i}\left(x_k, f_{x_k}\left(w_k^i\right)\right)$ [3]. When the secondary membership grades are unity, i.e., interval type-2 fuzzy sets, the union of type-2 embedded sets are defined as:

$$\tilde{A}_e^j \cup \tilde{B}_e^i = \bigcup_{k=1}^{N} \left[1 / \left(u_k^j \vee w_k^i\right)\right] / x_k \tag{15}$$

In [15] the arithmetic operations with type-2 fuzzy numbers such as addition of $\tilde{about}$ 2 and $\tilde{about}$ 5 is defined based on (14) in series of steps. Later in [3], computationally less expensive arithmetic operations, e.g., addition and multiplication with type-2 fuzzy numbers, is defined, which will be used in this paper to define the similarity between the sentences to construct the type-2 affinity matrix. with unity secondary membership grades.

Here we will only give a brief description of addition between two interval type-2 fuzzy numbers,

$$\tilde{about}\ 3 = \{0.3, 0.6, 0.7\}/2, \{1\}/3, \{0.3, 0.6, 0.7\}/4$$
$$\tilde{about}\ 5 = \{0.6, 0.7\}/4, \{1\}/5, \{0.6, 0.7\}/6$$

Since the secondary membership grades are unity, we obtain for each discrete primary membership value:

$$\mu_{about A_e^j + \tilde{about} B_e^i}(z) = sup\left(u_x^j \wedge w_y^i\right) \tag{16}$$

where $z = x \otimes y$ is performed for any discrete number defined in $\tilde{about}\ A$ and $\tilde{about}\ B$, on all embedded numbers $u_x^j = \mu_{about\ A_e^j}$, $j = 1..n_A$ and $w_x^i = \mu_{about\ B_e^i}$, $i = 1..n_B$. In (16), $\wedge$ is t-norm(minimum). Then the union of all the pairs are captured along domains of $z \in Z = X \otimes Y$. For potential duplicate values from the resultant of the $supremum$ operation in (16), the best practice is to select pair with the highest

[1] In this paper we deal with discrete fuzzy sets as naturally the fuzzy sets are discretized to do inference.

ISBN: 978-989-95079-6-8

membership value to form the resultant interval type-2 fuzzy number. From the addition operation on two type-2 fuzzy numbers, the resulting type-2 fuzzy number would yield:

$$\tilde{about}\ 8 = \{0.3, 0.6, 0.7\}/6, \{0.6, 0.7\}/7, \{1\}/8, \{0.6, 0.7\}/9, \{0.3, 0.6, 0.7\}/10$$

5 Novel Type-2 Semi-Supervised Learning

5.1 Motivation

In graph based SSL methods, one of the main problems is to construct the graph. Typically, k-nearest neighbor of each data point is identified and then each node is connected to their k nearest neighbors according to the edge weights using a Gaussian distance function as in (1). In cases where it is difficult to identify numerical characterization of similarities between two data points, one could identify imprecise similarities as type-2 fuzzy numbers (as explained in the previous section) and characterize these fuzzy numbers based on small perturbations on the learning parameters. Applications of Type-2 fuzzy modeling tools, e.g., [16], [17], [18] has shown to be effective methods for such cases.

In this work, we mainly focus on the imprecision of model parameters to define the edge weights. It is in this sense that we characterized the edge weights of a graph with type-2 fuzzy numbers, to be explained next.

5.2 The Learning Algorithm

We assume that the dataset $X = \{x_1, ..., x_n\}$ contain set of labeled and unlabeled data points which are represented as $X = \{X^u \cup X^l\}$. There are c number of classes based on the class labels of the labeled data and k represents the nearest neighbor count. For the labeled training data points the degree of memberships to any of classes is either 0 or 1 depending on the label of the class. After we obtain the partition matrix of the overall dataset X, we compute type-2 link membership values of each pair in the dataset.

We construct one KFKNN model and obtain one partition matrix $U^{sr} \subset \Re^{n \times k}$ for each pair of imprecise parameters, $\{\sigma^s, m^r\}$, $s = 1..ns$, $r = 1..nr$, where $\mu_{ij}^{sr} \in U^{sr}$ represents the degree of membership of data point x_j to any class i obtained using parameter set of $\{\sigma^s, m^r\}$. The next step is to merge all these partition matrices to construct type-2 fuzzy graph. Each vertex is connected to its k-nearest neighbor obtained from each KFKNN model. First we measure the similarities between each connected node, $w_{pq}^s \in W^s$ in each graph with Gaussian weighted distance measure as in (1). Note that the edge similarities are effected by the Gaussian variance parameter, σ^s. As a result, we characterize each edge weight, $\tilde{w}_{pq} = \{w_{pq}^1, ..., w_{pq}^{ns}\}$, $s = 1, .., ns$, with a discrete interval valued number based on σ^s.

The next step is to associate degree of similarity to each edge weight. Any vertex v_j, representing each data point x_j holds a degree of membership μ_{ij}^{sr} to each class label $i = 1...c$ (conditional degree of possibility given class labels). To construct the fuzzy graph, μ_i^{sr} we calculate the degree of relationship between each node using conditional possibilities via min-max rule. Let $\mu_{R(x_p,x_q)}$ be the degree of relationship (similarity) between paired data points, x_p and x_q, in the graph which is measured from the conditional possibility values of

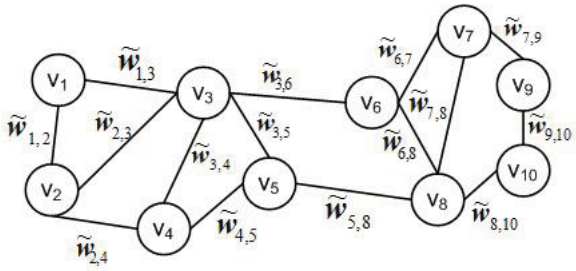


Figure 1: A type-2 fuzzy graph.

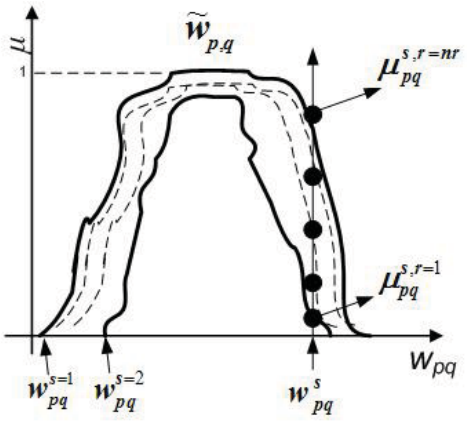


Figure 2: The interval type-2 fuzzy edge weight.

each vertex. Using the fuzzy $max-min$ rule of relations [13], we calculate the joint degree of fuzzy relations as follows:

$$\mu_{pq}(x_p, x_q) = \vee_{i=1,...,c} \left(\mu_i(x_p) \wedge \mu_i(x_q)\right) \qquad (17)$$

where $\vee$ and $\wedge$ denote max and min operations, respectively. Using (17) we define as many degree of similarity values, μ_{pq}^{sr}, as the number of discrete m^r, $r = 1..nr$, to characterize each discrete edge weight, w_{pq}^s. This would identify the edge weights as type-2 fuzzy numbers, see Figure 2.

In summary, we defined ns, $s = 1, ..., ns$ different variance values, σ^s to calculate ns different similarity values, w_{pq}^s, $s = 1.., ns.$ for any pair of connected nodes. These form the soft weight, $\tilde{w}_{pq} = w_{pq}^{s=1..ns}$ of the corresponding edge $e_{pq}, p, q = 1, ..., n$ of the graph. In addition, each discrete similarity value w_{pq}^s is represented with a degree of similarity from the $max-min$ operation, $\mu_{pq}^{sr}(x_p, x_q)$ using (17). These stretch out the degree of membership of each w_{pq}^s, and we identify discrete membership values for each weight, w_{pq}^s.

The fuzzy graph where the edges are represented with the soft type-2 weights is shown in Fig. 1, and Fig. 2 displays a sample edge weight that is in the form of type-2 fuzzy number.

The pseudo code for the type-2 fuzzy semi-supervised graph learning method is shown in Algorithm 2. Initially the partition matrix U is computed for unlabeled data points using the labeled data (Step-3). Note that the membership values of the labeled data points $\mu_i(x_j^l)$ to any of the c clusters are either 1 or 0 so they are merged to the partition matrix (Step-7).

After the conditional possibility values of each data point x_j to each class $\mu_i^{sr}(x_j)$ is obtained from every pair of parameter values, $\{\sigma^s, m^r\}$ we measure the joint possibilities of paired data points in step 9-10. We use the $max-min$ rule to obtain single degree of similarity between two nodes, x_p and x_q. In

ISBN: 978-989-95079-6-8

Algorithm 2 Forming Type-2 SSL Graph

procedure LEARNGRAPH(k, X^l, X^u, $X = \{X^l + X^u\}$)
  For $\sigma^s = \{\sigma^1, ..., \sigma^{ns}\}$ and $m^r = \{m^1, ..., m^r\}$
    -Execute Algorithm 1 using $\{\sigma^s, m^r\}$ to obtain
    each partition matrix $U^{sr} \subset \Re^{n \times c}$
    -Calculate similarity matrix $W^s = \{w^s_{pq}\}$ using σ^s
  EndFor
  $\mu^{sr}_j \in U^{sr} = U^{sr} + \{\mu_{i=1..c}(x^l_{i=1..n_l})\}$ where
  For $p, q = 1, ..., n,\ p \neq q$
    -Calculate degree of fuzzy relation, $\mu^{sr}_{pq}(x_p, x_q)$
    using (17)
    -Characterize each edge weight as type-2 fuzzy
    number $\tilde{w}_{pq} = \{w^{s=1..ns}_{pq}\} \in \tilde{W} \subset \Re^{n \times n}$
    using each μ^{sr}_{pq}.
    $\tilde{d}_p \in \tilde{D} \subset \Re^{n \times n} \leftarrow \sum_{q=1}^{n} \tilde{w}_{pq}$
  EndFor
  $\tilde{Ln} \leftarrow \tilde{D}^{-1/2} \tilde{W} \tilde{D}^{-1/2}$
end procedure

this way we represent the discrete membership values of the type-2 fuzzy number of the corresponding edge similarity.

Using the type-2 fuzzy number operations from section 4 [3], we obtain the fuzzy degree of each connected edge. Note that, instead of crisp edge weights, W, and graph node degrees, D, we characterized type-2 fuzzy edge matrix $\tilde{W}$ and degrees, $\tilde{D}$. To obtain the soft normalized graph Laplace, we compute the Step-16.

5.3 Inference with Soft Graphs

In the previous section we presented the algorithm to form a soft graph Laplacian matrix $\tilde{L}_n$, the members of which are represented with interval type-2 fuzzy numbers. It has been shown in recent literature [**?**, 9, 11, 10] that some of the uncertainties in system models can be captured during learning when we define interval values for the learning parameters instead of crisp values. For instance when fuzzy clustering methods are used to construct membership functions (fuzzy sets) defining an interval valued fuzziness parameter, m, instead of crisp value, we can identify the uncertainty interval of the membership functions as well as the perturbation effects on the local functions. The same is also true for the parameters of similarity functions. The variance parameter of a Gaussian kernel has a significant effect on the outcome [7]. By the analogy of these approaches, we use the interval valued Gaussian variance distance measures and fuzziness parameter of the KFKNN for identification of the soft graph Laplacian as explained in the previous section. The classifier function is learnt using the graph Laplacian. In standard SSL methods, briefly explained in section 2, one can use a label prorogation method to find the classifier function, f^*. In Type-2 SSL algorithm of this paper, we learn classifier by first type-reducing the type-2 graph Laplacian and then defuzzifying its values to obtain a single graph Laplace.

The graph Laplacian is a sparse matrix $L := \left(l^{n \times n}_{pq}\right)$ representing the difference between the degree matrix and weight matrix. The matrix elements are zero if there is no edge link, and 1 for the diagonal elements. Hence the matrix is sparse representing the graph where there are edge

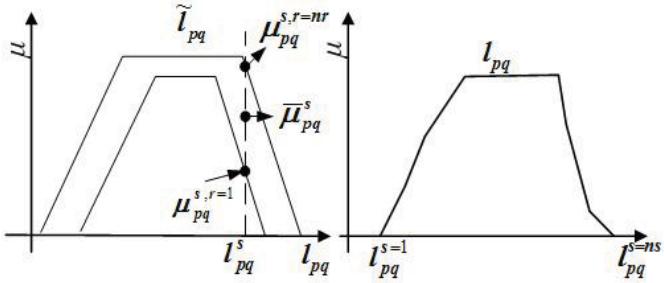


Figure 3: (Left Graph)Interval type-2 fuzzy laplace value. (Right Graph)Type-1 fuzzy laplace value.

links. We consider soft normalized graph Laplacian $\tilde{\mathcal{L}} = I - \tilde{D}^{-1/2} \tilde{W} \tilde{D}^{-1/2}$. Each matrix element $\tilde{l}_{pq}$ is an interval type-2 fuzzy number, represented as:

$$\tilde{l}_{pq} = \left\{\mu^{s=1,r=1}_{pq}, .., \mu^{s,r=nr}_{pq}\right\} / l^{s=1}_{pq} + \cdots + \left\{\mu^{s=ns,r=1}_{pq}, .., \mu^{s=ns,r=nr}_{pq}\right\} / l^{s=ns}_{pq}$$

We consider type-reducing the Laplace values via center of gravity (COG) rule as shown in Figure 3 as

$$COG(\mu^s_{pq}) = \sum_{j=1}^{nr} \mu^{s,r=j}_{pq} / nr \tag{18}$$

Then we defuzzify to obtain a single Laplace values:

$$l_{pq} = \frac{\sum_{t=1}^{ns} COG\left(\mu^s_{pq}\right) l^{s=t}_{pq}}{\sum_{i=1}^{ns} COG\left(\mu^s_{pq}\right)} \tag{19}$$

Using the defuzzified graph Laplacian, we obtain the classifier function using (6).

6 Textual Entailment Model for Question Answering

In this section we use a real question and answering (Q/A) dataset to test the performance of the type-2 SSL in comparison to the SSL method [7], and well-known SVM [19]. We used the TREC2001-2003 factoid questions for training and TREC04 questions for testing. The training and testing datasets comprise of 20 relevant sentences for each question that may or may not contain the true answer. We build the training and testing datasets using the relevant documents provided by NIST [2] and extracted 20 relevant sentences from TREC corpus using search engine [3]. We manually classified each retrieved sentence as true/false based on the provided gold-factoid-answers. In addition, we used additional question-sentence pairs from RTE-3, and RTE-4 datasets of RTE challenges [4]. We used only Q/A type sentence pairs and converted the hypothesis sentences into question forms to build the Q/A pairs. We sustained the true/false distributions in the training and testing datasets around 25% true-75% false.

Each of the Q/A pairs, i.e., from TREC and RTE, are used to extract features to indicate true/false entailment. Thus the

[2] available at http://trec.nist.gov/data/qa.html

[3] available at http://lucene.apache.org/java/docs/

[4] Recognizing Textual Entailment Challenge, datasets available at http://www.nist.gov/tac/tracks/2008/rte/index.html

ISBN: 978-989-95079-6-8

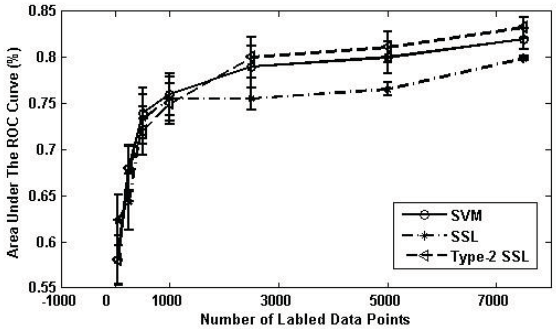


Figure 4: AUC performance on Testing Dataset.

aim of the model is to identify if the candidate sentence entails the question. To measure entailment we extracted 20 different features based on syntactic and semantic match between the question and answer sentences. Some of our features include syntactic matching (keyword and phrase match), named-entity match, headword match using hypernym, hyponym extensions of words, answer type match, which is determined from the question. An automatic hypernym relation extraction algorithm is implemented to capture semantic structures in sentences, such as 'a.k.a', 'such as', apposition relations, to list a few. In addition, we used semantic features such as sentence structure match, i.e., subject, object, headword match, named entity match based on our named entity module to match 50 fine objects in sentence pairs such as time, place, location, person, descriptions, reason, numbers, etc.

In this experiment we used around 600 training questions to compile around 7500 training data points. Similarly, using the entire 202 TREC04 factoid questions as testing, we compiled around 4000 testing data points. Among the training datasets, we used different number of labeled data points to compile series of experiments. Since the data points are randomly selected from the pool, we repeated the experiments 10 times and captured the standard errors to measure robustness. We measured the accuracy of the true/false entailments based on **A**rea **U**nder the Receiver Operating **C**urve statistic. We repeated the experiment using different training sets of different sizes. We built type-2 SSL models with parameters, i.e., $k = \{3, 5, 10, 50\}$, $m = \{1.4, 1.5, ..., 2.6\}$, $\sigma = \left\{10^{-2}, ..10^{2}\right\}$ and tested the AUC performance on the same testing dataset. We compared the performance of our models with the ones obtained from application of the same experiments using standard SSL method [20] with the same learning parameters as well as the SVM methods using learning parameters $C = \left\{2^{-1}, ..., 2^{8}\right\}$ and $\gamma = \{1 - default\}$. The AUC performance comparison with the well-known SSL and SVM method is shown Figure 4. The results of the experiment on QA datasets reveals that the new soft SSL outperforms the standard SSL method. The results are comparable with the state-of the art SVM tool. Based on the current experiment we can suggest that the presented method is an alternative method for linguistic analysis when there are not enough labeled data but more unlabeled data present.

7 Conclusions

We presented a new uncertainty modeling tool for graph-based semi-supervised learning methods using type2 fuzzy number arithmetic. We learn the interval valued degree of similarity values based on fuzzy k-nearest approach. Each edge linking weight is represented with a type-2 fuzzy number and arithmetic operations on type-2 fuzzy numbers are implemented to build the type-2 fuzzy graph Laplacian and learn a classifier function. Experiments on information extraction from unstructured text for question/answering task have shown promising results.

References

[1] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Scholkopf. Learning with local and global consistency. *Advances in Neural Information Processing Systems*, 16:321–328, 2004.

[2] X.-H. Wu and J.-J. Zhou. Kernel-based fuzzy k-nearest neigbor algorithm. 2:159–162, 2005.

[3] W. Blewitt, S.M. Zhou, and S. Coupland. A novel approach to type-2 fuzzy addition. *Proc. Intern. Conference on Fuzzy Systems*, pages 1–6, 2007.

[4] Lotfi A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[5] E.-M. Voorhees. The trec question answering track. *Natural Language Engineering*, 7(4):361–378, 2001.

[6] D. Zhou and B. Scholkopf. Learning from labeled and unlabeled data using random walks. pages 237–244, 2004.

[7] X. Zhu, J. Lafferty, and Z. Ghahramani. *Semi-supervised learning: From Gaussian Fields to Gaussian processes*. Technical Report, Carnegie Mellon University, 2003.

[8] J.M. Keller, M.R. Gray, and J.A. Givens Jr. A fuzzy k-nearest neighbor algorithm. *IEEE Trans. Systems Man and Cybernetics*, 15:580–585, 1985.

[9] A. Celikyilmaz and I.B. Turksen. Uncertainty bounds of fuzzy c-regression method. *IEEE-FUZZ Int. Conf. on Fuzzy Systems, Hong Kong*, pages 1098–7584, June 2008.

[10] A. Celikyilmaz and I. B. Turksen. Uncertainty modeling with evolutionary improved fuzzy functions approach. *IEEE Systems, Man, and Cybernetics- Part B*, 38(4):1098–1110, 2008.

[11] C. Hwang and F. C.-H. Rhee. Uncertainty fuzzy clustering: Interval type-2 fuzzy approach to c-means. *IEEE Trans. On Fuzzy Systems*, 15(1):107–120, 2007.

[12] J.C. Bezdek. *Pattern recognition with fuzzy objective functions*. Plenum Press, New York, 1981.

[13] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning - (i-ii-iii). *Information Sciences*, 8-9, 1975.

[14] J.M. Mendel and R.I. John. Type-2 fuzzy sets made simple. 10(2):117–127, 2002.

[15] S. Coupland and R. John. An approach to type-2 fuzzy arithmetic. *Proc. UK Workshop on Computational Intelligence*, pages 107–114, 2003.

[16] A. Celikyilmaz and I.B. Turksen. Uncertainty modeling with evolutionary improved fuzzy functions approach. *IEEE Systems, Man, and Cybernetics- Part B*, 38(4):1098–1110, 2008.

[17] O. Castillo and P. Melin. *Type-2 Fuzzy Logic: Theory and Applications*. Springer-Verlag, Heidelberg, Germany, 2008.

[18] M. B. Begian, W. Melek, and J.M. Mendel. Parametric design of stable type-2 tsk fuzzy systems. 2008.

[19] B.E. Boser, I. Guyon, and V. Vapnik. A training algorithm for optimal margin classifiers. *Proc. of Fifth Annual Workshop on Computational Learning Theory*, 3:144–152, 1992.

[20] X. Zhu and Z. Ghahramani. *Learning from labeled and unlabeled data with label propagation*. Technical Report CMU-CALD-02-107, Carnegie Mellon University, Pittsburgh, 2002.

ISBN: 978-989-95079-6-8

Can We Learn Algorithms from People Who Compute Fast: An Indirect Analysis in the Presence of Fuzzy Descriptions

Olga Kosheleva[1]

1.Department of Teacher Education
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
olgak@utep.edu

Abstract— *In the past, mathematicians actively used the ability of some people to perform calculations unusually fast. With the advent of computers, there is no longer need for human calculators – even fast ones. However, recently, it was discovered that there exist, e.g., multiplication algorithms which are much faster than standard multiplication. Because of this discovery, it is possible than even faster algorithm will be discovered. It is therefore natural to ask: did fast human calculators of the past use faster algorithms – in which case we can learn from their experience – or they simply performed all operations within a standard algorithm much faster? This question is difficult to answer directly, because the fast human calculators' self-description of their algorithm is very fuzzy. In this paper, we use an indirect analysis to argue that fast human calculators most probably used the standard algorithm.*

Keywords— Fast Fourier transform, fast human calculators, fast multiplication, fuzzy description

1 People Who Computed Fast: A Historical Phenomenon

In history, several people have been known for their extraordinary ability to compute fast. Before the 20 century invention of computers, their computational abilities were actively used.

For example, in the 19 century, Johann Martin Zaharias Dase performed computations so much faster than everyone else that professional mathematicians hired him to help with their calculations; see, e.g., [1, 3]. Dase:

- computed π with a record-breaking accuracy of 200 digits,
- calculated the logarithms table with 7 digit accuracy, and
- performed many other computational tasks.

Karl Friedrich Gauss himself recommended that Dase be paid by the Academy of Science to perform the calculations.

2 People Who Computed Fast: How They Computed?

Calculations have been (and are) important in many practical problems. Because of this practical importance, people have therefore always been trying to speed up computations. One natural way to speed up computations is to learn from the people who can do computations fast.

3 People Who Computed Fast: Their Self-Explanations Were Fuzzy and Unclear

In spite of numerous attempts to interview the fast calculators and to learn from them how they compute, researchers could not extract a coherent algorithm. One of the main reasons is that most of the fast calculators were *idiots savants*: their intellectual abilities outside computations were below average. Their explanations of their algorithms were always fuzzy and imprecise, formulated in terms of words of natural language rather than in precise mathematical terms.

To us familiar with fuzzy logic and its applications this fuzziness should not be surprising. It is normal that people in general – and not necessarily idiots savants – cannot precisely describe

- how they drive,
- how they operate machines,
- how they walk,
- how they translate from one language to another,
- how they recognize faces,
- how they control different situations, etc.

– this is why fuzzy control and other fuzzy techniques, techniques for translating this fuzzy description into a precise strategy, have so many practical applications; see, e.g., [4].

What may be somewhat unusual about fast computations is that while the self-description of fast calculators was fuzzy and imprecise, the results of the computations were always correct and precise.

4 With the Appearance of Computers, the Interest in Fast Human Calculators Waned

With the appearance of computers, the need for human calculators disappeared, and the interest in their skills waned. Such such folks may provide a good entertainment, they may be of interest to psychologists who study how we reason and how we perform menial tasks – but mathematicians were no longer interested.

Yes, fast human calculators can perform calculations faster than an average human being – but electronic computers can perform the same computations much much faster. From this viewpoint, fast human calculators remained a curiosity.

ISBN: 978-989-95079-6-8

5 Why Interest Waned: Implicit Assumptions

One of the main reasons why in the 1940s and 1950s the interest in fast human calculators waned is that it was implicitly assumed that the standard techniques of addition and multiplication are the best.

For example, it was assumed that the fastest way to add the two n-digit numbers is to add them digit by digit, which requires $O(n)$ operations with digits.

It was also implicitly assumed that the fastest way to multiply two n-digit numbers is the standard way to multiply the first number by each of the digits of the second numbers, and then to add the results. Each multiplication by a digit and each addition requires $O(n)$ steps, thus multiplication by all n digits and the addition of all n results require $n \cdot O(n) = O(n^2)$ computational steps.

From this viewpoint, the only difference between a normal human calculator, a fast human calculator, and an electronic computer is in the speed with which we can perform operations with digits. From this viewpoint, the only thing we can learn from fast human calculators is how to perform operations with digits faster. Once the electronic computers became faster than fast human calculators, the need to learn from the fast human calculators disappeared.

6 A Surprising 1960s Discovery of Fast Multiplication Algorithms

The above implicit assumption about arithmetic operations was not questioned until a surprising sequence of discoveries was made in the 1960s; see, e.g., [2].

These discoveries started with the discovery of the Fast Fourier Transform algorithm, an algorithm that enables us to compute the Fourier Transform

$$\hat{f}(\omega) = \frac{1}{\sqrt{2\pi}} \cdot \int f(t) \cdot \exp(\mathrm{i} \cdot \omega \cdot t)\, dt, \tag{1}$$

where $\mathrm{i} = \sqrt{-1}$, in time $O(n \cdot \log(n))$ – instead of the $n \cdot O(n) = O(n^2)$ time needed for a straightforward computation of each of n values of $\hat{f}(\omega)$ as an integral (i.e., in effect, a sum) over n different values $f(t)$.

The ability to compute Fourier transform fast lead to the ability to speed the computation of the convolution of two functions:

$$h(t) \stackrel{\text{def}}{=} \int f(s) \cdot g(t-s)\, ds. \tag{2}$$

A straightforward computation of the convolution requires that for each of the n values $h(t)$, we compute the integral (sum) of n different products $f(s) \cdot g(t-s)$ corresponding to n different values s. Thus, the straightforward computation requires $O(n^2)$ computational steps.

Indeed, it is known that the Fourier transform of the convolution is equal to the product of Fourier transforms. Thus, to compute convolution, we can do the following:

- first, we compute Fourier transforms $\hat{f}(\omega)$ and $\hat{g}(\omega)$;
- then, we compute the Fourier transform of h as

$$\hat{h}(\omega) = \hat{f}(\omega) \cdot \hat{g}(\omega); \tag{3}$$

- finally, we apply the inverse Fourier transform to the function $\hat{h}(\omega)$ and compute the desired convolution $h(t)$.

What is the computation time of this algorithm?

- Both Fourier transform and inverse Fourier transform can be computed in time $O(n \cdot \log(n))$.
- The point-by-point multiplication $\hat{h}(\omega) = \hat{f}(\omega) \cdot \hat{g}(\omega)$ requires n computational steps.

Thus, the overall computation time requires

$$O(n \cdot \log(n)) + n + O(n \cdot \log(n)) = O(n \cdot \log(n)) \tag{4}$$

steps, which is much faster than $O(n^2)$.

V. Strassen was the first to notice, in 1968, than this idea can lead to fast multiplication of long integers. Indeed, an integer x in a number system with base b can be represented as a sum

$$x = \sum_{i=1}^{n} x_i \cdot b^i. \tag{5}$$

In these notations, the product $z = x \cdot y$ of two integers $x = \sum\limits_{i=1}^{n} x_i \cdot b^i$ and $y = \sum_{i=1}^{n} y_i \cdot b^i$ can be represented as

$$z = \left(\sum_{i=1}^{n} x_i \cdot b^i\right) \cdot \left(\sum_{j=1}^{n} y_j \cdot b^j\right) =$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} x_i \cdot y_j \cdot b^i \cdot b^j = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i \cdot y_j \cdot b^{i+j}. \tag{6}$$

By combining terms at different values b^k, we conclude that

$$z = \sum_{k=1}^{n} z_k \cdot b^k, \tag{7}$$

where

$$z_k = \sum_{i} x_i \cdot y_{k-i}. \tag{8}$$

This is convolution, and we know that convolution can be computed in time $O(n \cdot \log(n))$.

The values z_k are not exactly the digits of the desired number z, since the sum (8) can exceed the base b. Thus, some further computations are needed. However, even with these further computations, we can multiply two numbers in almost the same time

$$O(n \cdot \log(n) \cdot \log(\log(n))). \tag{9}$$

The corresponding algorithm, first proposed by A. Schönhage and V. Strassen in their 1971 paper [5], remains the fastest known – and it is much faster than the standard $O(n^2)$ algorithm.

7 Fast Multiplication: Open Problems

Fast algorithms drastically reduced the computation time. Fast Fourier transform is one of the main tools for signal processing. These successes has led to the need to find faster and faster algorithms. From this viewpoint, it is is desirable to look for even faster algorithms for multiplying numbers.

The fact that researchers succeeded in discovering algorithms which are much faster than traditional multiplication gives us hope that even faster algorithms can be found.

ISBN: 978-989-95079-6-8

8 Interest in Fast Human Calculators Revived

Where to look for these algorithms? One natural source is folks who did compute fast. From this viewpoint, the fact that they were unable to clearly explain what algorithm they used becomes an advantage: maybe the algorithm that they actually used is some fast multiplication algorithm? This possibility revived an interest in fast human calculators.

So, the question is:

- did fast human calculators use fast multiplication algorithm(s), or
- they used the standard algorithm but simply performed operations with digits faster?

9 Direct Analysis Is Impossible

The most well-known fast human calculator, Johann Martin Zacharias Dase, died almost 150 years ago. Even when he was alive, his self-descriptions were not sufficient to find out how exactly he performed the computations. There might have been hope that our knowledge of fast multiplication algorithms can help in this understanding, but it did not work out. In other words, the direct analysis of Dase's behavior has been impossible – and it is still impossible. We therefore need to perform an *indirect* analysis.

10 Indirect Analysis: Main Idea

A natural way to check which algorithm is used by a computational device – be it a human calculator or an electronic computer – is to find out how the computation time changes with the size n (= number of digits) of the numbers that we are multiplying.

- If this computation time grows with n as n^2, then it is reasonable to conclude that the standard algorithm is used – since for this algorithm, the computation time grows as n^2.
- On the other hand, if the computation time grows with n as $\approx n \cdot \log(n)$ (or even slower), then it is reasonable to conclude that a fast multiplication algorithm is used.
  - It maybe a Strassen-type algorithm, for which the computation time grows as $n \cdot \log(n)$.
  - It may be a (yet unknown) faster algorithm in which case the computation time grows even slower than $n \cdot \log(n)$.

11 Data That We Can Use

Interestingly, there is a data on the time that Dase needed to perform multiplication of numbers of different size. This data comes from fact that Dase's performance was analyzed and tested by several prominent mathematicians of his time – including Gauss himself. Specifically:

- Dase multiplied two 8-digit numbers in 54 seconds;
- he multiplied two 20-digit numbers in 6 minutes;
- he multiplied two 40-digit numbers in 40 minutes; and
- he multiplied two 100-digit numbers in 8 hours and 45 minutes.

12 Analysis

For the standard multiplication algorithm, the number of computational steps grows with the numbers size n as n^2. Thus, for the standard multiplication algorithm, the computation time for performing the computation also grows as n^2:

$$t(n) = C \cdot n^2. \tag{10}$$

So, for this algorithm, two different number sizes $n_1 < n_2$, we would have $t(n_1) = C \cdot n_1^2$ and $t(n_2) = C \cdot n_2^2$ and thus,

$$\frac{t(n_2)}{t(n_1)} = \frac{n_2^2}{n_1^2} = \left(\frac{n_2}{n_1}\right)^2. \tag{11}$$

For a faster algorithm, e.g., for the algorithm that requires $O(n \cdot \log(n))$ running time, the corresponding ratio will be smaller:

$$\frac{t(n_2)}{t(n_1)} = \frac{n_2 \cdot \log(n_2)}{n_1 \cdot \log(n_1))} = \frac{n_2}{n_1} \cdot \frac{\log(n_2)}{\log(n_1)}. \tag{12}$$

Thus, to check whether a human calculators uses the standard algorithm or a faster one it is sufficient to compare the the corresponding time ratio $\frac{t(n_2)}{t(n_1)}$ with the square $\left(\frac{n_2}{n_1}\right)^2$:

- If the time ratio is smaller than the square, this means that the human calculator used an algorithm which is much faster than the standard one.
- On the other hand, if the time ratio is approximately the same as the square, thus means that the human calculators most probably used the standard algorithm – or at least some modification of it that does not drastically speed up the computations.
- If it turns out that the time ratio is larger than the square, this would mean, in effect, that a human calculator used an algorithm which is asymptotically even slower than the standard one – this can happen, e.g., if there is an overhead needed to store values etc.

According to the above data, we have:

- $t(8) = 0.9$ minutes,
- $t(20) = 6$ minutes,
- $t(40) = 40$ minutes, and
- $t(100) = 8 \cdot 60 + 45 = 525$ minutes.

Here, for $n_1 = 8$ and $n_1 = 20$, we have:

$$\frac{t(n_2)}{t(n_1)} = \frac{t(20)}{t(8)} = \frac{6}{0.9} \approx 6.7 > \left(\frac{n_2}{n_1}\right)^2 =$$

$$\left(\frac{20}{8}\right)^2 = 2.5^2 = 6.25. \tag{13}$$

For $n_1 = 8$ and $n_2 = 40$, we have

$$\frac{t(n_2)}{t(n_1)} = \frac{t(40)}{t(8)} = \frac{40}{0.9} \approx 44 > \left(\frac{n_2}{n_1}\right)^2 =$$

ISBN: 978-989-95079-6-8

$$\left(\frac{40}{8}\right)^2 = 5^2 = 25. \quad (14)$$

Finally, for $n_1 = 8$ and $n_2 = 100$, we have

$$\frac{t(n_2)}{t(n_1)} = \frac{t(100)}{t(8)} = \frac{525}{0.9} \approx 583 > \left(\frac{n_2}{n_1}\right)^2 = \left(\frac{100}{8}\right)^2 = 12.5^2 \approx 156. \quad (15)$$

In all these cases, the computation time of a human calculator grows faster then n^2 corresponding to the standard algorithm.

The same conclusion can be made if instead of comparing each value n with the smallest value n_1, we compare these values with each other. For $n_1 = 20$ and $n_2 = 40$, we have

$$\frac{t(n_2)}{t(n_1)} = \frac{t(40)}{t(20)} = \frac{40}{6} \approx 7 > \left(\frac{n_2}{n_1}\right)^2 = \left(\frac{40}{20}\right)^2 = 2^2 = 4. \quad (16)$$

For $n_1 = 20$ and $n_2 = 100$, we have

$$\frac{t(n_2)}{t(n_1)} = \frac{t(100)}{t(20)} = \frac{525}{6} \approx 88 > \left(\frac{n_2}{n_1}\right)^2 = \left(\frac{100}{20}\right)^2 = 5^2 = 25. \quad (17)$$

Finally, for $n_1 = 40$ and $n_2 = 100$, we have

$$\frac{t(n_2)}{t(n_1)} = \frac{t(100)}{t(40)} = \frac{525}{40} \approx 13 > \left(\frac{n_2}{n_1}\right)^2 = \left(\frac{100}{40}\right)^2 = 2.5^2 = 6.25. \quad (18)$$

Thus, it is reasonable to conclude that the fast human calculators did not use any algorithm which is faster than the standard one.

13 Possible Future Work

People who perform computations fast appear once in a while. It may be a good idea to record and analyze their computation time – and maybe record their fuzzy explanations and try to make sense of them.

Acknowledgment

The author is thankful to the anonymous referees for useful suggestions.

References

[1] P. Beckmann, *A History of* π, Barnes & Noble, New York, 1991.

[2] Th. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, MIT Press, Cambridge, MA, 2001.

[3] D. R. Hofstadter, *Godel, Escher, Bach: an Eternal Golden Braid*, Basic Books, 1999.

[4] G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic*, Prentice Hall, New Jersey, 1995.

[5] A. Schönhage and V. Strassen, "Schnelle Multiplikation grosser Zahlen", *Computing*, 1971, Vol. 7, pp. 281–292.

ISBN: 978-989-95079-6-8

A parametric method to solve quadratic programming problems with fuzzy costs

Ricardo C. Silva[1] Carlos Cruz[2] Akebo Yamakami[1]

1.Department of Telematics, School of Electrical and Computer Engineering,
University of Campinas, P.O.Box 6101, 13083-970, Campinas-SP, Brazil.
Email: {rcoelhos,akebo}@dt.fee.unicamp.br
2.Department of Computer Science and Artificial Intelligence,
University of Granada, E-18071, Granada, Spain
Email: carloscruz@decsai.ugr.es

Abstract— *This work describes a novel fuzzy-sets-based method to solve a particular class of quadratic programming problems which have vagueness coefficients in the objective function. Quadratic programming problems are of utmost importance in an increasing variety of practical fields. In addition, as the ambiguity and vagueness are natural and ever-present in real-life situations requiring solutions, it makes perfect sense to attempt to address them using fuzzy quadratic programming problems. Also, two other methods to solve this kind of problems are briefly described. The proposal uses two phases to solve fuzzy quadratic programming problems. In the first, phase we parametrize the fuzzy problem in several classical alpha-problems with different cutting levels. In the second, phase each of these alpha-problems is solved by using conventional solving techniques. The final fuzzy solution to the problem is obtained by integrating all of these particular alpha-solutions. The results obtained using these two methods are compared with the two-phased proposal outlined above.*

Keywords— Fuzzy sets, decision making, fuzzy mathematical programming, quadratic optimization.

1 Introduction

The conception of Artificial Intelligence (AI) in the early fifties was inspired by the wide variety of physical and mental tasks performed by humans without any measurements and any computations. Thus, this forward thinking described the definition of AI as it is widely known: "the study and design of intelligent agents [14]", where an intelligent agent is a system that perceives its environment and takes actions which maximize its chances of success[15]. The capability to compute and reason with perception-based information can be applied to real-world problems in which decision-relevant information is a mixture of measurements and perceptions [20], where in general, measurements are crisp, although perceptions are fuzzy.

The literature shows that the best way of modeling these types of problems is using Soft Computing methodologies. According to Verdegay et al.[17], Soft Computing is a family of problem-resolution methods headed by approximate reasoning, functional and optimization approximation methods, which now include search methods. More specifically, the search methods use thorough evaluations to find regions with approximate points of optimal solutions that are obtained by using optimization methods.

These methods, which are algorithms based on mathematical knowledge, are used to guarantee the convergence to the optimal solution. They are covered by the area of mathematical programming. This area has several sub-fields where the quadratic programming is a special type that optimizes a quadratic objective function of several variables subject to linear constraints. A precise definition of the constraints and the objective function to be optimized is necessary to obtain exact optimal solutions of quadratic programming problems.

Nevertheless, in most real practical applications (portfolio, game theory, engineering modeling, design and control, logistics, etc.) one lacks this kind of exact knowledge, and only approximate, vague and imprecise values are known. Moreover, these imprecise values can be dealt with fuzzy logic. In this case, the concept of fuzzy mathematical programming emerges when it is used.

With this in mind, the objective of this paper is to review some methods that are related to quadratic topic and to outline a soft computing model used to solve novel fuzzy quadratic programming problems.

The paper is organized as follows: Section 2 briefly introduces the different approaches to solving the quadratic programming problem with fuzzy costs. In this section, a novel approach is developed to solve quadratic programming problems with fuzzy costs where these are transformed into parametrical quadratic multiobjective programming problems. To clarify the above developments, three numerical examples are analyzed in section 3. Finally, conclusions are presented in Section 4.

2 Quadratic optimization with fuzzy costs

An optimization problem that is described with a quadratic objective function subject to linear constraints is called a "Quadratic Programming" problem. QP can be viewed both as a special case of the nonlinear programming and a generalization of the linear programming. Several applications and methods can be found in [3, 8, 9, 16, 19]. A quadratic programming problem can be formulated as

$$\begin{array}{ll} \min & \mathbf{c^t x} + \frac{1}{2}\mathbf{x^t Q x} \\ \text{s.t.} & \mathbf{A x} \leq \mathbf{b} \\ & \mathbf{x} \geq 0 \end{array} \tag{1}$$

where $\mathbf{x}$ and $\mathbf{c}$ are n vectores of real numbers, $\mathbf{Q}$ is an $n \times n$ symmetric matrix of real numbers, $\mathbf{A}$ is an $m \times n$ matrix of real numbers and $\mathbf{b}$ is an m vector of real numbers.

ISBN: 978-989-95079-6-8

In this work, quadratic programming problems with fuzzy costs are considered because there are cases of real world problems whose parameters are seldom known exactly and have to be estimated by the decision maker. Hence, the $n \times n$ symmetric matrix $\mathbf{Q}$ and the n vector $\mathbf{c}$ have fuzzy numbers in these components. The application of fuzzy logic is also a way to describe, mathematically, this vagueness as described in [2, 10, 13, 18]. The uncertainties of costs of the objective function can be dealt with by fuzzy numbers.

Thus, this set of problems with uncertain costs can be formalized in the following form:

$$\begin{array}{ll} \min & \tilde{\mathbf{c}}^{\mathbf{t}}\mathbf{x} + \frac{1}{2}\mathbf{x}^{\mathbf{t}}\tilde{\mathbf{Q}}\mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} \leq \mathbf{b} \\ & \mathbf{x} \geq 0 \end{array} \tag{2}$$

where the fuzzy numbers are characterized by membership functions that are defined by decision makers. The membership functions can be defined as

$$\mu_j, \mu_{ij} : \mathbb{R} \to [0,1], \;\; i,j \in \mathbb{J} = \{1,2,\ldots,n\}$$

In particular these membership functions will be supposed as:

$$\mu_j(y) = \begin{cases} 0 & \text{if } c_j^U \leq y \text{ or } y \leq c_j^L \\ h_j(y) & \text{if } c_j^L \leq y \leq c_j \\ g_j(y) & \text{if } c_j \leq y \leq c_j^U \end{cases} \qquad j \in \mathbb{J} \tag{3}$$

and

$$\mu_{ij}(y) = \begin{cases} 0 & \text{if } q_{ij}^U \leq y \text{ or } y \leq q_{ij}^L \\ h_{ij}(y) & \text{if } q_{ij}^L \leq y \leq q_{ij} \\ g_{ij}(y) & \text{if } q_{ij} \leq y \leq q_{ij}^U \end{cases} \qquad i,j \in \mathbb{J} \tag{4}$$

where $\mathbf{h}(\cdot)$ and $\mathbf{g}(\cdot)$ are assumed to be strictly increasing and decreasing continuous functions, respectively, $h_j(c_j) = g_j(c_j) = 1,\ j \in \mathbb{J}$ and $h_{ij}(q_{ij}) = g_{ij}(q_{ij}) = 1,\ i,j \in \mathbb{J}$.

Bellman and Zadeh describe the *fuzzy decision* in [4] as the intersection of goals and constraints which is formalized in the following definition:

Definition 1 *Assume that we are given a fuzzy goal μ_G and a fuzzy constraint μ_C in a space of alternative X. Then, μ_G and μ_C combine to form a* fuzzy decision*, μ_D, which is a fuzzy set resulting from intersection of μ_G and μ_C. In symbols*

$$\sup_{\mathbf{x} \in X} \mu_D(\mathbf{x}) = \sup_{\mathbf{x} \in X} [\mu_G(\mathbf{x}) \wedge \mu_C(\mathbf{x})]$$

Note that in Definition 1, the goals and the constraints enter into the expression for μ_D in the same way. Thus, it is able to find a maximizing decision to an extremum problem for a scalar function, as described in [13]. Let ϕ be denoted by $\phi : [0,1] \to [0,1]$ that implies $\phi(\alpha) = \sup_{\mathbf{x} \in \mathbf{X}(\alpha)} \mu_G(\mathbf{x})$ where $X(\alpha) = \{\mathbf{x} \in \mathbb{R}^n \mid \mu_X(\mathbf{x}) \geq \alpha\}$. If ϕ is continuous in [0,1] then ϕ has a *fix point* $\overline{\alpha}$ and, therefore,

$$\sup_{\mathbf{x} \in X} \mu_D(\mathbf{x}) = \sup_{\mathbf{x} \in X(\overline{\alpha})} \mu_G(\mathbf{x}) = \overline{\alpha}$$

.

A novel parametric approach to solve quadratic programming problems with fuzzy costs will be presented in the next sub-section. This approach transforms the original problem into a parametric multi-objective quadratic programming problem.

2.1 Multi-objective approach

A multi-objective approach to solve a linear programming problem with imprecise costs is described in [6, 7]. As the linear problem is a particular case of quadratic problem, this approach can be extended to solve quadratic programming problems with fuzzy costs.

The quadratic objective function can be divided into two parts, where the first one is a linear term and the second one is a quadratic term. According to this, the fuzzy costs can only be in the first part or only the second part or in both. In this section, these three ways will be presented separately.

The linear problem considered in [6] used trapezoid membership functions for the costs but here, for the sake of simplicity, they will be supposed to be like (3) and (4).

Then, by considering the $(1-\alpha)$-cut of every cost, $\alpha \in [0,1]$,

$$\begin{array}{l} \forall\, x \in \mathbb{R},\;\; \mu_j(x) \geq 1-\alpha \Leftrightarrow h_j^{-1}(1-\alpha) \leq x \leq g_j^{-1}(1-\alpha), \\ \forall\, x \in \mathbb{R},\;\; \mu_{ij}(x) \geq 1-\alpha \Leftrightarrow h_{ij}^{-1}(1-\alpha) \leq x \leq g_{ij}^{-1}(1-\alpha), \end{array}$$

where $i,j \in \mathbb{J} = \{1,2,\ldots,n\}$.

2.1.1 Quadratic programming problem with fuzzy costs vector of linear term

The approaches described in this section solve a quadratic programming problem with a vector $\tilde{\mathbf{c}}$ of fuzzy numbers. This vector describes the costs of the objective function and Problem 2 can be rewritten as follows:

$$\begin{array}{ll} \min & \tilde{\mathbf{c}}^{\mathbf{t}}\mathbf{x} + \frac{1}{2}\mathbf{x}^{\mathbf{t}}\mathbf{Q}\mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}. \end{array} \tag{5}$$

Thus, according to some results by Bitran [5], it has been shown that a fuzzy solution to (5) may be found from the parametric solution of the multi-objective parametric quadratic programming problem

$$\begin{array}{ll} \min & \left[(\mathbf{c^1})^{\mathbf{t}}\mathbf{x} + \frac{1}{2}\mathbf{x}^{\mathbf{t}}\mathbf{Q}\mathbf{x}, (\mathbf{c^2})^{\mathbf{t}}\mathbf{x} + \frac{1}{2}\mathbf{x}^{\mathbf{t}}\mathbf{Q}\mathbf{x}, \ldots, (\mathbf{c^{2^n}})^{\mathbf{t}}\mathbf{x} + \frac{1}{2}\mathbf{x}^{\mathbf{t}}\mathbf{Q}\mathbf{x}\right] \\ \text{s.t.} & \mathbf{A}\mathbf{x} \leq \mathbf{b},\;\; \mathbf{x} \geq \mathbf{0}, \\ & \mathbf{c^k} \in \mathbb{E}(1-\alpha),\;\; \alpha \in [0,1],\;\; k = 1,2,\ldots,2^n, \end{array} \tag{6}$$

where $\mathbb{E}(1-\alpha)$, for each $\alpha \in [0,1]$, is the set of vectors in $\mathbb{R}^n$ such that each of its components is either in the lower bound, $h_j^{-1}(1-\alpha)$, or in the upper bound, $g_j^{-1}(1-\alpha)$, of the respective $(1-\alpha)$-cut, that is, $\forall\, k = 1,2,\ldots,2^n$, and $\forall\, j \in \mathbb{J}$

$$c^k = (c_1^k, c_2^k, \ldots, c_n^k) \in \mathbb{E}(1-\alpha) \Leftrightarrow c_j^k = \begin{cases} h_j^{-1}(1-\alpha) & \text{or} \\ g_j^{-1}(1-\alpha). \end{cases}$$

Now observe that to find a parametric solution to (6), from which one can obtain a fuzzy solution to (5), one may use any classical multiobjective quadratic programming approach.

2.1.2 Quadratic programming problem with fuzzy costs matrix of quadratic term

The approach described in this section solves a quadratic programming problem with matrix $\tilde{\mathbf{Q}}$ of fuzzy numbers. This

ISBN: 978-989-95079-6-8

vector describes the costs of the objective function and Problem 2 can be rewritten as follows

$$\begin{aligned} \min \quad & \mathbf{c^t x} + \tfrac{1}{2}\mathbf{x^t \tilde{Q} x} \\ \text{s.t.} \quad & \mathbf{Ax} \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned} \tag{7}$$

As it was shown above, a fuzzy solution to (7) may be found from the parametric solution of the multiobjective parametric quadratic programming problem

$$\begin{aligned} \min \quad & \left[\mathbf{c^t x} + \tfrac{1}{2}\mathbf{x^t Q^1 x}, \mathbf{c^t x} + \tfrac{1}{2}\mathbf{x^t Q^2 x}, \ldots, \mathbf{c^t x} + \tfrac{1}{2}\mathbf{x^t Q^{2^{n^2}} x}\right] \\ \text{s.t.} \quad & \mathbf{Ax} \leq \mathbf{b}, \ \ \mathbf{x} \geq \mathbf{0}, \\ & \mathbf{Q^k} \in \mathbb{E}(1-\alpha), \ \ \alpha \in [0,1], \ \ k = 1,2,\ldots,2^{n^2}, \end{aligned} \tag{8}$$

where $\mathbb{E}(1-\alpha)$, for each $\alpha \in [0,1]$, is the set of vectors in $\mathbb{R}^n$ such that each of its components is either in the lower bound, $h_{ij}^{-1}(1-\alpha)$, or in the upper bound, $g_{ij}^{-1}(1-\alpha)$, of the respective $(1-\alpha)$-cut, that is, $\forall\, k = 1,2,\ldots,2^{n^2}$, and $\forall\, i,j \in \mathbb{J}$

$$\mathbf{Q^k} = (q_{11}^k, \ldots, q_{1n}^k, \ldots, q_{nn}^k) \in \mathbb{E}(1-\alpha) \Leftrightarrow$$
$$\Leftrightarrow q_{ij}^k = \begin{cases} h_{ij}^{-1}(1-\alpha) & \text{or} \\ g_{ij}^{-1}(1-\alpha). \end{cases}$$

Now observe that to find a parametric solution to (8), from which one can obtain a fuzzy solution to (7), one may use any classical multiobjective quadratic programming approach.

2.1.3 *Quadratic programming problem with all fuzzy costs*

Thus, according to the parametric transformations shown above, a fuzzy solution to (2) may be found from the parametric solution of the multiobjective parametric quadratic programming problem

$$\begin{aligned} \min \quad & \left[(\mathbf{c^1})^\mathbf{t}\mathbf{x} + \tfrac{1}{2}\mathbf{x^t Q^1 x}, (\mathbf{c^2})^\mathbf{t}\mathbf{x} + \tfrac{1}{2}\mathbf{x^t Q^1 x}, \ldots,\right. \\ & \ldots, (\mathbf{c^{2^n}})^\mathbf{t}\mathbf{x} + \tfrac{1}{2}\mathbf{x^t Q^1 x}, (\mathbf{c^1})^\mathbf{t}\mathbf{x} + \tfrac{1}{2}\mathbf{x^t Q^2 x}, \ldots, \\ & \left. \ldots, (\mathbf{c^{2^n}})^\mathbf{t}\mathbf{x} + \tfrac{1}{2}\mathbf{x^t Q^2 x}, \ldots, (\mathbf{c^{2^n}})^\mathbf{t}\mathbf{x} + \tfrac{1}{2}\mathbf{x^t Q^{2^{n^2}} x}\right] \\ \text{s.t.} \quad & \mathbf{Ax} \leq \mathbf{b}, \ \ \mathbf{x} \geq \mathbf{0}, \\ & \mathbf{c^k}, \mathbf{Q^p} \in \mathbb{E}(1-\alpha), \ \ \alpha \in [0,1], \\ & k = 1,2,\ldots,2^n \text{ and } p = 1,2,\ldots,2^{n^2}, \end{aligned} \tag{9}$$

where $\mathbb{E}(1-\alpha)$, for each $\alpha \in [0,1]$, is the set of vectors in $\mathbb{R}^n$ such that each of its components is either in the lower bound, $h_j^{-1}(1-\alpha)$, or in the upper bound, $g_j^{-1}(1-\alpha)$, of the respective $(1-\alpha)$-cut, that is, $\forall\, k = 1,2,\ldots,2^n$, and $\forall\, i,j \in \mathbb{J}$

$$c^k = (c_1^k, c_2^k, \ldots, c_n^k) \in \mathbb{E}(1-\alpha) \Leftrightarrow c_j^k = \begin{cases} h_j^{-1}(1-\alpha) & \text{or} \\ g_j^{-1}(1-\alpha). \end{cases}$$

and

$$\mathbf{Q^k} = (q_{11}^k, \ldots, q_{1n}^k, \ldots, q_{nn}^k) \in \mathbb{E}(1-\alpha) \Leftrightarrow$$
$$\Leftrightarrow q_{ij}^k = \begin{cases} h_{ij}^{-1}(1-\alpha) & \text{or} \\ g_{ij}^{-1}(1-\alpha). \end{cases}$$

The obtained parametrical solutions for any of multiobjective models above, to the different α values, generate a set of solutions and then we use the Representation Theorem to integrate all of these particular alpha-solutions.

2.2 *Liu's approach*

Transforming a fuzzy quadratic programming problem into a two-level mathematical programming problem for finding the bounds of the fuzzy values in the objective function and in the set of constraints is discussed in [11, 12]. These papers describe how to transform the two-level mathematical program into the conventional one-level quadratic program. However, such a method is also valid when only the objective is fuzzy, that is, when only the vector of costs are fuzzy, and for this reason it shall be considered here. According to the goal of this work, the coefficient $\mathbf{c}$ of Problem (5) is fuzzy.

The authors derive the membership function of the goal, and then they apply Zadeh's extension principle to transform the fuzzy quadratic problem into family of classical quadratic problems that can be solved by conventional optimization techniques. Thus, the membership function of the objective function can be defined as

$$\mu_{\tilde{Z}}(z) = \sup_{\mathbf{c}} \min\{\mu_{\tilde{c}_j}(c_j), \ \forall j | z = Z(\mathbf{c})\} \tag{10}$$

where $Z(\mathbf{c})$ is the goal of the conventional quadratic problem. Membership function $\mu_{\tilde{Z}}$ can be computed by finding the functions that describe the shape of the left and right sides of the fuzzy numbers. Then, it is possible to obtain the upper bound of the objective value Z_α^U and your lower bound Z_α^L to each value α. Thus, Z_α^U is the maximum and Z_α^L is the minimum of $Z(\mathbf{c})$, respectively, that can be described as

$$Z_\alpha^U = \max_{c_j} \left\{ \begin{array}{ll} \min_{\mathbf{x}} & \sum_{j=1}^n c_j x_j + \frac{1}{2}\sum_{j=1}^n \sum_{l=1}^n q_{jl} x_j x_l \\ \text{s.a} & \sum_{j=1}^n a_{ij} x_j \leq b_i, \ \ i = 1,\ldots,m \\ & x_j \geq 0, \ \ j = 1,\ldots,n \end{array} \right. \tag{11}$$

$$Z_\alpha^L = \min_{c_j} \left\{ \begin{array}{ll} \min_{\mathbf{x}} & \sum_{j=1}^n c_j x_j + \frac{1}{2}\sum_{j=1}^n \sum_{k=1}^n q_{jk} x_j x_k \\ \text{s.a} & \sum_{j=1}^n a_{ij} x_j \leq b_i, \ \ i = 1,\ldots,m \\ & x_j \geq 0, \ \ j = 1,\ldots,n \end{array} \right. \tag{12}$$

where $c_j \in [(c_j)_\alpha^L, (c_j)_\alpha^U]$, for any $\alpha \in [0,1]$ chosen by decision maker.

These two formulations above are solved by outer-level and inner-level programs, respectively. The outer-level program obtains the values c_j that are used with parameters by inner-level program. The inner-level program solves a classical quadratic programming problem with the data obtained by outer-level program. The authors state that the formulation of two-level quadratic problems is a generalization of the conventional parametrical quadratic programming problem. Thus, the first two-level mathematical program can be transformed into the following quadratic problem by dual formulation:

$$\begin{aligned} \max \quad & \sum_{i=1}^m b_i \lambda_i - \tfrac{1}{2}\sum_{i=1}^n \sum_{j=1}^n q_{ij} x_i x_j \\ \text{s.t.} \quad & \sum_{i=1}^m a_{ij}\lambda_i + \delta_j - \sum_{i=1}^n q_{ij} x_i = c_j, \ \ j = 1,2,\ldots,n \\ & (c_j)_\alpha^L \leq c_j \leq (c_j)_\alpha^U, \ \ j = 1,2,\ldots,n \\ & \lambda_i, \delta_j \geq 0, \ \ i = 1,2,\ldots,m \text{ and } j = 1,2,\ldots,n. \end{aligned} \tag{13}$$

Since both the inner-level program and the outer-level of the second program have the same minimization operation, they can be combined into a conventional one-level program with the constraints of the two programs considered simultaneously. Consequently, some points must be analyzed which

ISBN: 978-989-95079-6-8

are shown in [11, 12]. The second program can be described as

$$\begin{array}{ll}\max & \sum_{j=1}^{n}(c_j)_{\alpha}^{L}+\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}q_{ij}x_ix_j \\ \text{s.t.} & \sum_{i=1}^{n}a_{ij}x_i\leq b_i,\ \ i=1,2,\ldots,m \\ & x_j\geq 0,\ \ j=1,2,\ldots,n\end{array} \tag{14}$$

2.3 Ammar's approach

A method to solve a quadratic programming problem with fuzzy coefficients in the objective function and in the set of constraint is described in [1]. However, such a method is also valid when only the objective is fuzzy, that is, when only the matrix of costs are fuzzy, and for this reason it shall be considered here. This approach can be formulated in the following way:

$$\begin{array}{ll}\min & \mathbf{x^t}\tilde{\mathbf{Q}}\mathbf{x} \\ \text{s.t.} & \mathbf{Ax}\leq\mathbf{b} \\ & \mathbf{x}\geq\mathbf{0}.\end{array} \tag{15}$$

where all decision variables are non negatives and the linear term in objective function is not used. However, in this work only costs of the objective functions are imprecise.

Thus, Problem (15) can be reformulated by using α-cutting levels as

$$\begin{array}{lll}(P_\alpha): & \min & \mathbf{x^t}[\mathbf{Q}_\alpha^-,\mathbf{Q}_\alpha^+]\mathbf{x} \\ & \text{s.a} & \mathbf{Ax}\leq\mathbf{b} \\ & & \mathbf{x}\geq\mathbf{0}.\end{array} \tag{16}$$

where $\alpha\in(0,1]$. $\mathbf{Q}_\alpha^-$ represents a matrix with the lower values of the interval and $\mathbf{Q}_\alpha^+$ represents a matrix with the upper values of the interval.

[1] describes how Problem (16) can be written in two quadratic problems. Then, these problems are formulated as follows

$$\begin{array}{lll}(P_\alpha^-): & \min & \mathbf{x^t}\mathbf{Q}_\alpha^-\mathbf{x} \\ & \text{s.a} & \mathbf{Ax}\leq\mathbf{b} \\ & & \mathbf{x}\geq\mathbf{0}.\end{array} \tag{17}$$

$$\begin{array}{lll}(P_\alpha^+): & \min & \mathbf{x^t}\mathbf{Q}_\alpha^+\mathbf{x} \\ & \text{s.t.} & \mathbf{Ax}\leq\mathbf{b} \\ & & \mathbf{x}\geq\mathbf{0}.\end{array} \tag{18}$$

where the first problem uses the lower bound of the interval of the α-cutting level and the second problem uses the upper bound.

The two quadratic problems (17) and (18) can be solved by using the conditions of Karush-Kuhn-Tucker's optimality to each α value. The optimal solution of the original problem (16), (P_α), is inside the interval formed by the optimal solutions of each one problems (P_α^-) and (P_α^+).

3 Numerical example

In this section, we illustrate a quadratic programming problem that will be solved with three different way. Each way is represented by different fuzzy environment in each problem, i.e., the first problem has fuzzy costs in the linear term, while the second has fuzzy costs in the quadratic term, and the third has fuzzy costs in all coefficients in the objective function.

Example 1 *Consider the following quadratic programming problem with fuzzy coefficients in the linear term:*

$$\begin{array}{ll}\min & (-6,-5,-4)x_1+(1,1.5,2)x_2+2x_1^2-2x_1x_2+x_2^2 \\ \text{s.t.} & x_1+x_2\leq 2 \\ & 2x_1-x_2\leq 4 \\ & x_1,x_2\geq 0.\end{array} \tag{19}$$

The membership functions are:

$$\mu_{-5}(c_1)=\left\{\begin{array}{ll}c_1+6 & -6\leq c_1\leq -5\\ -c_1-4 & -5\leq c_1\leq -4\end{array}\right.$$

$$\mu_{1.5}(c_2)=\left\{\begin{array}{ll}2(c_2-1) & 1\leq c_2\leq 1.5\\ 2(2-c_2) & 1.5\leq c_2\leq 2\end{array}\right.$$

From the membership function for this costs an interval representation according to (3) can be given:

$$\begin{array}{ll}c_1^{0.2}=[-5.8,-4.2], & c_1^{0.4}=[-5.6,-4.4],\\ c_1^{0.6}=[-5.4,-4.6], & c_1^{0.8}=[-5.2,-4.8]\end{array}$$

$$\begin{array}{ll}c_2^{0.2}=[1.1,1.9], & c_2^{0.4}=[1.2,1.8],\\ c_2^{0.6}=[1.3,1.7], & c_2^{0.8}=[1.4,1.6]\end{array}$$

with $c_1^0=[-6,-4]$, $c_1^1=[-5,-5]$ *and* $c_2^0=[1,2]$, $c_2^1=[1.5,1.5]$, *and now* $\mathbb{M}=\{0,0.2,0.4,0.6,0.8,1\}$.

Thus, Problem (19) can be re-formulated as:

$$\begin{array}{ll}\min & (\alpha-6)x_1+\frac{1}{2}(\alpha+2)x_2+2x_1^2-2x_1x_2+x_2^2\\ \min & (\alpha-6)x_1+\frac{1}{2}(4-\alpha)x_2+2x_1^2-2x_1x_2+x_2^2\\ \min & (-\alpha-4)x_1+\frac{1}{2}(\alpha+2)x_2+2x_1^2-2x_1x_2+x_2^2\\ \min & (-\alpha-4)x_1+\frac{1}{2}(4-\alpha)x_2+2x_1^2-2x_1x_2+x_2^2\\ \text{s.t.} & x_1+x_2\leq 2\\ & 2x_1-x_2\leq 4\\ & x_1,x_2\geq 0.\end{array} \tag{20}$$

Problem (20) is solved by weighting objectives, with $\alpha=0.8$*, where an auxiliary problem can be obtained in the following way:*

$$\begin{array}{ll}\min & \omega_1[-5.2x_1+1.4x_2]+\omega_2[-5.2x_1+1.6x_2]+\\ & +\omega_3[-4.8x_1+1.4x_2]+\omega_4[-4.8x_1+1.6x_2]+\\ & +2x_1^2-2x_1x_2+x_2^2\\ s.t. & x_1+x_2\leq 2,\ \ 2x_1-x_2\leq 4,\ \ x_1,x_2\geq 0.\end{array} \tag{21}$$

Different optimal solutions to each of the weighting vectors is obtained by choosing the canonical base of $\mathbb{R}^4$ *as weighting vectors, as it is presented in Table 1.*

Table 1: Result of Problem 21 to each vector ω.

ω	x_1	x_2	FunObj
(1,0,0,0)	1.4600	0.5400	-3.8580
(0,1,0,0)	1.4800	0.5200	-3.7520
(0,0,1,0)	1.4200	0.5800	-3.2820
(0,0,0,1)	1.4400	0.5600	-3.1680

By taking $\omega_1=\omega_2=\omega_3=\omega_4=\frac{1}{4}$ *and solving (21) the optimal solution is* $x^*=[1.45,0.55]$ *and value of objective function is -3.5125.*

On the other hand, according to the method described in 2.2, Problem (19), which is the outer-level program, can be

ISBN: 978-989-95079-6-8

rewritten as:

$$\begin{array}{ll}\max & 2\lambda_1+4\lambda_2-\frac{1}{2}(4x_1^2-4x_1x_2+2x_2^2)\\ \text{s.t.} & \lambda_1+2\lambda_2-\delta_1-4x_1+2x_2=c_1\\ & \lambda_1-1\lambda_2-\delta_2+2x_1-2x_2=c_2\\ & -5.2\le c_1\le -4.8\\ & 1.4\le c_2\le 1.6\\ & x_1,x_2,\lambda_1,\lambda_2,\delta_1,\delta_2\ge 0.\end{array} \tag{22}$$

and the inner-level program can be written as:

$$\begin{array}{ll}\min & -5.2x_1+1.4x_2+\frac{1}{2}(4x_1^2-4x_1x_2+2x_2^2)\\ \text{s.t.} & x_1+x_2\le 2\\ & 2x_1-x_2\le 4\\ & x_1,x_2\ge 0.\end{array} \tag{23}$$

This model is a traditional quadratic program that can represent the two models described in 2.2. Hence, by solving the outer-level program the optimal solution is $x^*=[1.3,0]$ and value of objective function is -3.38. On the other hand, by solving the inner-level program the optimal solution is $x^*=[1.45,0.55]$ and value of objective function is -3.8575.

Example 2 *Consider the following quadratic programming problem with fuzzy coefficients in the quadratic term:*

$$\begin{array}{ll}\min & -5x_1+1.5x_2+(1,2,3)x_1^2+(1,2,3)x_1x_2+(0,1,2)x_2^2\\ \text{s.t.} & x_1+x_2\ge 2\\ & 2x_1-x_2\ge 4\\ & x_1,x_2\ge 0.\end{array} \tag{24}$$

The membership functions are

$$\mu_2(c_1)=\begin{cases} c_1-1 & 1\le c_1\le 2\\ 2-c_1 & 2\le c_1\le 3\end{cases}$$

$$\mu_1(c_2)=\begin{cases} c_2 & 0\le c_2\le 1\\ 2-c_2 & 1\le c_2\le 2\end{cases}$$

From the membership function for these costs an interval representation according to (4) can be given:

$$c_1^{0.2}=[1.2,2.8],\quad c_1^{0.4}=[1.4,2.6],$$
$$c_1^{0.6}=[1.6,2.4],\quad c_1^{0.8}=[1.8,2.2]$$

$$c_2^{0.2}=[0.2,1.8],\quad c_2^{0.4}=[0.4,1.6],$$
$$c_2^{0.6}=[0.6,1.4],\quad c_2^{0.8}=[0.8,1.2]$$

with $c_1^0=[1,3]$, $c_1^1=[2,2]$ and $c_2^0=[0,2]$, $c_2^1=[1,1]$, and now $\mathbb{M}=\{0,0.2,0.4,0.6,0.8,1\}$.

Thus, Problem (24) can be re-formulated as

$$\begin{array}{ll}\min & -5x_1+1.5x_2+(\alpha+1)x_1^2-(\alpha+1)x_1x_2+\alpha x_2^2\\ \min & -5x_1+1.5x_2+(\alpha+1)x_1^2-(\alpha+1)x_1x_2+(2-\alpha)x_2^2\\ \min & -5x_1+1.5x_2+(\alpha+1)x_1^2-(3-\alpha)x_1x_2+\alpha x_2^2\\ \min & -5x_1+1.5x_2+(\alpha+1)x_1^2-(3-\alpha)x_1x_2+(2-\alpha)x_2^2\\ \min & -5x_1+1.5x_2+(3-\alpha)x_1^2-(\alpha+1)x_1x_2+\alpha x_2^2\\ \min & -5x_1+1.5x_2+(3-\alpha)x_1^2-(\alpha+1)x_1x_2+(2-\alpha)x_2^2\\ \min & -5x_1+1.5x_2+(3-\alpha)x_1^2-(3-\alpha)x_1x_2+\alpha x_2^2\\ \min & -5x_1+1.5x_2+(3-\alpha)x_1^2-(3-\alpha)x_1x_2+(2-\alpha)x_2^2\\ \text{s.t.} & x_1+x_2\ge 2\\ & 2x_1-x_2\ge 4\\ & x_1,x_2\ge 0.\end{array} \tag{25}$$

Problem (25) is solved by weighting objectives, with $\alpha=0.8$, where an auxiliary problem like (21) can be obtained;

$$\begin{array}{ll}\min & -5x_1+1.5x_2+\omega 1(1.8x_1^2-1.8x_1x_2+0.8x_2^2)\\ & +\omega_2(1.8x_1^2-1.8x_1x_2+1.2x_2^2)+\omega_3(1.8x_1^2-2.2x_1x_2+0.8x_2^2)\\ & +\omega_4(1.8x_1^2-2.2x_1x_2+1.2x_2^2)+\omega_5(2.2x_1^2-1.8x_1x_2+0.8x_2^2)\\ & +\omega_6(2.2x_1^2-1.8x_1x_2+1.2x_2^2)+\omega_7(2.2x_1^2-2.2x_1x_2+0.8x_2^2)\\ & +\omega_8(2.2x_1^2-2.2x_1x_2+1.2x_2^2)\\ \text{s.t.} & x_1+x_2\ge 2,\ 2x_1-x_2\ge 4,\ x_1,x_2\ge 0.\end{array} \tag{26}$$

Different optimal solutions to each of the weighting vectors is obtained by choosing the canonical base of $\mathbb{R}^8$ as weighting vectors, as it is presented in Table 2.

Table 2: Result of Problem 21 to each vector ω.

ω	x_1	x_2	FunObj
(1,0,0,0,0,0,0,0)	1.5114	0.4886	-3.8506
(0,1,0,0,0,0,0,0)	1.5521	0.4479	-3.7630
(0,0,1,0,0,0,0,0)	1.4688	0.5312	-4.1547
(0,0,0,1,0,0,0,0)	1.5096	0.4904	-4.0505
(0,0,0,0,1,0,0,0)	1.3854	0.6146	-3.0130
(0,0,0,0,0,1,0,0)	1.2705	0.3279	-2.9303
(0,0,0,0,0,0,1,0)	1.3558	0.6442	-3.3582
(0,0,0,0,0,0,0,1)	1.4018	0.5982	-3.2040

On the other hand, the approach described in 2.3 can be obtained by using the first and last weighting vector of Table 2. Hence, by taking $\omega_L=(1,0,0,0,0,0,0,0)$ and solving (26) the optimal solution is $x^*=[1.5114,0.4886]$ and value of objective function is -3.8506. Besides, by taking $\omega_U=(0,0,0,0,0,0,0,1)$ and solving (26) the optimal solution is $x^*=[1.4018,0.5982]$ and value of objective function is -3.2040.

Example 3 *Consider the following quadratic programming problem with all fuzzy coefficients:*

$$\begin{array}{ll}\min & (-6,-5,-4)x_1+(1,1.5,2)x_2+(1,2,3)x_1^2+\\ & +(1,2,3)x_1x_2+(0,1,2)x_2^2\\ \text{s.t.} & x_1+x_2\ge 2\\ & 2x_1-x_2\ge 4\\ & x_1,x_2\ge 0.\end{array} \tag{27}$$

The membership functions are

$$\mu_{-5}(c_1)=\begin{cases} c_1+6 & -6\le c_1\le -5\\ -c_1-4 & -5\le c_1\le -4\end{cases}$$

$$\mu_{1.5}(c_2)=\begin{cases} 2(c_2-1) & 1\le c_2\le 1.5\\ 2(2-c_2) & 1.5\le c_2\le 2\end{cases}$$

$$\mu_2(c_1)=\begin{cases} c_1-1 & 1\le c_1\le 2\\ 2-c_1 & 2\le c_1\le 3\end{cases}$$

$$\mu_1(c_2)=\begin{cases} c_2 & 0\le c_2\le 1\\ 2-c_2 & 1\le c_2\le 2\end{cases}$$

From the membership function for these costs an interval

ISBN: 978-989-95079-6-8

representation according to (3) and (4) can be given:

$$c_1^{0.2} = [-5.8, -4.2], \quad c_1^{0.4} = [-5.6, -4.4],$$
$$c_1^{0.6} = [-5.4, -4.6], \quad c_1^{0.8} = [-5.2, -4.8]$$

$$c_2^{0.2} = [1.1, 1.9], \quad c_2^{0.4} = [1.2, 1.8],$$
$$c_2^{0.6} = [1.3, 1.7], \quad c_2^{0.8} = [1.4, 1.6]$$

$$c_3^{0.2} = [1.2, 2.8], \quad c_3^{0.4} = [1.4, 2.6],$$
$$c_3^{0.6} = [1.6, 2.4], \quad c_3^{0.8} = [1.8, 2.2]$$

$$c_4^{0.2} = [0.2, 1.8], \quad c_4^{0.4} = [0.4, 1.6],$$
$$c_4^{0.6} = [0.6, 1.4], \quad c_4^{0.8} = [0.8, 1.2]$$

with $c_1^0 = [-6, -4]$, $c_1^1 = [-5, -5]$, $c_2^0 = [1, 2]$, $c_2^1 = [1.5, 1.5]$, $c_3^0 = [1, 3]$, $c_1^3 = [2, 2]$ *and* $c_4^0 = [0, 2]$, $c_4^1 = [1, 1]$, *and now* $\mathbb{M} = \{0, 0.2, 0.4, 0.6, 0.8, 1\}$.

Different optimal solutions to each of the weighting vectors can be showed by the multi-objective method described choosing the canonical base of $\mathbb{R}^{32}$ *as weighting vectors. According to the previous examples, it can obtain the minimum and maximum values that form the optimal interval of this problem. The minimum value is -4.5021 where the optimal solution is* $[1.4792, 0.5208]$ *while the maximum value of this optimal interval is -2.6579 where the optimal solution is* $[1.1803, 0.2186]$.

4 Conclusions

This paper presented three approaches for solving fuzzy quadratic programming problems. A novel approach that transforms a quadratic programming problem with imprecise costs in the objective function into a parametric multiobjective quadratic programming problem. The other two approaches are methods published that are used in this work for comparison. Quadratic Programming is very important in theoretical and practical areas. When real-world applications are considered, vagueness appears in a natural way, and hence it makes perfect sense to think of fuzzy quadratic programming problems. In contrast to what happens in fuzzy linear programming problems, unfortunately until now not much research has been conducted to this important class of problems.

It was shown that the parametric multiobjective programming problem can obtain the optimal solutions of the other approaches because these solutions are a quadratic or linear combination of the optimal solutions obtained for each of weighting vectors. Hence, this shows that the parametric multiobjective approach contains the optimal solutions obtained by the others. Therefore, parametric approaches can be used as a general method to solve quadratic programming problems with uncertain costs in the objective function.

The authors aim to extend the line of investigation involving Fuzzy Quadratic Programming problems in order to try to solve practical real-life problems by facilitating the building of Decision Support Systems.

Acknowledgment

The authors want to thank the support provided by the Brazilian agency CAPES and the Spanish projects TIN2005-08404-C04-01 (70% of which are FEDER funds) and MINAS (TIC-00129-JA) for the support provided.

References

[1] E. Ammar and H. A. Khalifa. Fuzzy portfolio optimization a quadratic programming approach. *Chaos, Solitons and Fractals*, 18:1045–1054, 2003.

[2] S.S. Appadoo, S.K. Bhatt, and C.R. Bector. Application of possibility theory to investment decisions. *Fuzzy Optimization and Decision Making,*, 7:35–57, 2008.

[3] E. M. L. Beale. On quadratic programming. *Naval Research Logistics Quarterly*, 6:227–244, 1959.

[4] R. E. Bellman and L. A. Zadeh. Decision-marking in a fuzzy environment. *Management Science*, 17(4):B141–B164, 1970.

[5] G.R. Bitran. Linear multiple objective problems with interval coefficients. *Management Sciences*, 26(7):694–706, 1980.

[6] M. Delgado, J.L. Verdegay, and M.A. Vila. Imprecise costs in mathematical programming problems. *Control and Cybernetics*, 16(2):113–121, 1987.

[7] M. Delgado, J.L. Verdegay, and M.A. Vila. Relating different approaches to solve linear programming problems with imprecise costs. *Fuzzy Sets and Systems*, 37:33–42, 1990.

[8] C. A. Floudas, P.M. Pardalos, C. Adjiman, W. R. Esposito, Z. H. Gümüs, S. T. Harding, J. L. Klepeis, C. A. Meyer, and C. A. Sshweiger. *Handbook of test problems in local and global optimization*, volume 33 of *Nonconvex Optimization and its Applications*. Kluwer Academic Publishers, Dordrecht, 1999.

[9] W. Hock and K. Schittkowski. *Test examples for nonlinear programming codes*, volume 187 of *Lecture Notes in Economics and Mathematical Systems*. Spring-Verlag, 1981.

[10] Y. J. Lai and C. L. Hwang. *Fuzzy mathematical programming : methods and applications*, volume 394 of *Lecture notes in economics and mathematical systems*. Springer, Berlin, 1992.

[11] S. T. Liu. Quadratic programming with fuzzy parameters: A membership function approach. Chaos, Solitons and Fractals. Article in press. doi:10.1016/j.chaos.2007.07.054, 2007.

[12] S. T. Liu. Solving quadratic programming with fuzzy parameters based on extension principle. In *FUZZ-IEEE 2007*, pages 1–5, London, UK, 2007. IEEE International Conference Fuzzy Systems.

[13] C. V. Negoita and D. A. Ralescu. *Applications of fuzzy sets to systems analysis*. Birkhauser Verlag, Stuttgard, 1975.

[14] D. Poole, A. Mackworth, and R. Goebel. *Computational intelligence: A logical approach*. Oxford University Press, 1998.

[15] S. J. Russel and P. Norvig. *Artificial intelligence: A modern approach*. Prentice Hall, Upper Saddle River, NJ, second edition, 2003.

[16] K. Schittkowski. *More test examples for nonlinear programming codes*, volume 282 of *Lecture Notes in Economics and Mathematical Systems*. Spring-Verlag, 1987.

[17] J. L. Verdegay, R. R. Yager, and P. P. Bonissone. On heuristics as a fundamental constituent of soft computing. *Fuzzy Sets and Systems*, 159(7):846–855, 2008.

[18] S. Wang and S. Zhu. On fuzzy portfolio selection problems. *Fuzzy Optimization and Decision Making,*, 1:361–377, 2002.

[19] P. Wolfe. The simplex method for quadratic programming. *Econometrica*, 27:382–398, 1959.

[20] L. A. Zadeh. A new direction in AI: Toward a computational theory of perceptions. *AI Magazine*, 22(1):73–84, 2001.

ISBN: 978-989-95079-6-8

Comparison of Multicriteria Methods for Land-use Suitability Assessment

Jozo Dujmović[1] Guy De Tré[2] Suzana Dragićević[3]

1.Department of Computer Science, San Francisco State University,
1600 Holloway Ave, San Francisco, CA 94132, U.S.A.
2.Department of Telecommunications and Information Processing Ghent University,
St.-Pietersnieuwstraat 41, B-9000 Ghent, Belgium
3.Department of Geography, Simon Fraser University,
8888 University Drive, Burnaby, BC V5A 1S6, Canada
Email: jozo@sfsu.edu, Guy.DeTre@UGent.be, suzanad@sfu.ca

Abstract— *In this paper we investigate properties of multicriteria methods that are used for building land-use suitability assessment criteria. We identify and describe fundamental properties that are of interest in the land-use suitability analysis and the design of suitability maps. The existing multicriteria methods can be evaluated from the standpoint of their ability to support the desirable properties that affect the expressive power of evaluation methods. In this paper we investigate simple additive scoring, MAVT, MAUT, AHP, OWA, outranking methods and LSP.*

Keywords— GIS, land-use, MCDM.

1 Introduction

This paper has two main goals. The first goal is to identify properties that affect the expressive power of multicriteria decision methods (MCDM) that can be used to create suitability maps of a specific geographic region. The second goal is to investigate land-use suitability assessment methods that are based on simple additive scoring (SAS), multiattribute value technique (MAVT), multiattribute utility technique (MAUT), analytic hierarchy process (AHP), ordered weighted average (OWA), outranking methods (ELECTRE, PROMETHEE), and logic scoring of preference (LSP) in view of satisfaction of the identified properties.

We assume that the analyzed geographic region is divided into cells and that each cell with coordinates x, y is characterized by an array of suitability attribute values $a_1(x,y), \ldots, a_n(x,y)$, $n \geq 1$ [5]. The attributes are defined as quantitative parameters that affect the suitability of a cell for some specific land-use (e.g. housing, recreation, agriculture, industrial development, etc.). The set of attributes must be complete, i.e., it must include *all* relevant components. Generally, the attributes must be justifiable and not redundant with each other.

The purpose of MCDM is to provide a criterion function $\sigma : \mathbb{R}^n \rightarrow [0,1]$ for computing an overall degree of suitability $S(x,y) = \sigma(a_1(x,y), \ldots, a_n(x,y))$ that reflects the suitability of location x, y for specific land-use. The overall suitability is a matter of degree: $0 \leq S(x,y) \leq 1$. As in all soft computing models, here 0 denotes a completely unsuitable location, and 1 denotes the highest level of suitability. A suitability map is defined as a distribution of the overall suitability $S(x,y)$ for a specific geographic region [7].

Nowadays suitability maps are assumed to be dynamically created using data from GIS databases in a way illustrated in Fig. 1. The multicriteria decision model must be interfaced both with the user and with the GIS database. An attribute ETL interface is necessary to Extract, Transform, and Load the set of cell attribute values from the GIS database. The multicriteria decision model is used to implement the criterion function σ and to generate the overall suitability $S(x,y)$. The user input includes a specification of desired suitability attributes and parameters of the decision model. A suitability criterion interface is necessary for accepting user input and for rendering the resulting suitability map.

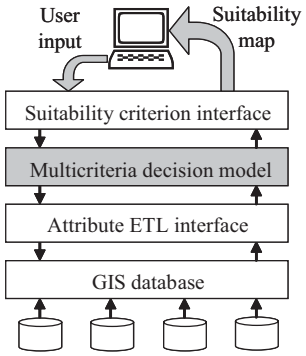


Figure 1: Dynamic generation of suitability maps using GIS.

Suitability maps introduced in GIS literature use a variety of decision models [16, 9, 11, 5, 20]. The emphasis of such efforts is primarily (and naturally) on the selection of attributes and the use of suitability maps. So far the GIS literature avoided problems of evaluating the credibility of MCDM used for the development of suitability maps.

Except for interfacing with GIS components, the land-use suitability assessment problems do not differ from evaluation problems in other areas. All multicriteria decision models are ultimately models of human decision making in the area of evaluation of complex alternatives. Therefore, the credibility of MCDM used in GIS context depends on their ability to express observable properties of human evaluation logic [4]. The primary purpose of this paper is to evaluate the expressive power of GIS-related MCDM based on the level of their ability to support the concepts of human evaluation logic.

The remainder of this paper is structured as follows. In Section 2, the logic properties of GIS-related MCDM are studied. In Section 3, the logic properties of the selected decision meth-

ISBN: 978-989-95079-6-8

ods are investigated. The major contributions of the paper are summarized and some conclusions are presented in Section 4.

2 Ten fundamental properties of MCDM

The properties of human evaluation logic are easily observable, and they are also observable in the context of land-use evaluation. Our goal is to first identify properties that are frequently encountered in all worth assessment problems. The most important properties include the following:

1. Ability to combine any number of attributes.
2. Ability to combine objective and subjective inputs.
3. Ability to combine absolute and relative criteria.
4. Flexible adjustment of relative importance of attributes.
5. Modelling of simultaneity requirements.
   (a) Modelling of soft simultaneity.
   (b) Modelling of hard simultaneity.
6. Modelling of replaceability requirements.
   (a) Modelling of soft replaceability.
   (b) Modelling of hard replaceability.
7. Modelling of balanced simultaneity/replaceability.
8. Modelling of mandatory, desired, and optional requirements.
9. Modelling of sufficient, desired, and optional requirements.
10. Ability to express suitability as an aggregate of usefulness and inexpensiveness (separation of the usefulness analysis and the cost analysis).

2.1 Ability to combine any number of attributes

We assume that the analyzed geographic region is divided into cells and that each cell is characterized by attributes $a_1, \ldots, a_n$, $n \geq 1$. In special cases $n = 1$. For example, the acoustic pollution (noise) maps presented in [11] uses in each cell a single scalar value, the level of noise. However, in all cases of more complex suitability maps we have $n > 1$. For example, the housing suitability maps discussed in [5] use $n = 11$. If we want to develop maps of complex suitability indicators that depend on variety of inputs it is reasonable to expect that the number of inputs can be large. So, in a general case, MCDM used to compute the land-use suitability must be able to support any number of inputs.

It should be noted that this requirement indirectly transforms into a requirement to use nonlinear MCDM models. Indeed, in the case of linear weighted aggregation models, the overall suitability $S(x, y)$ is a dot product of a vector of n suitability components $\mathbf{s} = [s_1(x, y), \ldots, s_n(x, y)]$ (one component for each attribute) and a corresponding vector of relative weights $\mathbf{W} = [W_1, \ldots, W_n]$: $S(x, y) = \sum_{i=1}^{n} W_i s_i(x, y)$ where $0 \leq s_i(x, y) \leq 1$, $0 < W_i < 1$, $i = 1, \ldots, n$, and $W_1 + \cdots + W_n = 1$.

Obviously, a large value of n yields a low significance of individual components. For example, if $n = 100$ the average significance (the ability to change the overall suitability S) of each input is only 1%, and in the case of unequal weights some inputs will have the significance considerably below 1% which means that such inputs may (and should) be neglected. This is not acceptable since credible decision models must be equally applicable for any number of inputs.

2.2 Ability to combine objective and subjective inputs

Some input attributes are objectively measurable values (e.g., distances, slopes, altitudes, temperatures, etc). Other inputs are subjective and must be assessed by experts (e.g., aesthetic quality of an area, the quality of educational and/or medical institutions, quality of public transportation, development costs, etc.). MCDM for GIS applications must be able to combine and aggregate objective and subjective inputs.

2.3 Ability to combine absolute and relative criteria

Each suitability attribute must satisfy specific requirements. The requirements can be defined as elementary criterion functions $g_i : \mathbb{R} \rightarrow [0, 1]$, $i = 1, 2, \ldots, n$. By definition, the attribute suitability $s_i(x, y) = g_i(a_i(x, y))$ is the degree of satisfaction of the attribute requirements. The suitability of an attribute is a component of the overall suitability of the evaluated location x, y.

Suppose that a decision maker (DM) needs a suitability map for housing in a rural area where it is possible to buy land and build a house. Let the attribute $a_i(x, y) = t$ denote the traveling time between location $L(x, y)$ and the closest elementary school (or hospital, or airport, or any other point of interest). An absolute criterion is a criterion that evaluates location L regardless of other competitive locations. E.g., a DM may specify the absolute requirements as threshold times t_{min} and t_{max} so that all times $t \geq t_{max}$ are considered unacceptable and all times $t \leq t_{min}$ are considered perfectly acceptable. Then the elementary criterion function might be defined as follows: $g_i(t) = \max(0, \min(1, (t_{max} - t)/(t_{max} - t_{min})))$. So, $g_i(t)$ specifies the attribute suitability for any value of t. If DM wants to compare locations L_1 and L_2, with respect to the access to school then the comparison can be based on $g_i(t_1)$ and $g_i(t_2)$. If t_{min} and t_{max} are well justifiable values, then $g_i(t_1)$ and $g_i(t_2)$ are very credible results.

Another approach is to define a relative criterion. In the case of k locations an example of such a criterion could be $g_i(t) = t_{min}/t$, $t_{min} = \min(t_1, \ldots, t_k)$. The relative criterion evaluates relationships between competitors regardless of DM's actual needs: the closest location is considered perfectly suitable ($g_i(t_{min}) = 1$) and a location where $t = 2t_{min}$ gets $g_i(t) = 0.5$. Obviously, the credibility of relative criteria is much less under DM's control than the credibility of corresponding absolute criteria. E.g., if only two locations are available, then the location that is 2 minutes from the closest school is likely to be equally attractive as the location that is 1 minute from the school, and locations that are 1 hour and 2 hours from the closest school might be equally unacceptable; in both cases the suitability levels of 1 and 0.5 are inappropriate. For some other values, e.g., $t_{min} = 7$ minutes, the presented relative criterion might be appropriate.

It is highly unlikely that DM knows what is better, but does not know what is good. In other words, if DM can specify a justifiable relative criterion, it is very likely that DM can also specify a justifiable absolute criterion. Absolute criteria can be used to evaluate a single alternative (e.g., a single location), while relative criteria can be applied only if we have multiple alternatives. Relative criteria are appropriate only in situations where expected values of attributes are unknown. Such situations are not frequent, but exist, and MCDM must be able to combine and aggregate absolute and relative attribute criteria.

ISBN: 978-989-95079-6-8

2.4 Flexible adjustment of the relative importance of attributes

In a general case, multiple attributes are not equally important. The relative importance of an attribute is usually expressed using multiplicative or implicative weights [3]. The relative importance has two roles in suitability criteria: it defines the level of contribution of an attribute to the overall suitability, and it defines compensatory properties between attributes. Assuming that the overall suitability is a function of attribute suitability degrees, $S = \lambda(s_1, \ldots, s_j, \ldots, s_k, \ldots, s_n)$, we expect that if the suitability s_j is more important than the suitability s_k, then $\partial S/\partial s_j \geq \partial S/\partial s_k$. Under the same assumptions, if some compensation is possible, then $\lambda(s_1, \ldots, s_j, \ldots, s_k, \ldots, s_n) = \lambda(s_1, \ldots, s_j - p, \ldots, s_k + q, \ldots, s_n)$ and $p < q$. In other words, a suitability decrement p of an attribute can be compensated by the suitability increment q of a less important attribute, but the increment q must be greater than the decrement p. These properties are essential in human reasoning and must be supported by MCDM.

2.5 Modelling of simultaneity requirements

Function $\lambda : [0,1]^n \to [0,1]$ that aggregates all attribute suitability degrees and computes the overall suitability degree is essentially a logic function. In human decision making the aggregation of suitability degrees is usually a stepwise process where small groups of related suitability degrees are aggregated and replaced by an aggregated suitability degree. The process of stepwise aggregation terminates when the suitability degrees of subsystems at the highest level are aggregated yielding the overall suitability degree S as a compound function of attribute suitability degrees.

Let us investigate an aggregation step where DM aggregates m suitability degrees using an aggregation function $\mu : [0,1]^m \to [0,1]$. In human decision making $\mu(s_1, \ldots, s_m)$ is most frequently a model of simultaneity. E.g., a homebuyer regularly prefers locations that are both close to schools for children *and* to jobs for parents. Such a simultaneity requirement can be modelled using some form of partial conjunction [3], i.e., μ is expected to have an adjustable degree of similarity with conjunction $s_1 \wedge \cdots \wedge s_m$. A normalized degree of proximity between $\mu(s_1, \ldots, s_m)$ and $s_1 \wedge \cdots \wedge s_m$ is called andness and it can be defined by: $\alpha = \frac{(s_1 \vee \cdots \vee s_m) - \mu(s_1, \ldots, s_m)}{(s_1 \vee \cdots \vee s_m) - (s_1 \wedge \cdots \wedge s_m)}$. Since $(s_1 \wedge \cdots \wedge s_m) \leq \mu(s_1, \ldots, s_m) \leq (s_1 \vee \cdots \vee s_m)$ it follows that $0 \leq \alpha \leq 1$ and $\mu(s_1, \ldots, s_m)$ is a model of simultaneity if $0.5 < \alpha \leq 1$ and $(s_1 \wedge \cdots \wedge s_m) \leq \mu(s_1, \ldots, s_m) < (s_1 + \cdots + s_m)/m$. This form of andness depends on input degrees of suitability and humans usually think globally using an average andness, e.g., $\overline{\alpha} = \int_{[0,1]^m} \alpha(s_1, \ldots, s_m) ds_1 \ldots ds_m$, $0 \leq \overline{\alpha} \leq 1$. If DM needs a high simultaneity (e.g., $\overline{\alpha} > 0.7$), that usually means that all inputs must be at least partially satisfied. In other words, $\mu(s_1, \ldots, s_m) = 0$, $s_i = 0$, $i \in \{1, \ldots, m\}$ and it is mandatory to satisfy all inputs. This kind of simultaneity is called a *hard simultaneity*. In other cases DM may need a *soft simultaneity*, where the average andness is slightly above 0.5 and $\mu(s_1, \ldots, s_m) > 0$, $s_i > 0$, $i \in \{1, \ldots, m\}$. Supporting models of simultaneity is an important requirement that MCDM must satisfy.

2.6 Modelling of replaceability requirements

Replaceability requirements are symmetrical and complementary to simultaneity requirements. Replaceability means that a high suitability in a group of attributes can be achieved using any one of the attributes (i.e., they can replace each other). E.g., a home location can be considered suitable for recreational activities if it is close to lake, *or* close to ski terrains. The intensity of replaceability can be determined using the orness indicator that is a complement of andness: $\overline{\omega} = 1 - \overline{\alpha}$. Replaceability aggregators satisfy conditions $0.5 < \overline{\omega} \leq 1$ and $(s_1 + \cdots + s_m)/m < \mu(s_1, \ldots, s_m) \leq (s_1 \vee \cdots \vee s_m)$. High orness means low andness and vice versa. Similarly to the case of simultaneity, a high level of replaceability (e.g., $\overline{\omega} > 0.7$) may be combined with the requirement for hard replaceability where $\mu(s_1, \ldots, s_m) = 1$, $s_i = 1$, $i \in \{1, \ldots, m\}$. Soft replaceability is any form of replaceability that does not satisfy the hard replaceability requirements. In the case of soft replaceability we have $\mu(s_1, \ldots, s_m) < 1$, $s_i < 1$, $i \in \{1, \ldots, m\}$.

2.7 Modelling of balanced simultaneity/replaceability

If the simultaneity and replaceability are balanced then $\alpha = \omega = 0.5$. In the case of two variables, from definition $\alpha = \frac{(s_1 \vee s_2) - \mu(s_1, s_2)}{(s_1 \vee s_2) - (s_1 \wedge s_2)} = 0.5$ it follows that $\mu(s_1, s_2) = ((s_1 \vee s_2) + (s_1 \wedge s_2))/2 = (s_1 + s_2)/2$. This result indicates that the arithmetic mean is a soft computing logic function that combines simultaneity and replaceability requirements in a balanced way: DM would like that all attributes are simultaneously satisfied, but at the same time s/he accepts that any attribute can compensate any other attribute. It is important to understand that the arithmetic mean represents a model of this specific logic condition and nothing more. MCDM that use the arithmetic mean are acceptable only in cases where DM can justify the use of this specific logic condition.

According to Malczewski [16], "GIS implementations of the weighted summation procedures are often used without full understanding of the assumptions underlying this approach."

2.8 Modelling of mandatory, desired, and optional requirements

Using models of simultaneity, replaceability and negation ($x \mapsto 1 - x$) it is possible to create a variety of compound soft computing logic functions that precisely reflect the needs of DM. A compound aggregator that is most frequent in human reasoning is used to combine mandatory and nonmandatory attributes. Most frequently there are one mandatory attribute and one or two nonmandatory attributes. E.g., a DM may reject home locations that do not have good ground transportation, but accept locations that are far from an international airport. In such cases the ground transportation is a mandatory requirement, and the vicinity of an international airport is desired, but not mandatory. Optional attributes are also nonmandatory and have lower significance than desired attributes. To model such requirements we need aggregators $\mu(s_{man}, s_{des}, s_{op})$ that satisfy the conditions $\mu(s_{man}, s_{des}, s_{op}) = 0$, $s_{man} = 0$, $s_{des} > 0$, $s_{op} > 0$ and if the compensation between s_{des} and s_{op} is possible, then $\mu(s_{man}, s_{des}, s_{op}) = \mu(s_{man}, s_{des} - p, s_{op} + q)$, $s_{man} > 0$, $p < q$. Optional attributes can be omitted, and in such cases

ISBN: 978-989-95079-6-8

we use only mandatory and desired inputs. An aggregator with these properties is a partial absorption function introduced in [2] and expanded in [19].

2.9 Modelling of sufficient, desired, and optional requirements

Sufficient, desired and optional requirements are a disjunctive counterpart of mandatory, desired and optional requirements. If the sufficient input is completely satisfied then the desired and optional inputs have no effect. If the sufficient attribute has low or even zero suitability degree, this can be partially compensated by the desired and optional attributes. The corresponding aggregator $\mu(s_{suf}, s_{des}, s_{op})$ satisfies conditions $\mu(s_{suf}, s_{des}, s_{op}) = 1, s_{suf} = 1, s_{des} < 1, s_{op} < 1$, and if the compensation between s_{des} and s_{op} is possible, then $\mu(s_{suf}, s_{des}, s_{op}) = \mu(s_{suf}, s_{des} - p, s_{op} + q)$, $s_{suf} < 1$, $p < q$.

2.10 Ability to express suitability as an aggregate of usefulness and inexpensiveness

Land-use is regularly related to a variety of costs (e.g., the cost of land, the cost of building infrastructure and objects, the cost of financing, etc.). The overall suitability depends on two simultaneous requirements: finding locations that are very useful for specific purpose and at the same time inexpensive. A convenient way to solve that problem is to define *usefulness* as a non-financial part of the overall suitability, and *inexpensiveness* as an overall result of cost analysis (an aggregate of cost components only). Then, the overall suitability can be conveniently expressed as an aggregate of usefulness and inexpensiveness. Separation of cost and usefulness attributes reflects human reasoning where the overall cost is compared with the corresponding overall satisfaction. Of course, there are cases where the overall suitability does not depend on cost. In such cases the overall suitability reduces to usefulness.

The presented list of properties is not proved to be necessary and sufficient in all cases. However, the presented conditions are relevant for many land use decision problems and show that logic aggregation of attribute suitabilities is a frequently needed property. For detailed analysis of mathematical conditions see [4], and for sample applications see [4, 5].

The necessary properties of MCDM's do not change if attributes of a cell are functions of time, or functions of the values of attributes in other cells.

3 Properties of GIS related MCDM

Several GIS related MCDM approaches have been presented in literature [15, 16]. Among the decision methods used in these approaches we selected a representative set that consists of the following techniques: *simple additive scoring* (SAS) [8, 6], *multiattribute value technique* (MAVT) [22, 13, 14], *multiattribute utility technique* (MAUT) [13, 10], *analytic hierarchy process* (AHP) [18, 1], *ordered weighted average* (OWA) [21, 17], *outranking methods* [12] and *logic scoring of preference* (LSP) [4, 5].

The ten fundamental features presented in the previous section can be used to investigate the selected decision methods in view of their appropriateness for land-use evaluation and suitability map construction.

3.1 Simple additive scoring

The SAS technique [8, 6] is based on the concept of a weighted average in which weights are used to denote relative importance of suitability attributes. A DM directly assigns a weight w_i to each suitability attribute a_i, $i = 1, \ldots, n$. These assigned weights are rescaled to normalized weights W_i, $i = 1, \ldots, n$, such that $\sum_{i=1}^{n} W_i = 1$. The overall score or overall degree of suitability $S(x, y)$ of each cell x, y is then computed by:

- Determining the n suitability components $s_1(x, y), \ldots, s_n(x, y)$ that are obtained from the evaluation of the n attributes for the cell.
- Multiplying each suitability component $s_i(x, y)$ with the normalized weight W_i of its corresponding attribute.
- Summing the products over all attributes, i.e., $S(x, y) = \sum_{i=1}^{n} W_i s_i(x, y)$.

Therefore, SAS uses a simple linear weighted aggregation model.

3.2 Multiattribute value technique

In MAVT [22, 13, 14] suitability attributes a_i, $i = 1, \ldots, n$ are evaluated using value functions that aim to mathematically represent human judgements. A single-attribute value function translates the performance of the alternative attribute values into a value score which represents the degree to which a decision objective is achieved. As such, value function v_i associates a number (or 'value') $v_i(a)$ with each alternative value a of attribute a_i in such a way that a preference order on the alternatives consistent with DM value judgements is obtained.

For aggregation, more complex value functions are used. The most commonly used function is the simple additive weighting function $S(x, y) = \sum_{i=1}^{n} W_i v_i(a_i(x, y))$ where W_i is the weight of suitability attribute a_i and $v_i(a_i(x, y))$ is the value of suitability attribute value $a_i(x, y)$ (of cell x, y). This approach is valid if suitability attributes are preferentially independent.

The weights W_i are scaling constants that have to be derived with reference to the attribute ranges. These need to be elicited through questions which capture acceptability of trade-offs (e.g., 'how many units of one suitability attribute are worth how many of units of another suitability attribute?'). Weights must sum up to 1, i.e., $\sum_{i=1}^{n} W_i = 1$.

As such, the MAVT approach is similar to the 'scoring method', except that the scores $s_i(x, y)$ are replaced by values $v_i(a_i(x, y))$ that are obtained with the value function v_i.

3.3 Multiattribute utility technique

MAUT [13, 10] is used and treated separately from MAVT when 'risks' or 'uncertainties' have a significant role in the definition and assessment of alternatives. The attitude of the DM toward risk is incorporated into the assessment of a single-attribute utility function u_i, which is obtained through utility analysis and translates the values of suitability attribute a_i into 'utility units'. A 'utility unit' is a relative value between 0 and 1 (where 0 and 1 resp. denote the worst and best values). The concept of a utility function is inherently probabilistic in nature.

ISBN: 978-989-95079-6-8

Aggregation is done by an overall utility function. A simple additive weighting function is most commonly used. In such a case $S(x,y) = \sum_{i=1}^{n} W_i u_i(a_i(x,y))$. Suitability attributes must be preferentially independent. The weights W_i, $i = 1, \ldots, n$ have to sum up to 1, i.e., $\sum_{i=1}^{n} W_i = 1$.

From aggregation point of view, the MAUT approach is similar to the MAVT and 'scoring' approaches.

3.4 Analytic hierarchy process

AHP [18, 1] uses a different approach. Cognitive psychology has found that people are poor at assimilating large quantities of information on problems. The subsequent steps in AHP can be summarized as follows:

- Model the problem as a hierarchy containing the decision goal, the alternatives for reaching it, and the suitability attributes for evaluating the alternatives.
- Establish priorities (normalized weights) among the elements of the hierarchy by making a series of judgments based on pairwise comparisons of the elements.
- Synthesize these judgments to yield a set of overall weights for the hierarchy. This is done by means of a sequence of multiplications of the matrices of relative weights at each level of the hierarchy.
- Check the consistency of the judgments.
- Come to a final decision based on the results of this process.

Suitability attributes are subdivided in subattributes and hierarchically structured. Overall suitability $s(x,y)$ for an attribute a is computed by an additive weighting function $s(x,y) = \sum_{j=1}^{m} W_{s_j(x,y)} s_j(x,y)$ where $s_j(x,y)$, $j = 1, \ldots, m$ denote the suitability components of the subattributes of a. All weights $W_{s_j(x,y)}$ must to sum up to 1, i.e., $\sum_{i=j}^{n} W_{s_j(x,y)} = 1$.

3.5 Ordered weighted averaging

In an OWA approach [17], the DM specifies the decision-relevant suitability attributes to be used as evaluation criteria; identifies preferred criteria values on a qualitative scale; and defines the relative importance of each criterion by assigning weights. The weighted criterion values are then combined using an OWA aggregation operator [21], resulting in an evaluation score for each cell. OWA allows the DM to specify a decision strategy that reflects his decision-related preferences.

The OWA operators [21], provide a parameterized class of mean type aggregation operators. Many notable mean operators such as the max, arithmetic average, median and min, are members of this class. OWA operators allow to model linguistically expressed aggregation instructions.

An OWA operator of dimension n is a mapping function $F : [0,1]^n \rightarrow [0,1]$ that has an associated collection of weights $\mathbf{W} = [W_1, \ldots, W_n] \in [0,1]^n$, for which it holds that $\sum_{i=1}^{n} W_i = 1$, and with $F(s_1, \ldots, s_n) = \sum_{i=1}^{n} W_i s'_i$ where s'_i is the ith largest value of the s_i.

By choosing different $\mathbf{W}$, different aggregation operators can be implemented. The OWA operator is a non-linear operator as a result of the process of determining the values s'_i.

3.6 Outranking methods

Outranking methods, such as variants of ELECTRE and PROMETHEE are used in the areas of GIS and environmental planning [12]. The basic idea of these methods is to use strictly relative criteria that express a range from indifference to strong preference of one alternative over another alternative separately for all individual attributes. The attribute preferences are then averaged using the arithmetic mean to generate the credibility of the outranking relation of two alternatives (PROMETHEE). In such a case no input is mandatory. The overall degree of outranking of two alternatives can also be computed using a product (ELECTRE III), making all attributes mandatory.

3.7 Logic scoring of preference

With the LSP method [4, 5] suitability maps are created with the following main steps:

1. Creation of an attribute tree. This tree contains and structures all parameters that affect the overall suitability and is build by the DM.
2. Definition of elementary criteria. The DM has to provide an elementary criterion for each attribute involved in the decision process. These criteria will be evaluated during suitability map construction. For each analyzed cell x, y, the evaluation of each attribute a_i, $i = 1, \ldots, n$ will result in an elementary satisfaction degree $s_i(x,y)$.
3. Creation of the aggregation structure. For each analyzed cell, all associated elementary satisfaction degrees must be aggregated. Therefore, the DM has to create an aggregation structure, which adequately reflects his domain knowledge and reasoning.
4. Computation of the overall suitability degree. Once the attribute tree, the elementary criteria and the aggregation structure are available, the suitability map construction can start. The elementary criteria are evaluated and their resulting elementary satisfaction degrees can be aggregated in order to compute the overall satisfaction degree of each analyzed cell.

Aggregation in LSP is done via the aggregation structure, build by the DM. The basic building blocks of the aggregation structure are the simple and compound LSP aggregators. The DM can use them to construct an easily understandable aggregation schema which is consistent with observable properties of human reasoning in the area of evaluation.

The simple LSP aggregators are all graded preference logic functions and based on a superposition of the fundamental *Generalized Conjunction/Disjunction* (GCD) function. Most frequently, GCD is implemented by the weighted power mean function $GCD(s_1, \ldots, s_n) = (W_1 s_1^r + \cdots + W_n s_n^r)^{1/r}$ where $r \in [-\infty, +\infty]$ and $0 < W_i < 1$ such that $\sum_{i=1}^{n} W_i = 1$ [5]. The parameter r determines the logical behavior of the function. As such a continuous variety of logical functions ranging from full conjunction to full disjunction can be modelled. The simple LSP aggregators can be used to construct more complex, compound operators like the conjunctive partial absorption which can be used to aggregate a mandatory and a desired criterion.

For suitability map construction, the overall suitability degree of each analyzed cell x, y must be computed. This is

ISBN: 978-989-95079-6-8

done in two steps: First, the elementary satisfaction degrees $s_1(x,y),\ldots,s_n(x,y)$ are aggregated using the logic aggregation structure. This results in the overall satisfaction degree $s(x,y)$ of the cell. Second, cost (if applicable) is taken into account. Cost is dealt with separately. This better reflects human reasoning and allows for more efficient cost/satisfaction studies. Cost is considered to be a function c of the analyzed cells. For each cell x,y the cost function returns the associated cost $c(x,y)$ of the cell. If the importance of high satisfaction of criteria is the same as the importance of low cost, then the overall suitability degree $S(x,y)$ of the cell can be computed by $S(x,y)=s(x,y)/c(x,y)$. Alternative definitions of $S(x,y)$ are possible.

4 Conclusions

All multicriteria decision methods should be models of human decision making and the only way to prove their credibility is to show their proximity to human evaluation logic. Our analysis shows that the majority of existing MCDM are not derived with an explicit goal to model observable properties of human reasoning. Indeed, human reasoning is neither restricted to the use of arithmetic mean nor restricted to relative criteria only. Humans use a spectrum of absolute and relative elementary criteria and a spectrum of soft computing logic aggregators, such as soft and hard simultaneity and replaceability with nonmandatory, mandatory and sufficient attributes, conjunctive and disjunctive partial absorption, as well as other compound aggregators.

Decision methods that are used in land-use evaluation problems cannot be randomly selected, without appropriate justification. The justification for using a specific evaluation method in a GIS environment should be based on investigating the capability of the method to support features that are proved to characterize human decision making. Many oversimplifications that are frequent in GIS literature [16] (particularly those based on simple additive models) are based on the fact that they are "easy-to-understand and intuitively appealing." Unfortunately, that is not enough. Before using mathematical models it is first necessary to prove that they are appropriate.

LSP is a method developed with the goal to support logic operators observed in human reasoning. Consequently, it is realistic to expect that the LSP method can provide highly accurate and justifiable models for GIS applications, such as land-use evaluation, suitability maps, and natural resources planning.

References

[1] R. Banai. Fuzziness in geographic information systems: contributions from the analytic hierarchy process. *International Journal of Geographical Information Systems*, Vol. 7, No. 4, pp. 315–329, 1993.

[2] J. Dujmović. Partial Absorption Function. *Journal of the University of Belgrade, EE Dept., Series Mathematics and Physics*, No. 659, pp. 156–163, 1979.

[3] J. Dujmović and H.L. Larsen. Generalized Conjunction / Disjunction. *International Journal of Approximate Reasoning*, Vol. 46, pp. 423–446, 2007.

[4] J. Dujmović. Preference Logic for System Evaluation. *IEEE Transactions on Fuzzy Systems*, Vol. 15, No. 6, pp. 1082–1099, 2007.

[5] J. Dujmović, G. De Tré and N. Van de Weghe. Suitability Maps Based on the LSP Method. In *Proc. of the 5th MDAI conference (Modeling Decisions for Artificial Intelligence)*, Sabadell, Spain, 2008. Lecture Notes in Artificial Intelligence, vol. 5285, pp. 15–25, Springer-Verlag, Heidelberg, Germany, 2008.

[6] I. Heywood, J. Oliver and S. Tomlinson. Building an exploratory multi-criteria modelling environment for spatial decision support. In P. Fisher (Ed.) *Innovations in GIS 2*, pp. 127–136, Taylor & Francis, London, UK, 1995.

[7] L.D. Hopkins. Methods for Generating Land Suitability Maps: A Comparative Evaluation. *JAPA*, Vol. 43, No. 4, 1977.

[8] P. Jankowski and L. Richard. Integration of GIS-based suitability analysis and multicriteria evaluation in a spatial decision support system for route selection. *Environment and Planning B*, Vol. 21, No. 3, pp. 326–339, 1994.

[9] P. Jankowski, N. Andrienko and G. Andrienko. Map-centered exploratory approach to multiple criteria spatial decision making. *Int. J. Geographical Information Science*, Vol. 15, No. 2, pp. 101–127, 2001.

[10] R. Janssen and P. Rietveld. Multicriteria analysis and geographical information systems: an application to agricultural land use in the Netherlands. In H.J. Scholten and J.C.H. Stillwell (Eds.) *Geographical information systems for urban and regional planning*, pp. 129–139, Kluwer Academic Publishers, Dordrecht, The Netherlands.

[11] F. Joerin, M. Thériault and A. Musy. Using GIS and outranking multicriteria analysis for land-use suitability assessment. *Int. J. Geographical Information Science*, Vol. 15, No. 2, pp. 153–174, 2001.

[12] A. Kangas, J. Kangas and J. Pykäkäinen. Outranking Methods As Tools in Strategic Natural Resources Planning. *Silva Fennica*, Vol. 35, No. 2, pp. 215–227, 2001.

[13] R.L. Keeney and H. Raiffa. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs (2nd Edition)*. Cambridge University Press, Cambridge, 1993.

[14] R.L. Keeney. Value-focused thinking: identifying decision opportunities and creating alternatives. *European Journal of Operational Research*, Vol. 92, pp. 537–549, 1996.

[15] J. Malczewski. *GIS and multicriteria decision analysis*. John Wiley & Sons Inc., New York, USA, 1999.

[16] J. Malczewski. GIS-based multicriteria analysis: a survey of the literature. *International Journal of Geographic Information Science*, Vol. 20, No. 7, pp. 703–726, 2006.

[17] C. Rinner and M. Raubal. Personalized Multi-Criteria Decision Strategies in Location-Based Decision Support. *Journal of Geographic Information Sciences*, Vol. 10, No. 2, pp. 149–156, 2004.

[18] T.L. Saaty. *The Analytic Hierarchy Process*. McGraw Hill, New York, USA, 1980.

[19] S.Y.W. Su, J. Dujmović, D.S. Batory, S.B. Navathe and R. Elnicki. A Cost-Benefit Decision Model: Analysis, Comparison, and Selection of Data Management Systems. *ACM Transactions on Database Systems*, Vol. 12, No. 3, pp. 472–520, 1987.

[20] L.J. Wood and S. Dragićević. GIS-based multicriteria evaluation and fuzzy sets to identify priority sites for marine protection. *Biodivers. Conserv.*, Vol. 16, pp. 2539–2558, 2007.

[21] R.R. Yager. On ordered weighted averaging aggregation operators in multi-criteria decision making. *IEEE Trans. on Systems, Man and Cybernetics*, Vol. 18, No. 1, pp. 183–190, 1988.

[22] M. Zeleny. *Multiple Criteria Decision Making*. McGraw Hill, New York, USA, 1982.

ISBN: 978-989-95079-6-8

Computing with Actions: The case of driving a car in a simulated car race

Sergio Guadarrama *
Fundamentals for Soft Computing Unit
European Centre for Soft Computing
Mieres, Spain
sergio.guadarrama@softcomputing.es

Abstract— *Describing and modeling a complex system is very hard and time consuming, so, we propose to use a linguistic approach introduced by Zadeh. By taking a linguistic approach we can use linguistic variables and approximate rules to characterize approximately phenomenons which are too complex or too ill-defined to be done in numerical terms.*

Starting from the results obtained in our previous works, in this paper we propose to decompose complex behaviors in simple behaviors, and context dependent rules for the behaviors coordination, which are described by set of fuzzy rules. The use of fuzzy rules for behavior coordination and behavior description has allowed us to have a shared framework that makes easy to develop and to compute with linguistic descriptions of actions.

We propose to use a mixture of competitive and cooperative co-evolutionary algorithms to manage the big search space and tackle a complex and decomposable problem.

We use the Fuzzy-IEEE 2007 Car Racing Competition as an example of a problem that allows to show the benefits of our approach. In the example presented the results obtained are better and the flexibility is increased with respect to previous works.

Keywords— Computing with Words, Computing with Actions, Fuzzy Linguistic Descriptions, Fuzzy Controllers, Co-evolutionary Algorithms, Behavior coordination.

1 Introduction

Real-life problems with solutions that are complex and unknown in advance require to build and test several prototypes before being able to build a good enough solution. Prototypes allow designers to devise what an appropriate solution might look like while they work at intermediate stages of development. In this way, designers require to test potential solutions, following a design-test-update cycle before reaching a good enough solution (as suggested in the Agile software development [1]). Computers are useful tools to develop and test these prototypes, which in turn serve designers to explore, build and test different ideas.

Describing and modeling a complex system is very hard and time consuming. So, we propose to use a linguistic approach introduced by Zadeh in [2, 3, 4, 5, 6] to deal with complex systems and complex concepts. By taking a linguistic approach we can use linguistic variables and approximate rules to characterize approximately phenomenons which are too complex or too ill-defined to be done in numerical terms [7, 8]. We use linguistic descriptions to describe the desired solution, problem decomposition techniques and fuzzy controllers to build the solution, and co-evolutionary algorithms to tune and improve the solutions.

*His work has been supported by the Spanish Department of Science and Innovation (MICINN) under program Juan de la Cierva JCI-2008-3531, and the European Social Fund.

Indeed, in this work we are introducing a tentative proposal to compute with actions by computing with linguistic descriptions of actions, paralleling Zadeh's Computational Theory of Perceptions [9], in which he proposed to compute with perceptions by means of computing with linguistic descriptions of perceptions.

To cope with these problems, we combine our previous ideas, improvements and results in the areas of computing with words [10], fuzzy controllers [11] and co-evolutionary algorithms [12].

The rest of the paper is organized as follows. In section 2, we provide details of the problem tackled, the FUZZ-IEEE 2007 Car Racing Competition. In section 3, the solution based on linguistic description of the behaviors is presented. The coordination between behaviors is presented in section 5. The mixture of co-evolutionary algorithms is presented in section 6. In section 7 we present the results and finally, we summarize our work and provide conclusions in section 8.

2 The problem: FUZZ-IEEE 2007 Car Racing Competition

Car Racing Competition is a well-known sport that gets attention from many people in a number of different scenarios, ranging from popular real-life Fomula One to virtual computer games. Particularly, we tackle the challenge of building a car pilot for the FUZZ-IEEE 2007 Car Racing Competition.

Driving is a complex problem and how to find a general solution is unknown. Everyone need a lot practice before becoming a good driver. Actually, people become driving experts by practicing in many different scenarios. So we will use the simulator provided by the FUZZ-IEEE 2007 Car Competition to simulate different races that will allow us to learn and test solutions in many different scenarios.

In this problem, the challenge was to develop the best fuzzy car driver controller for a basic form of car race. A competition in which a pair of simulated cars race through a series of waypoints within a 2d plane, and compete between them to reach the maximum number of waypoints. Basic aspects of the car race are the following:

1. A car only sees the next three waypoints at a given time.
2. Waypoints have to be visited in order and the next one to be visited is highlighted.
3. Waypoints are randomly generated on a 2d plane.
4. The car that first reaches the next waypoint scores.
5. Cars cannot collide with each other, so two cars can follow the same path at the same time.

ISBN: 978-989-95079-6-8

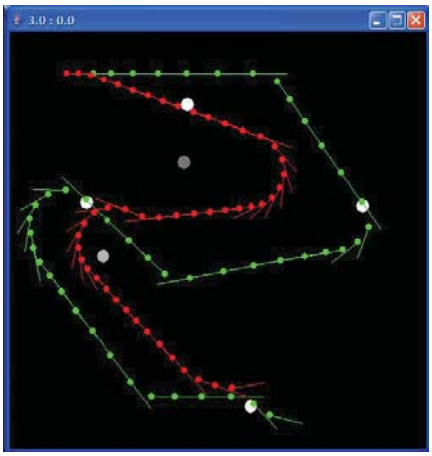


Figure 1: Car Racing Simulation on a 2d Plane.

6. Cars have imperfect sensors and actuators: gaussian noise is used to simulate the lack of perfect observations on position and velocity.

7. General driving ability rather than specialization to a particular track it is tested.

3 The solution: Behaviors described by linguistic actions

We describe a desired behavior by describing linguistically a set of actions modeled by fuzzy rules. So each behavior will be described by a set of fuzzy rules, and a solution will be composed by one or more behaviors coordinated appropriately.

Fuzzy rules are linguistically interpretable units that can serve our purposes of designing and testing prototypes easily and quickly, since they are easily human-interpretable, as well as, adjustable.

We have used fuzzy controllers [13, 14, 15] to implement individual behavior units. Fuzzy controllers incorporate heuristic control knowledge in the form of if-then rules, whose membership functions are later tuned by the co-evolutionary algorithm, similarly as it was proposed by Karr in [16].

We have used Xfuzzy 3.0, a fuzzy system development environment, to design our fuzzy inference system [17]. This system contains a set of tools that allows not only to easily describe a fuzzy inference model but also to implement it, and export it to programming languages such as JAVA, C and C++.

In this section we present the three different behaviors that we introduced in our previous paper [12]. The first behavior "Go straight to next target" was a local solution in the sense that the other targets were not considered (the other two way-points). The second behavior "Follow a spline curve" takes into account the next three way-points, thus following a more global approach. The third behavior "Avoiding ellipses around targets" cope with unwanted situation of going around a target indefinitely.

3.1 Behavior 1: Go straight to the target

Basically, this fuzzy controller takes as input variables: velocity, distance to next target and angle to next target; and has as output variables: acceleration and turn angle. The final controller is composed of two rule-bases:

- Acceleration rules:
  If Distance is Large then Acceleration is Very Positive.
  If Distance is Medium then Acceleration is Positive.
  If Distance is Small and Velocity is Very Positive then Acceleration is Very Negative.
  If Distance is Small and Velocity is Positive then Acceleration is Negative.
  If Distance is not Small and Velocity is Zero then Accelerations is Positive.

- Turn rules:
  If Angle is Zero then Turn Zero
  If Angle is Positive then Turn Positive
  If Angle is Negative then Turn Negative
  If Angle is not Zero and Distance is Very Small then Turn Zero

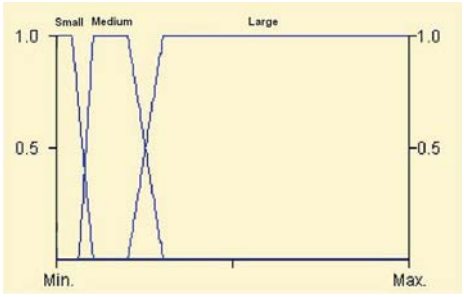


Figure 2: Distance Linguistic Variable

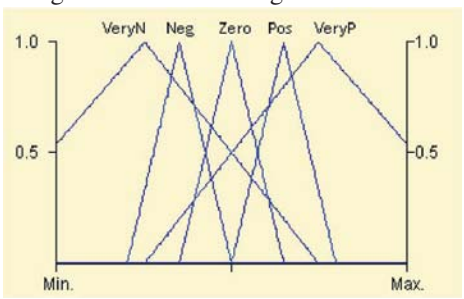


Figure 3: Acceleration Linguistic Variable

We impose the following ordering constrains [18] between the linguistic labels of the linguistic variables to maintain the interpretability of the rules, although these don't impose that they should form a partition:

- Distance = { Small $\prec$ Medium $\prec$ Large }
- Velocity = { Very Negative $\prec$ Negative $\prec$ Zero $\prec$ Positive $\prec$ Very Positive }
- Acceleration = { Very Negative $\prec$ Negative $\prec$ Zero $\prec$ Positive $\prec$ Very Positive }

After tuning parameters of the fuzzy membership functions using a genetic algorithm (see figures 2 and 3) of the fuzzy rules this behavior was able to defeat the heuristic controller in all races (see table 2 for comparison). So from now on we took it as the new baseline for comparison with the rest of the solutions.

ISBN: 978-989-95079-6-8

3.2 Behavior 2: Follow a spline curve

We observed that the previous behavior was able to align with the first target (next target) quite well, but it did not consider the positions of second and third targets. In other words, direction of the car was perfect for the next target, but it was not exploiting the knowledge about the rest other targets.

To do that, we use a spline curve instead of following straight lines, in order to follow a global strategy and gain advantage over the opponent, as soon as possible. We built a spline curve that goes through four points on a 2d plane: the position of our car, and the next three targets. We use the spline curve to calculate/approximate intermediate points between our car and the next target, and then follow them using these fuzzy rules:

- Acceleration rules:
  If Distance is Large then Acceleration is Very Positive.
  If Distance is Medium then Acceleration is Positive.
  If Distance is not Small and Velocity is Zero then Accelerations is Positive.
- Turn rules:
  If Angle is Zero then Turn Zero
  If Angle is Positive then Turn Positive
  If Angle is Negative then Turn Negative

On the negative side of this approach we can say that the distance that the car is covering is higher than in our previous behavior, that is, without using a spline curve, so when we set up a competitive co-evolutionary strategy this behavior (B2) was always loosing, unable to defeat the previous one (B1) (see table 3 for comparison).

3.3 Behavior 3: Avoid ellipses around targets

We also noticed that some races were ended when both cars were unable to reach a target. This situation occurred when a new target appeared and both cars were: 1, too close to the new target; and 2, with such a high angle to the target that it could not be corrected enough, given the short distance and the velocity of the cars. As a result, cars spent their last moments turning around the target.

To avoid that, we add this behavior (B3), that detects the unwanted situation and takes a different action. In order to detect that unwanted situation, we keep track of what happened in last n iterations. Particularly, we consider three facts during the last n iterations: 1, the target remained the same; 2, the angle to target was around 80 degrees; 3, the distance to target was small.

To address the described scenario, when the unwanted situation is detected then this new behavior is activated during a number m of iterations. This new behavior was built following the same rules that *"Go straight to the target"* with one sole difference: it did not go towards the target but in the opposite direction (it has the same rules as behavior 1 but with opposite sign).

4 Tuning Fuzzy Membership Functions by Co-evolution

Developing a good prototype from scratch is very time consuming and difficult since it involves a good understanding of the problem to do a good design, and a good understanding of the solution to adjust all its parameters. We did this initially manually and then used genetic algorithms, and co-evolutionary algorithms.

Table 1: Summary of previous results

	HC	GA	$Comp_1$	$Coop_\alpha$
HC	-	15	0	0
GA	83	-	1	5
$Comp_1$	100	99	-	47
$Coop_\alpha$	100	81	47	-

To avoid complexity and exploit the features of co-evolutionary algorithms we defined different species with different attributes to optimize. Actually, we divided our population into three species. Species number 1 consisted of individuals of behavior 1 *"Go straight to target"*, species number 2 was formed by individuals belonging to behavior 2 *"Follow a spline curve"*, and species number 3 was formed by individuals of behavior 3 *"Avoid ellipses"*.

So, we built a population with three different species with 20 individuals each, which evolve during 1000 generations. These species only met when racing (to calculate the fitness functions of individuals), while for the rest (cross, mutation, selection of individuals) were kept apart.

In [12] we compared how the use of a competitive or a cooperative co-evolutionary algorithm affects to the final solution. Obtaining that both approaches can find very good solutions, once the appropriate crossover operator is chosen. Table 6 recall the summary of results.

1. *HC*, hand coded model;
2. *GA*, complete model including three pilots, automatically tuned using a genetic algorithm;
3. $Comp_1$, complete model including three pilots, automatically tuned using a competitive coevolutionary approach, with 1-point cross operator;
4. $Coop_\alpha$, complete model including three pilots, automatically tuned using a cooperative coevolutionary approach, with BLX-alpha cross operator.

5 Behavior coordination

The main two problems of behavior coordination are:

1. Arbitration: deciding which behavior should be activated in each situation. This can be decided a priori or dynamically by establishing variable priorities. The simplest solution is a switching scheme that selects one behavior and ignores the rest. More flexible arbitration can be obtained using fuzzy context rules, that allow to activate concurrently several behaviors with different degrees of activation.
2. Command fusion: If only one behavior is active at a time, then there is not need to do it, but if there are more than

ISBN: 978-989-95079-6-8

one behavior active simultaneously, then it is needed to combine in some way the different commands generated by them. It is needed to fuse, or aggregate, properly these commands in order not to send a set of contradictory commands but to send a coherent fused command.

Context dependent blending is a flexible and general form of solving the problem of behavior coordination, since it solves both arbitration and command fusion by using a common scheme, fuzzy context rules and fuzzy aggregation. It was initially proposed by Ruspini in [19] and later expanded by Saffioti and himself in [20, 21].

5.1 Context dependent behavior coordination

In our previous work we defined a fix arbitration scheme to switch between behaviors. And compared competitive and cooperative co-evolutionary algorithms to tune the membership functions of the linguistic variables used in the fuzzy rules.

In this work we use context dependent blending, using fuzzy rules for both behavior coordination and behavior description, since that allowed us to have a shared framework that made easier to develop complex behaviors.

To address the coordination problem, we defined context rules for each behavior, so they could be activated in different situations, and serve to merge their outputs when are activated at the same time.

- Context Rule 1:
  If Distance is Small then "Go Straight to the target".
- Context Rule 2:
  If Distance is Large then "Follow a spline curve".
- Context Rule 3:
  If Distance is Small and Angle is approx. 80^o then "Avoid ellipses".

This allows to follow a local strategy when the car is near to the next target, a global strategy when the car is far to the next target, and avoiding ellipses around targets

We set up a mixture of competitive and cooperative co-evolutionary strategy, the individual behaviors cooperate to form the complex behavior (B123) while competing with the best solution obtained previously (B1). (see table 5 for comparison).

In figure 4 we can see a diagram of how the coordination of the different behaviors works. Using a weighted fusion between the commands of behaviors 1 and 2 and a switching scheme with behavior 3.

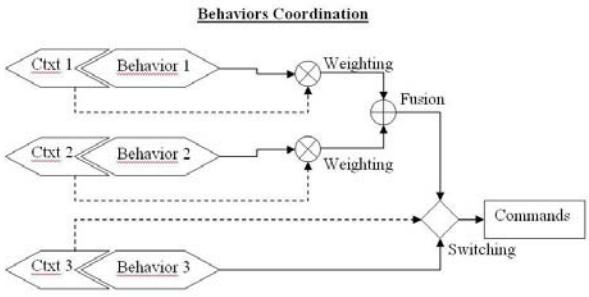


Figure 4: Behaviors Coordination

6 Mixture of Co-evolutionary algorithms

In some way, approaches pointing in the same direction that ours have already been proposed. In fact, many papers on fuzzy systems tuned with genetic algorithms have been published (for a good review see [22]).

Genetic algorithms, of single-population algorithms, are not well suited to problems which have solutions that are long, complex and interdependent [23]. In the last years, co-evolutionary algorithms have shown their ability to manage big search spaces and to tackle complex problems. Specially when the solution to the problem may be decomposable, or when there exist different competitive solutions [24, 25].

Co-evolutionary algorithms can be seen as an extension of traditional evolutionary algorithms, where there are, at least, two species that are simultaneously evolved and their fitness functions are coupled [24, 25]. Therefore, each species has its own coding scheme and reproduction scheme, but each individual is evaluated considering individuals from other species. Depending on the type of interaction between species, we can distinguish between competitive or cooperative species.

6.1 Competitive Co-evolution

In this context, each individual represents a candidate solution to the problem and competes with individuals of other species to find better solutions. Thus, if one individual increases its fitness, then fitness functions from the rest individuals decrease [24]. This may be seen as a kind of struggle between species, which presses individuals to compete between them for finding better solutions.

6.2 Cooperative Co-evolution

In this case, each individual represents a component of the solution and requires the rest components to build a complete solution. Therefore, an individual needs to cooperate with others to find a better solutions [25]. Thus, an individual fitness is linked with fitness values of its cooperative individuals. This is often seen as a kind of symbiosis between species to preserve existence, leading individuals to cooperate for finding better solutions.

6.3 Mixture of Competitive and Cooperative Co-evolution

We have different solutions that compete or cooperate to obtain better solutions. Following the two co-evolutionary approaches we have defined a mixed environment, where alternative solutions were competing to obtain better solutions, and where complementary solutions were cooperating to obtain better solutions. At the same time that the three different behaviors evolve in a cooperative way, they are competing with the best previous solutions.

The main features of the co-evolutionary algorithm used to tune the parameters of the membership functions are the following:

generations = 1000.
population size = 20.
crossover probability = 0.9.
mutation probability = 0.1.
crossover operator = BLX-α.
$\alpha = 0.7$.
species = 3.
co-evolutionary strategy = mixed.

ISBN: 978-989-95079-6-8

Each specie have its own coding as follows:

- Go Straight to Target (51 genes) + Context 1 (2 genes)
- Following a Spline Curve (39 genes) + Context 2 (2 genes)
- Avoiding Ellipses (51 genes) + Context 3 (5 genes)

7 Results

In this section it is shown the sum of the results we obtained during the development of the system and a comparative evaluation of the obtained solutions (for previous results see [12]).

Table 2:
Heuristic vs Behavior 1

Race	HC	B1
1	18	37
2	15	35
3	12	42
4	16	44
5	11	43
6	19	41
7	13	46
8	12	42
9	20	44
10	17	39
Total	153	413

Table 3:
Behavior 1 vs Behavior 2

Race	B1	B2
1	47	4
2	44	2
3	44	5
4	43	7
5	42	6
6	44	3
7	43	9
8	42	8
9	42	5
10	44	5
Total	435	54

Table 2 shows the number of way-points (targets) got in 10 races by the hand coded (HC) against the behavior 1 (B1) "Go straight to the next target", in which we can see that the behavior 1 is much better than the hand coded controller, and it is wining all the races by a big difference.

Table 3 shows the number of way-points (targets) got in 10 races by the behavior 1 (B1) "Go straight to the next target" against the behavior 2 (B2) "Follow the spline curve", in which we can see that behavior 2 is performing very bad against behavior 1, loosing all the races by a huge difference of points, and performing even worse than the hand coded controller.

These results indicate that the best individual behavior is "Go straight to the next target", and that to improve these results we need to combine it with other behaviors, as we did in the next experiments.

Table 4 shows the number of way-points (targets) got in 10 races by the behavior 1 (B1) "Go straight to the next target" against the combined behavior 1 and 2 (B12) "Go straight to the next target; Follow the spline curve", in which we can see that the combined behavior 12 is much better that the behavior 1 alone, and it is wining all the races by some difference. Table 5 shows the number of way-points (targets) got in 10 races by the behavior 1 (B1) "Go straight to the next target" against the combined behavior 1, 2 and 3 (B123) "Go straight to the next target; Follow the spline curve; Avoid ellipses", in which

Table 4:
Behavior 1 vs Behavior 12

Race	B1	B12
1	26	35
2	27	37
3	25	32
4	31	32
5	24	38
6	32	31
7	28	36
8	26	36
9	28	35
10	30	40
Total	277	352

Table 5:
Behavior 1 vs Behavior 123

Race	B1	B123
1	19	35
2	18	45
3	13	43
4	22	33
5	18	32
6	20	34
7	16	37
8	26	28
9	24	32
10	20	36
Total	196	355

we can see that behavior 123 is performing very well against behavior 1 wining all the races by a big difference of points, and performing even better than the combined behavior 1 and 2 (B12).

These results indicate that the best behavior is the combined behavior 123. But after these comparisons and to evaluate the impact of the context blending behavior coordination, we compared it with the best solutions from our previous work $Comp_1$ and $Coop_\alpha$.

We set up a league competition among all of them, by playing 100 races between each pair of solutions. Table 6 shows the results of the five solutions considered: *Hand-coded*, $Comp_1$, $Coop_\alpha$ and $B123$, and their associated wins for each one in the 100 races between each pair of solutions.

Table 6: Wins Against Each Other

	HC	$Comp_1$	$Coop_\alpha$	$B123$	Total
HC	-	0	0	0	0
$Comp_1$	100	-	47	20	167
$Coop_\alpha$	100	51	-	31	182
$B123$	**100**	**74**	**63**	–	**237**

Table 6 shows the associated wins of each solution. Both behaviors, $Comp_1$ and $Coop_\alpha$, obtained good performance but the new solution $B123$ outperformed them with a greater number of wins than the other approaches, which indicates that: 1, the context dependent blending provides a great advantage compared to fixed arbitration coordination; 2, automatically tuning of fuzzy membership functions yield better results than using hand-coded solutions.

8 Conclusions

In the particular context of the FUZZ-IEEE 2007 Car Racing Competition we have introduced a tentative proposal to compute with actions by computing with linguistic descriptions of behaviors. In this context we have identified and tackled the

ISBN: 978-989-95079-6-8

following key aspects: how to define different behaviors by describing them linguistically and how to coordinate several behaviors using context dependent rules.

We have set up a mixture of competitive and cooperative co-evolutionary algorithms to model the membership functions of the linguistic variables used. On one hand, competitive co-evolution allows to find alternative solutions that compete among them to find better solutions and to avoid being trapped in local minima. On the other hand, cooperative co-evolution allows to find complementary solutions that cooperate among them to find complex and better solutions.

By following an incremental approach, we were able to decompose the solution to this problem in simple behaviors with a lower degree of complexity. The use of fuzzy rules for behavior coordination and behavior description has allowed us to have a shared framework that makes easy to develop complex behaviors and to compute with linguistic descriptions of actions. The use of co-evolutionary algorithms has allowed us to test, compare, combine and tune the different behaviors appropriately and obtain a better solution.

In general, it is not possible to predict a priori all the scenarios, problems, or solutions. So a "learning by doing" approach is a need in these cases. The know-how obtained by testing and playing with different solutions is irreplaceable in order to build systems that compute with actions described by words. This approach allows to mix designer's knowledge and knowledge learned from the data, maintaining a good interpretability of the solution. This represents a very important step-stone that would allow to change and improve the system in front of new environments and situations.

Acknowledgment

The author want to thank to Claudio Moraga, David Perez and to the anonymous reviewers for their comments and fruitful discussion about the paper.

References

[1] J. Highsmith and A. Cockburn. Agile software development: The business of innovation. *Computer*, 34(9):120–127, 2001.

[2] L.A. Zadeh. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transaction on Systems, Man and Cybernetics*, 3:28–44, 1973.

[3] L.A. Zadeh. On the analysis of large-scale systems. In *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers*, pages 195–209. World Scientific Pub Co Inc, 1996.

[4] L.A. Zadeh. A fuzzy-algorithmic approach to the definition of complex or imprecise concepts. *Int. J. Man-Machine Studies*, 8:249–291, 1976.

[5] L.A. Zadeh. Precisiated natural language (PNL). *AI MAGAZINE*, 25(3):74–91, 2004.

[6] L.A. Zadeh. A New Frontier in Computation? Computation with Information Described in Natural Language. In *Proceedings Symposium on Fuzzy Systems in Computer Science FSCS*, pages 1–2, 2006.

[7] L. A. Zadeh. Fuzzy logic = computing with words. *IEEE Trans. Fuzzy Systems*, 4:103–111, 1996.

[8] L.A. Zadeh. From computing with numbers to computing with words.from manipulation of measurements to manipulation of perceptions. In P.P. Wang, editor, *Computing with words*, pages 35–68. John Wiley & Sons, 2001.

[9] L.A. Zadeh. A New Direction in AI: Toward a Computational Theory of Perceptions. *AI Magazine*, 22(1):73, 2001.

[10] S. Guadarrama. *A Contribution to Computing with Words / Perceptions*. PhD thesis, Universidad Politécnica de Madrid, Madrid, April 2007. http://oa.upm.es/448/.

[11] C. Moraga, E. Trillas, and S. Guadarrama. Multiple-valued logic and artificial intelligence: Fundamentals of fuzzy control revisited. *Artif. Intell. Rev.*, 20(3-4):169–197, 2003.

[12] S. Guadarrama and R. Vazquez. Tuning a fuzzy racing car by coevolution. In *Genetic and Evolving Systems, 2008. GEFS 2008. 3rd International Workshop on*, pages 59–64, 2008.

[13] L.A. Zadeh. A rationale for fuzzy control. *Jr. of Dynamic Systems, Measurements and Control*, 34:3–4, 1972.

[14] S. Assilian and E.H. Mamdani. Learning control algorithms in real dynamic systems. In *Int. IFAC/IFIP Conf. on Digital Computer Appl. to Process Control*, ZÜrich, 1974. IFAC/IFIP.

[15] E.H. Mamdani. Advances in the linguistic synthesis of fuzzy controllers. *Int. Jr. Man-Machine Studies*, 8:669–678, 1976.

[16] C. Karr. Genetic algorithms for fuzzy controllers. *AI Expert*, 6(2):2633, 1991.

[17] F.J. Moreno-Velo, I. Baturone, S. Sanchez-Solano, and A. Barriga. XFUZZY 3.0: A Development Environment for Fuzzy Systems. In *Proc. International Conference in Fuzzy Logic and Technology*, pages 93–96, 2000.

[18] JJ Saade and H. Schwarzlander. Ordering fuzzy sets over the real line: an approach based on decisio making under uncertainty. *Fuzzy sets and systems*, 50(3):237–246, 1992.

[19] E.H. Ruspini. Fuzzy logic in the flakey robot. In *Procs. of the Int. Conf on Fuzzy Logic and Neural Networks*, pages 767–770, Iizuka, Japan, 1990.

[20] A. Saffiotti, K. Konolige, and E.H. Ruspini. A multivalued logic approach to integrating planning and control. *Artificial Intelligence*, 76(1-2):481–526, 1995.

[21] A. Saffiotti, E.H. Ruspini, and K. Konolige. Blending reactivity and goal-directedness in a fuzzy controller. In *Fuzzy Systems, 1993., Second IEEE International Conference on*, pages 134–139, 1993.

[22] O. Cordon, F. Herrera, F. Gomide, F. Hoffmann, and L. Magdalena. Ten years of genetic fuzzy systems: current framework and new trends. *Fuzzy Sets and Systems*, 141(1):5–31, 2004.

[23] C.A. Pena-Reyes and M. Sipper. Fuzzy CoCo: a cooperative-coevolutionary approach to fuzzy modeling. *Fuzzy Systems, IEEE Transactions on*, 9(5):727–737, 2001.

[24] C.D. Rosin and R.K. Belew. New Methods for Competitive Coevolution. *Evolutionary Computation*, 5(1):1–29, 1997.

[25] M.A. Potter and K.A.D. Jong. Cooperative Coevolution: An Architecture for Evolving Coadapted Subcomponents. *Evolutionary Computation*, 8(1):1–29, 2000.

ISBN: 978-989-95079-6-8

Analysis of Similarity Measures for Atanassov's Intuitionistic Fuzzy Sets

Eulalia Szmidt and Janusz Kacprzyk

Systems Research Institute, Polish Academy of Sciences,
ul. Newelska 6, 01–447 Warsaw, Poland
Email: {szmidt, kacprzyk}@ibspan.waw.pl

Abstract— *We consider some existing similarity measures for Atanassov's intuitionistic fuzzy sets (A-IFSs, for short). We show that neither similarity measures treating an A-IF as a simple interval values fuzzy set, nor straightforward generalizations of the similarity measures well-known for the classic fuzzy sets work under reasonable circumstances. Next, expanding upon our previous works, we consider a family of similarity measures constructed by taking into account both all the three functions (the membership, non-membership and hesitation) describing an A-IF, and the complements of the elements we compare to each other. That is, we use all kinds and fine shades of information available. We point out their proper behavior and an intuitive appeal.*

Keywords— Atanassov's intuitionistic fuzzy sets, similarity measures.

1 Introduction

Atanassov's intuitionistic fuzzy sets (Atanassov [1], [2], [3]) – to be called A-IFSs, for short – can be viewed as a tool that may better model and process imperfect information. The use of positive and (independently) negative information that is the core of A-IFSs is natural in real life and is also well-known, advocated and studied in psychology and other social and behavioral sciences [e.g., [21], [14]]. It also attracted much attention and research interest in soft computing. It would be difficult to deal with machine learning (making use of examples and counter-examples), modeling of preferences or voting without taking into account positive and (independently) negative testimonies or opinions. Although from a mathematical point of view A-IFSs are equipotent, under some assumptions, to interval-valued fuzzy sets (cf. Atanassov and Gargov in 1989 [4]), from the point of view of solving problems (starting from the stage of collecting data), they are different as A-IFSs force the user to explicitly consider positive and negative information independently. On the other hand, while employing the interval-valued fuzzy sets, the user's attention is focused on positive information (in an interval form) only. Notably, Dubois [9] noticed recently that A-IFSs correspond to an intuition that differs from that behind the interval valued fuzzy sets.

The fact that people tend to notice and take into account only most obvious aspects (e.g. advantages only) when making decision is well known in psychology (cf. Kahneman [14]), Sutherland [21]) and may often lead to improper decisions. In this context, A-IFSs ("forcing" a decision maker to take into account both negative and positive aspects of the decisions) may be seen as belonging to modern and promising means of knowledge representation and processing.

We consider here some major similarity measures for A-IFSs. First we present a whole array of similarity measures (known from the literature) for A-IFSs viewed in terms of single intervals. Second, we consider measures that are straightforward generalization of the similarity measures for the conventional fuzzy sets. Unfortunately, both do not meet our expectations, and we provide some counter-intuitive examples. It seems to be an indirect hint that A-IFSs are not functionally equivalent to the interval valued fuzzy sets.

Next, we reconsider our (cf. Szmidt and Kacprzyk [41]) concept of a similarity measure between A-IFSs taking into account all three functions (the membership, non-membership and hesitation), and explicitly add to the above three functions the complements of the elements we compare to each other.

2 A brief introduction to A-IFSs

One of the possible generalizations of a fuzzy set in X (Zadeh [48]), given by

$$A^{'} = \{< x, \mu_{A^{'}}(x) > | x \in X\} \tag{1}$$

where $\mu_{A^{'}}(x) \in [0,1]$ is the membership function of the fuzzy set $A^{'}$, is Atanassov's intuitionistic fuzzy set (Atanassov [1], [2], [3]), A-IF, A given by

$$A = \{< x, \mu_A(x), \nu_A(x) > | x \in X\} \tag{2}$$

where: $\mu_A : X \rightarrow [0,1]$ and $\nu_A : X \rightarrow [0,1]$ such that

$$0 \leq \mu_A(x) + \nu_A(x) \leq 1 \tag{3}$$

and $\mu_A(x)$, $\nu_A(x) \in [0,1]$ denote the degree of membership and a degree of non-membership of $x \in A$, respectively.

Each fuzzy set may be represented by the following A-IF

$$A = \{< x, \mu_{A^{'}}(x), 1 - \mu_{A^{'}}(x) > | x \in X\} \tag{4}$$

For each intuitionistic fuzzy set in X, we will call

$$\pi_A(x) = 1 - \mu_A(x) - \nu_A(x) \tag{5}$$

an *intuitionistic fuzzy index* (or a *hesitation margin*) of $x \in A$ and, it expresses a lack of knowledge of whether x belongs to A or not (cf. Atanassov [3]). It is obvious that $0 \leq \pi_A(x) \leq 1$, for each $x \in X$.

The hesitation margin turns out to be important while considering the distances (Szmidt and Kacprzyk [26], [29], [39], entropy (Szmidt and Kacprzyk [31], [40]), similarity (Szmidt and Kacprzyk [41]) for the A-IFSs, etc. i.e., the measures that play a crucial role in virtually all information processing tasks.

In our further considerations we will use the following distances between fuzzy sets A, B in $X = \{x_{1,}, \ldots, x_n\}$ Szmidt

ISBN: 978-989-95079-6-8

and Baldwin [23], [24], Szmidt and Kacprzyk [29], [39]: the normalized Hamming distance

$$l_{IFS}(A,B) = \frac{1}{2n}\sum_{i=1}^{n}(|\mu_A(x_i) - \mu_B(x_i)| + |\nu_A(x_i) - \nu_B(x_i)| + |\pi_A(x_i) - \pi_B(x_i)|) \quad (6)$$

and the normalized Euclidean distance:

$$q_{IFS}(A,B) = (\frac{1}{2n}\sum_{i=1}^{n}(\mu_A(x_i) - \mu_B(x_i))^2 + (\nu_A(x_i) - \nu_B(x_i))^2 + (\pi_A(x_i) - \pi_B(x_i))^2)^{\frac{1}{2}} \quad (7)$$

For distances (6), and (7) we have $0 \leq l_{IFS}(A,B) \leq 1$, and $0 \leq q_{IFS}(A,B) \leq 1$. Clearly these distances satisfy the conditions of the metric.

In our further considerations we will use the notion of the complement elements, which definition is a simple consequence of a complement set A^C

$$A^C = \{< x, \nu_A(x), \mu_A(x) > | x \in X\} \quad (8)$$

The use of A-IFSs instead of fuzzy sets implies the introduction of another degree of freedom (non-memberships) into the set description. Such a generalization of fuzzy sets gives us an additional possibility to represent imperfect knowledge which leads to describing many real problems in a more adequate way. Applications of intuitionistic fuzzy sets to group decision making, negotiations, voting and other situations are presented in Szmidt and Kacprzyk [25], [27], [28], [30], [32], [34], [33], [35], [38], Szmidt and Kukier [42], [43].

3 Some counter-intuitive results given by the traditional similarity measures

In the literature there is a multitude of similarity measures both for A-IFSs (Atanassov [1, 2, 3], and vague sets (Gau and Buehrer [10]) which have also been proved to be equivalent to A-IFSs (Bustince and Burillo [5]). Here we adopt the notation for A-IFSs but we will consider the measures originally introduced for vague sets, too.

Chen [6, 7] considered similarity measures between two A-IFSs A and B as

$$S_C(A,B) = 1 - \frac{\sum_{i=1}^{n}|S_A(x_i) - S_B(x_i)|}{2n} \quad (9)$$

where $S_A(x_i) = \mu_A(x_i) - \nu_A(x_i)$, $S_A(x_i) \in [-1,1]$ and $S_B(x_i) = \mu_B(x_i) - \nu_B(x_i)$, $S_B(x_i) \in [-1,1]$.

But, as Hong and Kim [11] noticed

$$\mu_A(x_i) - \nu_A(x_i) = \mu_B(x_i) - \nu_B(x_i) \Rightarrow S_C(A,B) = 1 \quad (10)$$

which is counterintuitive as, e.g., for $A = (x,0,0)$ and $B = (x,0.5,0.5)$, we have $S_C(A,B) = 1$.

To overcome the problem of S_C (9), Hong and Kim [11] proposed the similarity measures S_H and S_L

$$S_H(A,B) = 1 - (\sum_{i=1}^{n}|\mu_A(x_i) - \mu_B(x_i)| + |\nu_A(x_i) - \nu_B(x_i)|)/2n \quad (11)$$

$$S_L(A,B) = 1 - \frac{1}{4n}((\sum_{i=1}^{n} S_A(x_i) - S_B(x_i)) - (\sum_{i=1}^{n}|\mu_A(x_i) - \mu_B(x_i)| + |\nu_A(x_i) - \nu_B(x_i)|)) \quad (12)$$

Since $S_H(A,B)$ takes into account the absolute values, it does not distinguish the positive from negative differences, e.g., for $A = \{(x,0.3,0.3)\}$, $B = \{(x,0.4,0.4)\}$, $C = \{(x,0.3,0.4)\}$, and $D = \{(x,0.4,0.3)\}$, we obtain from (11) that $S_H(A,B) = S_H(C,D) = 0.9$ which seems counter-intuitive.

$S_L(A,B)$ also gives counter-intuitive results, e.g. for $A = \{(x,0.4,0.2)\}$, $B = \{(x,0.5,0.3)\}$, $C = \{(x,0.5,0.2)\}$, we obtain from (12) $S_L(A,B) = S_L(A,C) = 0.95$ which seems counter-intuitive.

The same problem like with S_H occurs with the similarity measure (cf. Li et al. [16]):

$$S_O(A,B) = 1 - (1/2n)^{0.5}(\sum_{i=1}^{n}(\mu_A(x_i) - \mu_B(x_i))^2 + (\nu_A(x_i) - \nu_B(x_i))^2)^{0.5} \quad (13)$$

Dengfeng and Chuntian [15] considered the similarity measure:

$$S_{DC}(A,B) = 1 - (1/n^{1/p})(\sum_{i=1}^{n}(|m_A(x_i) - m_B(x_i)|)^p)^{1/p} \quad (14)$$

where $m_A(x_i) = (\mu_A(x_i) + 1 - \nu_A(x_i))/2$, $m_B(x_i) = (\mu_B(x_i) + 1 - \nu_B(x_i))/2$, $1 \leq p < \infty$. Unfortunately, as for (14), the medians of two intervals are compared only, it is rather easy to point out the counter-intuitive examples, e.g., $A = (x,0.4,0.2)$, $B = (x,0.5,0.3)$, then $S_{DC}(A,B) = 1$, for each p.

Mitchell [18] modified Dengfeng and Chuntian's measure S_{DC} (14) using a statistical approach by interpreting A-IFSs as families of ordered fuzzy sets. Let $\rho_\mu(A,B)$ and $\rho_\nu(A,B)$ denote a similarity measure between the "low" membership functions μ_A and μ_B, and between the "high" membership functions $1 - \nu_A$ and $1 - \nu_B$, respectively, as:

$$\rho_\mu(A,B) = S_{DC}(\mu_A, \mu_B) = 1 - (1/n^{1/p})(\sum_{i=1}^{n}(|\mu_A(x_i) - \mu_B(x_i)|)^p)^{1/p}$$

$$\rho_\nu(A,B) = S_{DC}(1 - \nu_A, 1 - \nu_B) = 1 - (1/n^{1/p})(\sum_{i=1}^{n}(|\nu_A(x_i) - \nu_B(x_i)|)^p)^{1/p}$$

Then the modified similarity measure between A and B is

$$S_{HB}(A,B) = (\rho_\mu(A,B) + \rho_\nu(A,B))/2 \quad (15)$$

Unfortunately, S_{HB} gives the same counter-intuitive results as S_H, for $p = 1$ and for one-element sets.

To overcome the drawbacks of S_{DC}, Liang and Shi [17] proposed $S_e^p(A,B)$, $S_s^p(A,B)$, $S_h^p(A,B)$ as:

ISBN: 978-989-95079-6-8

$$S_e^p(A,B) = 1 - (1/n^{1/p})(\sum_{i=1}^{n}(\phi_{\mu AB}(x_i) - \phi_{\nu AB}(x_i)|)^p)^{1/p} \quad (16)$$

where $\phi_{\mu AB}(x_i) = |\mu_A(x_i) - \mu_B(x_i)|/2$, $\phi_{\nu AB}(x_i) = |(1 - \nu_A(x_i))/2 - (1 - \nu_B(x_i))|/2$. But, for $p = 1$ and for one-element sets, $S_e^p(A,B)$=S_{HB}=S_H, which are again the same counter-intuitive results.

$$S_s^p(A,B) = 1 - (1/n^{1/p})(\sum_{i=1}^{n}(\varphi_{s1}(x_i) - \varphi_{s2}(x_i))^p)^{1/p} \quad (17)$$

where: $\varphi_{s1}(x_i) = |m_{A1}(x_i) - m_{B1}(x_i)|/2$,
$\varphi_{s2}(x_i) = |m_{A2}(x_i) - m_{B2}(x_i)|/2$,
$m_{A1}(x_i) = (\mu_A(x_i) + m_A(x_i))/2$,
$m_{A2}(x_i) = (m_A(x_i) + 1 - \nu_A(x_i))/2$,
$m_{B1}(x_i) = (\mu_B(x_i) + m_B(x_i))/2$,
$m_{B2}(x_i) = (m_B(x_i) + 1 - \nu_B(x_i))/2$,
$m_A(x_i) = (\mu_A(x_i) + 1 - \nu_A(x_i))/2$,
$m_B(x_i) = (\mu_B(x_i) + 1 - \nu_B(x_i))/2$.

S_s^p (17) avoids the problematic results obtained from S_{DC} (14) (for the intervals with equal medians) but, again, a problem of counter-intuitive results remains. For example, for $A = \{(x, 0.4, 0.2)\}$, $B = \{(x, 0.5, 0.3)\}$, $C = \{(x, 0.5, 0.2)\}$, we obtain $S_s^p(A,B) = S_s^p(A,C) = 0.95$ which seems difficult to accept.

S_h^p is given as

$$S_h^p(A,B) = 1 - (1/(3n)^{1/p})(\sum_{i=1}^{n}(\eta_1(i) + \eta_2(i) + \eta_3(i))^p)^{1/p} \quad (18)$$

where $\eta_1(i) = \phi_\mu(x_i) + \phi_\nu(x_i)$ (the same as for S_e^p),
$\eta_2(i) = m_A(x_i) - m_B(x_i))$ (the same as for S_{DC}),
$\eta_3(i) = \max(l_A(i), l_B(i)) - \min(l_A(i), l_B(i))$,
$l_A(i) = (1 - \nu_A(x_i) - \mu_A(x_i))/2$,
$l_B(i) = (1 - \nu_B(x_i) - \mu_B(x_i))/2$. But, again, there are counter-intuitive cases for this measure. For $A = (x, 0.3, 0.4)$, and $B = (x, 0.4, 0.3)$, i.e., for an element and its complement, we obtain $S_h^p(A,B) = 0.933$ (which seems to be rather too big a similarity for an element and its complement).

Hung and Yang [12] proposed the similarity measures S_{HY}^1, S_{HY}^2, S_{HY}^3 in which Hausdorff distances were employed:

$$S_{HY}^1(A,B) = 1 - d_H(A,B) \quad (19)$$

$$S_{HY}^2(A,B) = 1 - (e^{d_H(A,B)} - e^{-1})/(1 - e^{-1}) \quad (20)$$

$$S_{HY}^3(A,B) = (1 - d_H(A,B))/(1 + d_H(A,B)) \quad (21)$$

where
$d_H(A,B))$=$\sum_{i=1}^{n} \max(|\mu_A(x_i)$–$\mu_B(x_i)|, |\nu_A(x_i)$–$\nu_B(x_i)|)$
Unfortunately, (19)–(21) give counter-intuitive results (implied by the calculation of $d_H(A,B)$ - cf. Szmidt [22]). For example if $A = (x, 0.4, 0.5)$, $B = (x, 0.5, 0.4)$, $C = (x, 0.5, 0.3)$, $D = (x, 0.6, 0.4)$, $E = (x, 0.6, 0.3)$, $F = (x, 0.4, 0.3)$ then $S_{HY}^1(A,B) = 0.9$ (a counter-intuitive large similarity for A and its complement as $B = A^C$), and also $S_{HY}^1(C,D) = S_{HY}^1(C,E) = S_{HY}^1(C,F) = 0.9$. Next, $S_{HY}^2(A,B) = S_{HY}^2(C,D) = S_{HY}^2(C,E) = S_{HY}^2(C,F) = 0.85$, and also $S_{HY}^3(A,B) = S_{HY}^3(C,D) = S_{HY}^3(C,E) = S_{HY}^3(C,F) = 0.85$.

A straightforward attempt to calculate the similarity between A-IFSs just by adding the non-memberships values to the existing similarity measures for fuzzy sets is due to Hung and Yang [13]. Their measures (22) and (23) are the extension of Wang's measures [46]:

$$S_{w1}(A,B) = (1/n) \sum_{i=1}^{n} \frac{\min(\mu_A(x_i), \mu_B(x_i)) + \min(\nu_A(x_i), \nu_B(x_i))}{\max(\mu_A(x_i), \mu_B(x_i)) + \max(\nu_A(x_i), \nu_B(x_i))} \quad (22)$$

But, it is easy to give counter-examples again. For example, for $A = \{(x, 0, 0.5)\}$, $B = \{(x, 0.1, 0.5)\}$, $C = \{(x, 0, 0.6)\}$, we obtain $S_{w1}(A,B) = S_{w1}(A,C) = 0.8(3)$ (for different B and C we obtain the same result), for $A = \{(x, 0, 0.5)\}$, $B = \{(x, 0.18, 0.5)\}$, $C = \{(x, 0, 0.68)\}$, we obtain $S_{w1}(A,B) = S_{w1}(A,C) = 0.735$ (again - for different B and C the same result) etc., which seems to be difficult to accept (S_{w1} is not bijective).

$$S_{w2}(A,B) = (1/n) \sum_{i=1}^{n}(1 - 0.5(|\mu_A(x_i) - \mu_B(x_i)| + |\nu_A(x_i) - \nu_B(x_i)|)) \quad (23)$$

But for $A = \{(x, 0, 0.5)\}$, $B = \{(x, 0, 0.4)\}$, $C = \{(x, 0, 0.6)\}$, we obtain $S_{w2}(A,B) = S_{w2}(A,C) = 0.95$ (again - for different B and C the same similarity), which seems to be difficult to accept (S_{w2} is not bijective).

Another straightforward extensions of fuzzy similarity measures proposed by Hung and Yang [13] for A-IFSS are measures (24), (25) and (26) (they are extensions of Pappis and Karacapilidis' [20] measures for fuzzy sets).

$$S_{pk1}(A,B) = \frac{\sum_{i=1}^{n}(\min(\mu_A(x_i), \mu_B(x_i)) + \min(\nu_A(x_i), \nu_B(x_i)))}{\sum_{i=1}^{n}(\max(\mu_A(x_i), \mu_B(x_i)) + \max(\nu_A(x_i), \nu_B(x_i)))} \quad (24)$$

For (24) the counter-intuitive examples are like for (22).

$$S_{pk2}(A,B) = 1 - 0.5(\max_i(|\mu_A(x_i) - \mu_B(x_i)|) + \max_i(|\nu_A(x_i) - \nu_B(x_i)|)) \quad (25)$$

For (25) it is easy to give counter-examples again - especially for one-element sets. For example, for $A = \{(x, 0, 0.5)\}$, $B = \{(x, 0.1, 0.5)\}$, $C = \{(x, 0, 0.6)\}$, we obtain $S_{pk2}(A,B) = S_{pk2}(A,C) = 0.95$ (for different B and C just the same result).

$$S_{pk3}(A,B) = 1 - \frac{\sum_{i=1}^{n}(|\mu_A(x_i) - \mu_B(x_i)| + |\nu_A(x_i) - \nu_B(x_i)|)}{\sum_{i=1}^{n}(|\mu_A(x_i) + \mu_B(x_i)| + |\nu_A(x_i) + \nu_B(x_i)|)} \quad (26)$$

But for $A = \{(x, 0, 0.5)\}$, $B = \{(x, 0, 0.26)\}$, $C = \{(x, 0, 0.965)\}$, we obtain $S_{pks}(A,B) = S_{pk3}(A,C) = 0.68$ (again, for different B and C the same similarity), which seems to be difficult to accept.

ISBN: 978-989-95079-6-8

3.1 Why the measures presented may yield counter-intuitive results?

It is worth noticing that all the above measures were constructed to satisfy the following conditions:

$$S(A,B) \in [0,1] \quad (27)$$

$$S(A,B) = 1 \Longleftrightarrow A = B \quad (28)$$

$$S(A,B) = S(B,A) \quad (29)$$

$$If A \subseteq B \subseteq C, \quad then \quad S(A,C) \leq S(A,B) \; and \; S(A,C) \leq S(B,C) \quad (30)$$

Conditions (27)–(29) are obvious. The problem lies in (30) as this condition is meant as:

$$A \subset B \text{ iff } \forall x \in X, \; \mu_A(x) \leq \mu_B(x) \; and \; \nu_A(x) \geq \nu_B(x) \quad (31)$$

Unfortunately, (31) is not constructive and operational for A-IFSs as for many cases it can not be used. For example, for the elements $(x : (\mu, \nu, \pi))$: x_1: $(0.12, 0.4, 0.48)$ and x_2: $(0.1, 0.3, 0.6)$ we can not come to a conclusion. Moreover, element x: $(0, 0, 1)$ seems to be always beyond consideration in the sense of (31) which is very specific, and mostly practically irrelevant.

Furthermore, most of the similarity measures shown above are in fact similarity measures comparing just two intervals (each interval representing one of the A-IFSs under comparison). But we should bear in mind that elements of A-IF are described via the membership and non-membership function and the hesitation margin. In other words, in the terms of intervals, we have both the membership in an interval, and the non-membership in an interval so that we should represent an A-IF via two (not one) intervals.

Considering the representation of A-IFSs as single intervals implies some problems while calculating distances. Distances used in the (counter-intuitive) similarity measures mentioned in the previous section are calculated without taking into account the hesitation margins as the membership and non-membership functions only are taken into account. The counter-intuitive results obtained in such a case are in Szmidt and Kacprzyk [29], [39], Szmidt [22].

4 Some examples of intuitively justified and operational similarity measures

First we recall the measure of similarity between A-IFSs presented by Szmidt and Kacprzyk [37], [36]).

In the simplest situations we calculate the similarity of any two elements X and F belonging to an A-IF (A-IFSs). The proposed measures indicate if X is more similar to F or to F^C, where F^C is the complement of F. In other words, the proposed measures answer the question: is X more similar or more dissimilar to F?

Definition 1

$$Sim_{rule}(X,F) = \frac{l_{IF}(X,F)}{l_{IF}(X,F^C)} \quad (32)$$

where: $l_{IFS}(X,F)$ *is a distance from* $X(\mu_X, \nu_X, \pi_X)$ *to* $F(\mu_F, \nu_F, \pi_F)$*,*
$l_{IFS}(X,F^C)$ *is a distance from* $X(\mu_X, \nu_X, \pi_X)$ *to* $F^C(\nu_F, \mu_F, \pi_F)$*,*
F^C *is a complement of* F*, distances* $l_{IFS}(X,F)$ *and* $l_{IFS}(X,F^C)$ *are calculated from (6).*

For (32) we have

$$0 \leq Sim_{rule}(X,F) \leq \infty \quad (33)$$

$$Sim_{rule}(X,F) = Sim_{rule}(F,X)$$

The similarity has typically been assumed to be symmetric. Tversky [44], however, has provided some empirical evidence that the similarity should not always be treated as a symmetric relation. We stress this to show that a similarity measure (32) may have some features which can be useful in some situations but are not welcome in others (see Cross and Sudkamp [8], Wang et al. [47], Veltkamp [45]).

It is obvious (cf. Szmidt and Kacprzyk [36]) that the formula (32) can also be stated as

$$Sim_{rule}(X,F) = \frac{l_{IFS}(X,F)}{l_{IFS}(X,F^C)} = \frac{l_{IFS}(X^C,F^C)}{l_{IFS}(X,F^C)} = \frac{l_{IFS}(X,F)}{l_{IFS}(X^C,F)} = \frac{l_{IFS}(X^C,F^C)}{l_{IFS}(X^C,F)} \quad (34)$$

It is worth noticing that
– $Sim_{rule}(X,F) = 0$ means the identity of X and F.
– $Sim_{rule}(X,F) = 1$ means that X is to the same extent similar to F and F^C (i.e., values bigger than 1 mean a closer similarity of X and F^C to X and F).
– When $X = F^C$ (or $X^C = F$), i.e. $l_{IFS}(X,F^C) = l_{IFS}(X^C,F) = 0$ means the complete dissimilarity of X and F (or in other words, the identity of X and F^C), and then $Sim_{rule}(X,F) \to \infty$.
– When $X = F = F^C$ means the highest possible entropy (see [31]) for both elements F and X i.e. the highest "fuzziness" – not too constructive a case when looking for compatibility (both similarity and dissimilarity).
In other words, while applying the measure (32) to analyze the similarity of two objects, one should be interested in the values $0 \leq Sim_{rule}(X,F) < 1$.

The proposed measure (32) was constructed for selecting objects which are more similar than dissimilar [and well-defined in the sense of possessing (or not) attributes we are interested in]. In Szmidt and Kacprzyk [37] it was shown that a measure of similarity defined as above, (32), between $X(\mu_X, \nu_X, \pi_X)$ and $F(\mu_F, \nu_F, \pi_F)$ is more powerful than a simple distance between them. The following conclusion were drawn:

– when a distance between two (or more) objects/elements, or sets, is large, then it means for sure that the similarity does not occur.
– when a distance is small, we can say nothing for sure about similarity just on the basis of a distance between two objects [when not taking into account complements of the objects as in (32)]. The distance between objects can be small and the compared objects can be more dissimilar than similar.

We have shown on a simple example (cf. Szmidt and Kacprzyk [37]) that the measure (32) gives reasonable results when applied to assessing agreement in a group of experts. The only disadvantage of the proposed measure is that it does

ISBN: 978-989-95079-6-8

Table 1: Example results obtained from the similarity measures (35)–(38)

1	2	3	4	5
$X = (\mu, \nu, \pi)$	$(0.3, 0.4, 0.3)$	$(0.4, 0.2, 0.4)$	$(0.4, 0.2, 0.4)$	$(0, 0, 0)$
$F = (\mu, \nu, \pi)$	$(0.4, 0.3, 0.3)$	$(0.5, 0.3, 0.2)$	$(0.5, 0.2, 0.3)$	$(0.5, 0.5, 0)$
Sim_1	0	0.6	0.75	0.5
Sim_2	0	0.43	0.6	0.33
Sim_3	0	0.72	0.88	0.6
Sim_4	0	0.48	0.65	0.38

not follow the range of the usually assumed values for the similarity measures. But it is possible to construct a whole array of similarity measures following the philosophy, and preserving the advantages of the measure (32), and whose numerical values are consistent with the common scientific tradition (i.e. belonging to $[0, 1]$). For example:

$$Sim_1(X, F) = Sim_1(l_{IFS}(X, F), l_{IFS}(X, F^C)) = 1 - \frac{l_{IFS}(X, F)}{l_{IFS}(X, F) + l_{IFS}(X, F^C)} \quad (35)$$

$$Sim_2(X, F) = Sim_2(l_{IFS}(X, F), l_{IFS}(X, F^C)) = \frac{1 - f(l_{IFS}(X, F), l_{IFS}(X, F^C))}{1 + f(l_{IFS}(X, F), l_{IFS}(X, F^C))} \quad (36)$$

$$Sim_3(X, F) = Sim_3(l_{IFS}(X, F), l_{IFS}(X, F^C)) = \frac{1 - f(l_{IFS}(X, F), l_{IFS}(X, F^C))^2}{1 + f(l_{IFS}(X, F), l_{IFS}(X, F^C))^2} \quad (37)$$

$$Sim_4(X, F) = Sim_4(l_{IFS}(X, F), l_{IFS}(X, F^C)) = \frac{e^{-f(l_{IFS}(X,F), l_{IFS}(X,F^C))} - e^{-1}}{1 - e^{-1}} \quad (38)$$

where

$$f(l_{IFS}(X, F), l_{IFS}(X, F^C)) = = \frac{l_{IFS}(X, F)}{l_{IFS}(X, F) + l_{IFS}(X, F^C)} \quad (39)$$

and $0 \leq f(l_{IFS}(X, F), l_{IFS}(X, F^C)) \leq 1$.

The measures (35) – (38) give intuitive results. Some examples, being troublesome for other measures, are presented in Table 1. It is worth noticing that each measure assigns similarity equal 0 for an element $(0.3, 0.4, 0.3)$ and its complement $(0.4, 0.3, 0.3)$. In general, similarity measures (35) – (38) satisfy the following properties:

$$Sim_i(X, F) \in [0, 1], \quad (40)$$

$$Sim_i(X, X) = 1, \quad (41)$$

$$Sim_i(X, X^C) = 0, \quad (42)$$

$$Sim_i(X, F) = Sim_i(F, X), \quad (43)$$

for $i = 1, \ldots, 4$.

The similarity measures introduced in this section assess similarity of any two elements (X and F) belonging to an intuitionistic fuzzy set (or sets). The counterpart similarity measures for A-IFSs A and B containing n elements each, are:

$$Sim_k(A, B) = \frac{1}{n} \sum_{i=1}^{n} Sim_k(l_{IFS}(X_i, F_i), l_{IFS}(X_i, F_i^C)) \quad (44)$$

for $k = 1, \ldots, 4$.

5 Conclusions

We considered two groups of similarity measures between A-IFSs. First, we dealt with similarity measures constructed as if an A-IFSs was equal to a simple interval valued fuzzy set, or similarity measures being straightforward generalizations of those well known for the fuzzy sets. Unfortunately, in some situations, both approaches give counter-intuitive results. Second, we considered similarity measures accounting for all three functions describing an A-IF (the membership, non-membership, and hesitation margin) which is different from viewing an A-IF as a single interval. Next, we also took into account the complements of the elements compared. That is, we employed all kinds and fine shades of information available. It seems that these last measures are the most promising because, first of all, they help avoid some strong counter-intuitive results. This is crucial for both theory and applications.

References

[1] Atanassov K. (1983), Intuitionistic Fuzzy Sets. VII ITKR Session. Sofia (Centr. Sci.-Techn. Libr. of Bulg. Acad. of Sci., 1697/84) (in Bulgarian).

[2] Atanassov K. (1986) Intuitionistic Fuzzy Sets. Fuzzy Sets and Systems, 20, 87–96.

[3] Atanassov K. (1999), Intuitionistic Fuzzy Sets: Theory and Applications. Springer-Verlag.

[4] Atanassov K. and Gargov G. (1989), Interval-valued intuitionistic fuzzy sets. Fuzzy sets and Systems, 31 (3), 343–349.

[5] Bustince H., Burillo P. (1996) Vague sets are intuitionistic fuzzy sets. Fuzzy Sets and Systems 67, 403–405.

[6] Chen S. M. (1995) Measures of similarity between vague sets. Fuzzy Sets and Systems 74(2), 217–223.

[7] Chen S. M. (1997) Similarity measures between vague sets and between elements. IEEE Trans. Syst. Mn Cybernet. 27(1), 153–158.

[8] Cross V. and Sudkamp T. (2002) Similarity and Compatibility in Fuzzy Set Theory. Physica-Verlag.

[9] Dubois D. On degrees of truth, partial ignorance and contradiction. Magdalena L., Ojeda-Aciego M., Verdegay J.M. (eds): Proc. IPMU08, pp. 31–38.

[10] Gau W.L., Buehrer D.J. (1993) Vague sets. IEEE Trans. Systems Man Cybernet 23, 610–614.

[11] Hong D.H. and Kim C. (1999) A note on similarity measures between vague sets and between elements. Inform Science 115, 83–96.

ISBN: 978-989-95079-6-8

[12] Hung W-L. and Yang M-S. (2004) Similarity measures of intuitionistic fuzzy sets based on Hausdorff distance. Pattern recognition Letters 25, 1603–1611.

[13] Hung W.L and Yang M.S. (2008) On similarity measures between intuitionistic fuzzy sets. International Journal of Intelligent Systems 23 (3): 364–383.

[14] Kahneman D. (2002) Maps of bounded rationality: a perspective on intuitive judgment and choice. Nobel Prize Lecture.

[15] Li D.F, Cheng C.T. (2002) New similarity measures of intuitionistic fuzzy sets and application to pattern recognitions. Pattern Recognition Letters 23, 221–225.

[16] Li Y., Zhongxian C., Degin Y. (2002) Similarity measures between vague sets and vague entropy. J. Computer Sci. 29(12), 129–132.

[17] Liang Z. and Shi P. (2003) Similarity measures on intuitionistic fuzzy sets. Pattern Recognition Lett. 24, 2687–2693.

[18] Mitchell H.B. (2003) On the Dengfeng-Chuntian similarity measure and its application to pattern recognition. Pattern Recognition Lett. 24, 3101–3104.

[19] Montero J., Gómez D. and Bustince H. (2007): *On the relevance of some families of fuzzy sets*. Fuzzy Sets and Systems 158, 2429–2442.

[20] Pappis C. P. and Karacapilidis N. (1993) A comparative assessment of measures of similarity of fuzzy values. Fuzzy Sets and Systems 56, 171–174.

[21] Sutherland S. (1994) Irrationality. The Enemy Within. Penguin Books.

[22] Szmidt E. (2006) On a problem with the distances based on the Hausdorff metric. Report RB/52/2006, IBS PAN.

[23] Szmidt E. and Baldwin J. (2003) New Similarity Measure for Intuitionistic Fuzzy Set Theory and Mass Assignment Theory. Notes on IFSs, 9 (3), 60–76.

[24] Szmidt E. and Baldwin J. (2004) Entropy for Intuitionistic Fuzzy Set Theory and Mass Assignment Theory. Notes on IFSs, 10(3), 15-28.

[25] Szmidt E. and Kacprzyk J. (1996c) Remarks on some applications of intuitionistic fuzzy sets in decision making, Notes on IF, 2(3), 22–31.

[26] Szmidt E. and Kacprzyk J. (1997) On measuring distances between intuitionistic fuzzy sets, Notes on IF, 3(4), 1–13.

[27] Szmidt E. and Kacprzyk J. (1998a) Group Decision Making under Intuitionistic Fuzzy Preference Relations. IPMU'98, 172–178.

[28] Szmidt E. and Kacprzyk J. (1998b) Applications of Intuitionistic Fuzzy Sets in Decision Making. EUSFLAT'99, 150–158.

[29] Szmidt E. and Kacprzyk J. (2000) Distances between intuitionistic fuzzy sets, Fuzzy Sets and Systems, 114(3), 505–518.

[30] Szmidt E. and Kacprzyk J. (2000) On Measures on Consensus Under Intuitionistic Fuzzy Relations. IPMU 2000, 1454–1461.

[31] Szmidt E., Kacprzyk J. (2001) Entropy for intuitionistic fuzzy sets. Fuzzy Sets and Systems, 118 (3), 467–477.

[32] Szmidt E. and Kacprzyk J. (2001) Analysis of Consensus under Intuitionistic Fuzzy Preferences. Proc. Int. Conf. in Fuzzy Logic and Technology. Leicester, UK, 79–82.

[33] Szmidt E. and Kacprzyk J. (2002) An Intuitionistic Fuzzy Set Based Approach to Intelligent Data Analysis (an application to medical diagnosis). In A. Abraham, L.Jain, Kacprzyk (Eds.) Recent Advances in Intelligent Paradigms and Applications. Springer-Verlag, 57-70.

[34] Szmidt E. and Kacprzyk J. (2002) Analysis of Agreement in a Group of Experts via Distances Between Intuitionistic Fuzzy Preferences. IPMU 2002, Annecy, France, 1859–1865.

[35] Szmidt E. and J. Kacprzyk J. (2002b) An Intuitionistic Fuzzy Set Based Approach to Intelligent Data Analysis (an application to medical diagnosis). In A. Abraham, L. Jain, J. Kacprzyk (Eds.): Recent Advances in Intelligent Paradigms and Applications. Springer-Verlag, 57–70.

[36] Szmidt E. and Kacprzyk J. (2004) Similarity of intuitionistic fuzzy sets and the Jaccard coefficient. IPMU 2004, 1405–1412.

[37] Szmidt E., Kacprzyk J. (2004) A Concept of Similarity for Intuitionistic Fuzzy Sets and its use in Group Decision Making. 2004 IEEE Conf. on Fuzzy Systems, Budapest, 1129–1134.

[38] Szmidt E. and Kacprzyk J. (2006) An Application of Intuitionistic Fuzzy Set Similarity Measures to a Multi-criteria Decision Making Problem. ICAISC 2006, LNAI 4029, Springer-Verlag, 314–323.

[39] Szmidt E. and Kacprzyk J. (2006) Distances Between Intuitionistic Fuzzy Sets: Straightforward Approaches may not work. 3rd Int. IEEE Conf. Intelligent Systems IS06, 716–721.

[40] Szmidt E. and Kacprzyk J. (2007). Some problems with entropy measures for the Atanassov intuitionistic fuzzy sets. Applications of Fuzzy Sets Theory. LNAI 4578, 291–297. Springer-Verlag.

[41] Szmidt E. and Kacprzyk J. (2007a). A New Similarity Measure for Intuitionistic Fuzzy Sets: Straightforward Approaches may not work. 2007 IEEE Conf. on Fuzzy Systems, 481–486.

[42] Szmidt E. and Kukier M. (2006). Classification of Imbalanced and Overlapping Classes using Intuitionistic Fuzzy Sets. 3rd International IEEE Conference on Intelligent Systems IS'06, London, 2006, 722-727.

[43] Szmidt E. and Kukier M. (2008) A New Approach to Classification of Imbalanced Classes via Atanassov's Intuitionistic Fuzzy Sets. In: Hsiao-Fan Wang (Ed.): Intelligent Data Analysis : Developing New Methodologies Through Pattern Discovery and Recovery. Idea Group, 85–101.

[44] Tversky A. (1977). Features of similarity. Psychol. Rev., 84, 327–352.

[45] Veltkamp R.C. (2001) Shape matching: similarity measures and algorithms. Proc. Shape Modelling International, Genova, Italy, IEEE Press, 188–197.

[46] Wang W.J. (1997) New similarity measures on fuzzy sets and on elements. Fuzzy Sets and Systems, 85, 305–309.

[47] Wang X., De Baets B., and Kerre E. (1995). A comparative study of similarity measures. Fuzzy Sets and Systems, 73 (2), 259–268.

[48] Zadeh L.A. (1965). Fuzzy sets. Information and Control, 8, 338–353.

ISBN: 978-989-95079-6-8

Non-commutative EQ-logics and their extensions

Vilém Novák, Martin Dyba

Institute for Research and Applications of Fuzzy Modelling, University of Ostrava
Ostrava, Czech Republic
Email: Vilem.Novak,martin.dyba@osu.cz

Abstract— *We discuss a formal many-valued logic called EQ-logic which is based on a recently introduced special class of algebras called EQ-algebras. The latter have three basic binary operations (meet, multiplication, fuzzy equality) and a top element and, in a certain sense, generalize residuated lattices. The goal of EQ-logics is to present a possible direction in the development of mathematical logics in which axioms are formed as identities. In this paper we propose a basic EQ-logic and three extensions which end up with a logic equivalent to the MTL-logic.*

Keywords— EQ-algebra, fuzzy equality, residuated lattice, MTL-logic, fuzzy logic.

1 Introduction

One of possible directions of the development of mathematical logic which can be tracked down to Leibnitz, Wittgenstein, and Ramsey (cf. [1]) is to develop it on the basis of identity (equality) as the principal connective. This direction has been crowned in a noble way by Henkin in [2] who developed the type theory (a higher order logic) using identity as a sole primitive constant. This is nice because it enables to treat equality between elements on the basis of the same principles not depending on their character. Thus, equality between truth values is, in principle, a relation of the same kind as an equality between other kinds of elements. Hence, we will speak about equality instead of equivalence even when speaking about truth values.

Note that there are works developing classical boolean logic on the basis of equivalence as the main (not sole) connective (a recent book presenting logic in this way is [3]). Moreover, the logic is developed there in the "equational style", i.e. proofs proceed as sequences of equations (in fact, equivalences).

This gives rise to an idea whether also fuzzy (many-valued) logic could be developed on the basis of fuzzy equality as the main connective where by fuzzy equality we mean, in fact, a fuzzy equivalence, i.e. a generalization of the classical logical equivalence. We will prefer the term "fuzzy equality", though. Recall that this idea is supported by the mentioned Henkin successful type theory (for a detailed development of this theory, see also [4]) as well as by the development of fuzzy type theory (see [5, 6]). On the other hand, unlike classical type theory, it seems impossible to have fuzzy equality a sole connective. At least conjunction is also necessary as follows from the considerations concerning the structure of the algebra of truth values. Namely, two approaches are apparent.

In the first approach, the structure of truth values is assumed to form a residuated lattice where the basic operations are, besides the lattice ones, the multiplication "$\otimes$" and its residuum "$\rightarrow$". The latter is in fuzzy logic a natural interpretation of implication which is a primary operation while the equivalence is interpreted by a biresiduation $a \leftrightarrow b = (a \rightarrow b) \wedge (b \rightarrow a)$ which is a derived operation. Since the basic connective, in our case, should be the fuzzy equality, it seems unnatural to interpret it by a derived operation.

The second approach follows an idea presented in [7, 8] where a specific kind of algebra called EQ-algebra is developed. Unlike the residuated lattices, the basic operation in EQ-algebras is a fuzzy equality "$\sim$" while implication is derived from it. Of course, a crucial question remains, which properties such an equality should have. At first sight it seems clear that these properties should be reflexivity, symmetry, and transitivity. We are convinced, however, that the fuzzy equality should be characterized by a deeper property reflecting some basic structure so that both symmetry as well as especially transitivity would follow. The most essential is, in our opinion, the "substitution principle" stating that if we replace an object by another one equal to the former then the result is not changed; in the case of fuzzy equality, the replacement should not make the structure in concern "worse" in some reasonable sense.

What basic structure should we consider in fuzzy logic? Clearly, this is *ordering* since the basic idea standing behind fuzziness is introducing degrees to enable measuring imprecision stemming from vagueness. Since ordering is in no way included in the equality, it must be introduced explicitly. Therefore, we will consider the basic structure to be a $\wedge$-semilattice with the greatest element $\mathbf{1}$. At the same time, we must "fuse" in some way statements about ordering and equality. The fusion is in our algebra realized by a third operation of multiplication "$\otimes$" whose properties, however, should be as weak as possible because we think that it should not participate on the ordering but enable only to "fuse things together". This is the main motivation behind axiom (E4) below. The symmetry and transitivity of $\sim$ are then consequences of it. The fusion $\otimes$ needs not be commutative and, perhaps, nor even associative.

Step by step we arrive at 8 axioms which determine the EQ-algebra. In its "crude form", there is no need to introduce the smallest element $\mathbf{0}$. Morever, the axioms make it possible to have two different truth values equivalent in the degree $\mathbf{1}$ and so, the fuzzy equality generalizes classical equivalence in an essential way. From the algebraic point of view, EQ-algebras generalize, in a certain sense, residuated lattices and so, they look as interesting class of algebraic structures.

When attempting to develop a formal logical calculus it came out, surprisingly, that the crude structure though reflecting the basic principles of fuzzy equality, is not strong enough. Namely, the logic enforced adding an axiom of "goodness" (E10) (or, as a logical formula (EQ1)) below which says that

ISBN: 978-989-95079-6-8

an element a is equal to $\mathbf{1}$ always in a degree a. This axiom implies that the algebra is separated, i.e. that two elements equal in the degree $\mathbf{1}$ must be identical in the classical sense. We conclude that the logic allows only a slight generalization of the classical equivalence. Of course, the goodness axiom has many important algebraic consequences (see [9, 8]). Among them is also the fact that each good EQ-algebra gives rise to a BCK-algebra. Putting all together we expect that EQ-algebras may put a different light on residuated lattices and on their general role in algebra and logic.

This paper is an attempt to introduce a formal logic called EQ-logic*) whose truth values are taken from a good EQ-algebra. Several basic properties of this logic demonstrating that it behaves reasonably are proved. In comparison with residuated fuzzy logics, however, our logic seems to be more limited. For example, the deduction theorem even in its weaker form probably does not hold and we guess that its validity is equivalent with residuation. Still, it is interesting to learn how far can we go with the development of EQ-logic. Besides others, it enables us to separate properties, which stem from the fuzzy equality itself, from the properties for which residuation is necessary. It may also shed light on classical logic, especially on its equational variant ([3]).

2 Definition and fundamental properties of EQ-algebras

Definition 1
A semicopula-based EQ-algebra $\mathscr{E}$ is an algebra of type (2, 2, 2, 0), i.e.

$$\mathscr{E} = \langle E, \wedge, \otimes, \sim, \mathbf{1} \rangle\,, \tag{1}$$

where for all $a, b, c, d \in E$:

(E1) $\langle E, \wedge, \mathbf{1}\rangle$ *is a commutative idempotent monoid (i.e. $\wedge$-semilattice with top element $\mathbf{1}$) where the ordering $a \le b$ is defined in standard way as $a \wedge b = a$,*

(E2) $\otimes$ *is a binary multiplication operation isotone w.r.t. $\le$, i.e. $a \le b$ implies both $a \otimes c \le b \otimes c$ as well as $c \otimes a \le c \otimes b$ and $\mathbf{1}$ is a neutral element, i.e. $a \otimes \mathbf{1} = \mathbf{1} \otimes a = a$.*

(E3) $a \sim a = \mathbf{1}$,

(E4) $((a \wedge b) \sim c) \otimes (d \sim a) \le c \sim (d \wedge b)$,

(E5) $(a \sim b) \otimes (c \sim d) \le (a \sim c) \sim (b \sim d)$,

(E6) $(a \wedge b \wedge c) \sim a \le (a \wedge b) \sim a$,

(E7) $(a \wedge b) \sim a \le (a \wedge b \wedge c) \sim (a \wedge c)$,

(E8) $a \otimes b \le a \sim b$.

The operation $\sim$ is called a fuzzy equality.

For all $a, b \in E$ we put

$$a \to b = (a \wedge b) \sim a. \tag{2}$$

This is a derived operation called *implication*†).

Note that axioms (E6) and (E7) can be rewritten as

$$a \to (b \wedge c) \le a \to b\,, \tag{E6'}$$
$$a \to b \le (a \wedge c) \to b\,. \tag{E7'}$$

The following theorem confirms that $\sim$ is indeed a fuzzy equality.

Theorem 1
Let $\mathscr{E}$ be an EQ-algebra. The following holds for all $a, b, c \in E$:

(a) Symmetry: $a \sim b = b \sim a$,

(b) Transitivity: $(a \sim b) \otimes (b \sim c) \le a \sim c$,

(c) Transitivity of implication: $(a \to b) \otimes (b \to c) \le a \to c$.

We say that the multiplication $\otimes$ is *$\to$-isotone* if

$$a \to b = \mathbf{1} \quad \text{implies} \quad a \otimes c \to b \otimes c = \mathbf{1} \tag{3}$$

for all $a, b, c \in E$. Let us remark that the $\to$-isotonicity of $\otimes$ is usually fulfilled. There exist EQ-algebras, however, which have not this property.

A semicopula-based EQ-algebra has the operation $\otimes$, in general neither commutative nor associative. If the latter is only non-commutative but associative then we will speak about a *non-commutative* EQ-algebra.

Theorem 2
Let $\mathscr{E} = \langle E, \wedge, \otimes, \sim, \mathbf{1}\rangle$ is a semicopula-based EQ-algebra and put $b \bar{\otimes} a = a \otimes b$. Then $\mathscr{E}' = \langle E, \wedge, \bar{\otimes}, \sim, \mathbf{1}\rangle$ is also a semicopula-based EQ-algebra.

The following lemmas summarizes some of the properties of EQ-algebras useful below. Its proof can be found in [9, 8].

Lemma 1
Let $\mathscr{E}$ be an EQ-algebra. For all $a, b \in E$ such that $a \le b$ it holds that

(a) $a \to b = \mathbf{1}$,

(b) $a \sim b = b \to a$,

(c) $c \to a \le c \to b$ *and* $b \to c \le a \to c$.

Lemma 2
Let $\mathscr{E}$ be an EQ-algebra. For all $a, b, c, d, a', b', c', d' \in E$ it holds that

(a) $a \otimes b \le a, a \otimes b \le a \wedge b, c \otimes (a \wedge b) \le (c \otimes a) \wedge (c \otimes b)$,

(b) $a \sim b \le a \to b, a \to a = \mathbf{1}$,

(c) $(a \to b) \otimes (b \to a) \le a \sim b$,

(d) $a = b$ *iff* $a \sim b = \mathbf{1}$,

(e) $a = \mathbf{1} \to a$ *and* $a \to \mathbf{1} = \mathbf{1}$,

(f) $a \otimes (a \sim b) \le b$,

(g) $b \le a \to b$,

(h) $((a \wedge b) \sim (c \wedge d)) \otimes (a \sim a') \otimes (b \sim b') \otimes (c \sim c') \otimes (d \sim d') \le (a' \wedge b') \sim (c' \wedge d')$.

*) This logic has been first introduced in Barcelona, October 2008 at the occasion of Prof. F. Esteva 65th birthday.

†) Note that this operation has been, in fact, introduced already by G. W. Leibnitz (cf. [10])

ISBN: 978-989-95079-6-8

Note that if $\mathscr{E}$ is separated, then $a \to b = \mathbf{1}$ implies $a \leq b$. Indeed, $a \to b = (a \wedge b) \sim a = \mathbf{1}$ implies $a \wedge b = a$, i.e. $a \leq b$.

Let $\mathscr{E}$ contain also the bottom element $\mathbf{0}$. Then we put

$$\neg a = a \sim \mathbf{0}, \qquad a \in E \tag{4}$$

and call $\neg a$ a *negation* of $a \in E$.

Definition 2
Let $\mathscr{E}$ be an EQ-algebra and $a, b, c, d \in L$. We say that $\mathscr{E}$ is:

(i) separated *if for all $a \in E$,*

(E9) $a \sim b = \mathbf{1}$ *implies* $a = b$.

(ii) good *if*

(E10) $a \sim \mathbf{1} = a$.

(iii) residuated *if for all $a, b, c \in E$,*

(E10) $(a \otimes b) \wedge c = a \otimes b$ *iff* $a \wedge ((b \wedge c) \sim b) = a$.

(iv) involutive *(IEQ-algebra) if for all $a \in E$,*

(E11) $\neg\neg a = a$.

(v) prelinear *if for all $a, b \in E$,*

(E12) $\sup\{a \to b, b \to a\} = \mathbf{1}$.

(vi) lattice EQ-algebra *(ℓEQ-algebra) if it is a lattice and*

(E13) $((a \vee b) \sim c) \otimes (d \sim a) \leq (d \vee b) \sim c$.

(vii) complete *if it's $\wedge$-semilattice reduct is complete.*

(viii) EQ(R)-algebra *if for all $a, b \in E$ there exist*

$$a/b = \max\{c \in E \mid a \otimes c \leq b\}$$
$$a\backslash b = \max\{c \in E \mid c \otimes a \leq b\}$$

(the latter equalities will be referred to as (R)-condition).

It is easy to show that good EQ-algebras are *separated*. Obviously, $\otimes$ in good EQ-algebras is $\to$-isotone. As already discussed, the "goodness property" appeared to be indispensable for logic and so, from now on, *by EQ-algebra we will always mean a good one*.

Remark 1
Let $\mathscr{E}$ be a semicopula-based EQ-algebra. A set $F \subseteq E$ is a filter if $\mathbf{1} \in F$; $a, b \in F$ implies $a \otimes b \in F$; $a, a \to b \in F$ implies $b \in F$; $a \to b \in F$ implies $(a \otimes c) \to (b \otimes c) \in F$ and $(c \otimes a) \to (c \otimes b) \in F$. Then it can be proved that a factor algebra $\mathscr{E}|F$ is a separated semicopula-based EQ-algebra. Surprizingly, there exist EQ-algebras with no proper filter and so, we cannot embed each EQ-algebra into a separated one.

The concept of EQ(R)-algebra is due to El Zekey [11]. Obviously, it inserts residuation inside EQ-algebra. In case that $E = [0, 1]$, it simply requires the operation $\otimes$ to be left-continuous.

Lemma 3
Let $\mathscr{E}$ be an IEQ-algebra. Then

(a) $\mathscr{E}$ is good ℓEQ-algebra with join defined by

$$a \vee b = \neg(\neg a \wedge \neg b).$$

(b) $a \leq b$ *iff* $\neg b \leq \neg a$.

(c) $a \sim b = \neg a \sim \neg b$.

(d) $a \otimes \neg a = \mathbf{0}$.

Theorem 3 ([11])
Let $\mathscr{E}$ be a good non-commutative EQ(R)-algebra. Then the following is equivalent:

(a) $\mathscr{E}$ is representable (i.e. subdirectly embeddable into a product of linearly ordered good EQ(R)-algebras[‡]).

(b) $\mathscr{E}$ satisfies the formula

$$(a \to b) \vee (d \to (d \otimes (c \to (b \to a) \otimes c))) = \mathbf{1}$$

for all $a, b, c, d \in E$.

3 Basic EQ-logic

In this section we introduce a propositional EQ-logic which we will call basic. This logic seems to be the simplest logic definable on the basis of EQ-algebras. Of course, we obtain more sophisticated logics when adding further axioms. From this point of view, all core (residuated) fuzzy logic in the sense of [12]) are extensions of the basic EQ-logic.

Definition 3
The language of EQ-logic consists of propositional variables $p_1, p_2, \ldots$, binary connectives $\wedge, \&, \equiv$ and a truth (logical) constant $\top$. Formulas are defined in the obvious way: each propositional variable is a formula, $\top$ is a formula and if A, B are formulas, then $A \wedge B, A \& B, A \equiv B$ are formulas. Implication is defined as a short

$$A \Rightarrow B := (A \wedge B) \equiv A. \tag{5}$$

Let J be a language of EQ-logic, F_J a set of all formulas in the language J and $\mathscr{E} = \langle E, \wedge, \otimes, \sim, \mathbf{1}\rangle$ be a good EQ-algebra. A truth evaluation $e : F_J \longrightarrow E$ is defined as usual: if $p \in J$ is a propositional variable then $e(p) \in E$. Furthermore, we put

$$e(\top) = \mathbf{1},$$
$$e(A \wedge B) = e(A) \wedge e(B)$$
$$e(A \& B) = (e(A) \otimes e(B)),$$
$$e(A \equiv B) = (e(A) \sim e(B))$$

for all formulas $A, B \in F_J$.

A formula $A \in F_J$ is a tautology if $e(A) = \mathbf{1}$ for each truth evaluation $e : F_J \longrightarrow E$.

Note that our logic has a non-commutative fusion connective $\&$ but only one implication because it is derived from equivalence (i.e. its interpretation is not residual operation adjoint with multiplication). We are convinced that this is an advantage.

[‡] A recent result of El-Zekey (personal communication) implies that the (R)-condition is unnecessary for the representability. This may lead to simplification of the structure of EQ-logics discussed below.

ISBN: 978-989-95079-6-8

3.1 Logical axioms and inference rules

The following formulas are axioms of the EQ-logic:

(EQ1) $(A \equiv \top) \equiv A$

(EQ2) $A \wedge B \equiv B \wedge A$

(EQ3) $(A \wedge B) \wedge C \equiv A \wedge (B \wedge C)$

(EQ4) $A \wedge A \equiv A$

(EQ5) $A \wedge \top \equiv A$

(EQ6) $A \,\&\, \top \equiv A$

(EQ7) $\top \,\&\, A \equiv A$

(EQ8a) $((A \wedge B) \,\&\, C) \Rightarrow (B \,\&\, C)$

(EQ8b) $(C \,\&\, (A \wedge B)) \Rightarrow (C \,\&\, B)$

(EQ9) $((A \wedge B) \equiv C) \,\&\, (D \equiv A) \Rightarrow (C \equiv (D \wedge B))$

(EQ10) $(A \equiv B) \,\&\, (C \equiv D) \Rightarrow (A \equiv C) \equiv (D \equiv B)$

(EQ11) $(A \Rightarrow (B \wedge C)) \Rightarrow (A \Rightarrow B)$

(EQ12) $(A \Rightarrow B) \Rightarrow (A \wedge C) \Rightarrow B$

Remark 2
The above axioms have been written using $\Rightarrow$ for better readability because formulas with this connective are more usual than those with $\equiv$. However, technically we apply equational-style proofs in the sense of [3], which are sequences of formulas of the form $A_1 \equiv A_2, \ldots, A_{n-1} \equiv A_n$ such that each of the individual theorems $A_i \equiv A_{i+1}$ has an independent individual proof. Therefore, a fussy presentation would require to rewrite all the axioms using $\equiv$ and the definition (5) only.

By $A[p := B]$ we denote a formula resulting from A by replacing all occurrences of a propositional variable p in A by the formula B. Then we introduce the following inference rules of EQ-logic:

$$(EA)\quad \frac{A, A \equiv B}{B} \qquad (L)\quad \frac{B \equiv C}{A[p := B] \equiv A[p := C]}$$

The rule (EA) is the *equanimity rule* and (L) is the *Leibnitz rule* for formulas (cf. [3] and elsewhere). The notion of provability is classical. A formal theory is any subset $T \subseteq F_J$. As usual, we suppose that T is defined by a set of special axioms.

3.2 Main properties

The proof of the following three lemmas is technical. They demonstrate several reasonable properties of the basic EQ-logic.

Lemma 4
The following are special derived rules:

(a) $A \equiv \top \vdash A$. *(rule (T1))*

(b) $A \vdash A \equiv \top$. *(rule (T2))*

(c) $A \wedge D \equiv C, A \equiv B \vdash B \wedge D \equiv C$. *(rule (C))*

(d) $(A \equiv D) \equiv C, A \equiv B \vdash (B \equiv D) \equiv C$. *(rule (E))*

(e) $A \,\&\, D \equiv C, A \equiv B \vdash B \,\&\, D \equiv C$. *(rule (F1))*

(f) $D \,\&\, A \equiv C, A \equiv B \vdash D \,\&\, B \equiv C$. *(rule (F2))*

Lemma 5
(a) $\vdash A \equiv A$,

(b) $A \equiv B \vdash B \equiv A$,

(c) $A \equiv B, B \equiv C \vdash A \equiv C$,

(d) $A, A \Rightarrow B \vdash B$, *(Modus Ponens)*

(e) $\vdash (\top \Rightarrow A) \equiv A$,

(f) $A \Rightarrow B, B \Rightarrow C \vdash A \Rightarrow C$,

(g) $A, B \vdash A \,\&\, B$,

(h) $\vdash (A \equiv B) \equiv (B \equiv A)$,

(i) $A \Rightarrow (B \equiv C), B \vdash A \Rightarrow C$,

(j) $\vdash (A \equiv D) \Rightarrow ((A \equiv B) \equiv (D \equiv B))$,

(k) $A \Rightarrow (B \equiv C), B \equiv D \vdash A \Rightarrow (D \equiv C)$,

(l) $A \Rightarrow (B \equiv C), C \equiv D \vdash A \Rightarrow (B \equiv D)$,

(m) $\vdash (A \equiv B) \Rightarrow (A \Rightarrow B)$,

(n) $A \equiv B, C \equiv D \vdash A \,\&\, C \equiv B \,\&\, D$,

(o) $A \Rightarrow B, C \Rightarrow D \vdash A \,\&\, C \Rightarrow B \,\&\, D$,

(p) $\vdash ((A \equiv B) \equiv C) \,\&\, (A \equiv D) \Rightarrow ((D \equiv B) \equiv C)$,

(q) $\vdash (A \equiv B) \,\&\, (C \equiv D) \Rightarrow (A \equiv C) \equiv (B \equiv D)$,

(r) $\vdash (A \Rightarrow B) \,\&\, (B \Rightarrow A) \Rightarrow (A \equiv B)$,

(s) $\vdash B \Rightarrow (A \Rightarrow B)$.

Lemma 6
(a) $\vdash (A \equiv B) \,\&\, (B \equiv C) \Rightarrow (A \equiv C)$,

(b) $\vdash (A \,\&\, (A \equiv B)) \Rightarrow B$

(c) $\vdash (A \Rightarrow B) \,\&\, (B \Rightarrow C) \Rightarrow (A \Rightarrow C)$,

(d) $\vdash (A \,\&\, (A \Rightarrow B)) \Rightarrow B$

(e) $\vdash B \Rightarrow (B \equiv \top)$,

(f) $A \Rightarrow (B \Rightarrow C) \vdash (A \,\&\, B) \Rightarrow C$,

(g) $\vdash A \,\&\, B \Rightarrow A \equiv B$,

(h) $\vdash (A \wedge B) \Rightarrow A$,

(i) $\vdash (C \Rightarrow A) \,\&\, (C \Rightarrow B) \Rightarrow (C \Rightarrow (A \wedge B))$.

By the straightforward verification we can prove the following two lemmas.

Lemma 7
All axioms of EQ-logic are tautologies.

Lemma 8
The deductive rules of EQ-logic are sound in the following sense. Let $e : F_J \longrightarrow E$ be a truth evaluation:

ISBN: 978-989-95079-6-8

(a) *If* $e(A) = \mathbf{1}$ *and* $e(A \equiv B) = \mathbf{1}$ *then* $e(B) = \mathbf{1}$.

(b) *If* $e(B \equiv C) = \mathbf{1}$ *then* $e(A[p := B] \equiv A[p := C]) = \mathbf{1}$ *for any formula* A.

The following is a standard Lindenbaum-Tarski technique. Put

$$A \approx B \quad \text{iff} \quad \vdash A \equiv B, \quad A, B \in F_J.$$

It follows from Lemmas 5(a), (h) and 6(a) that $\approx$ is an equivalence on F_J. Let us denote by $[A]$ an equivalence class of A and put $\bar{E} = \{[A] \mid A \in F_J\}$. If we define

$$\begin{aligned} \mathbf{1} &= [\top], \\ [A] \wedge [B] &= [A \wedge B], \\ [A] \otimes [B] &= [A \,\&\, B], \\ [A] \sim [B] &= [A \equiv B], \end{aligned}$$

then we obtain an algebra $\bar{\mathscr{E}} = \langle \bar{E}, \wedge, \otimes, \sim, \mathbf{1} \rangle$.

Lemma 9
The algebra $\bar{\mathscr{E}}$ *is a good semicopula-based EQ-algebra.*

Theorem 4 (Soundness)
The basic EQ-fuzzy logic is sound.

PROOF: This is a consequence of Lemmas 7 and 8. □

Theorem 5 (Completeness)
The following is equivalent for every formula A*:*

(a) $\vdash A$,

(b) $e(A) = \mathbf{1}$ *for every good semicopula-based EQ-algebra* $\mathscr{E}$ *and a truth evaluation* $e : F_J \longrightarrow E$.

PROOF: The implication (a) to (b) is soundness.

(b) to (a): By Lemma 9 the algebra $\bar{\mathscr{E}}$ of equivalence classes of formulas is a good semicopula-based EQ-algebra. Thus, if (b) holds then it holds also for $e : F_J \longrightarrow \bar{E}$. If $e(A) = \mathbf{1}$ then it means that $[A] = [\top]$, i.e. $\vdash A \equiv \top$ and so, $\vdash A$ by rule (T2). □

4 Extensions of the basic EQ-logic

In this section we will briefly discuss some extensions of the basic EQ-logic.

4.1 Involutive EQ-logic

This logic (we will speak about IEQ-logic) contains falsity and is characterized by the property of double negation which leads to some simplifications.

The language J is the same as that of basic EQ-logic with the exception that $\top$ is replaced by $\bot$. Furthermore, we introduce the following shorts of formulas:

$$\top := \bot \equiv \bot, \tag{6}$$

$$\neg A := A \equiv \bot, \tag{7}$$

$$A \vee B := \neg(\neg A \wedge \neg B). \tag{8}$$

(clearly, (7) is a *negation* and (8) a disjunction.

Axioms of IEQ-logic are (EQ2)–(EQ12) plus the following ones:

(EQ13) $(A \,\&\, B) \,\&\, C \equiv A \,\&\, (B \,\&\, C)$,

(EQ14) $(A \wedge \bot) \equiv \bot$,

(EQ15) $\neg\neg A \equiv A$.

Axiom (EQ14) characterizes the basic property of $\bot$ and it can be written as $\bot \Rightarrow A$ (ex falso quodlibet). Axiom (EQ1) is in this logic provable.

A contradiction is a formula $A \,\&\, \neg A$. Then we say that a theory T is *contradictory* if $T \vdash A \,\&\, \neg A$ for some formula $A \in F_J$.

A straightforward proof will give the following (classical) theorem:

Theorem 6
A theory T *is contradictory iff* $T \vdash A$ *for all* $A \in F_{J(T)}$.

Semantics of this logic is formed by IEQ-algebras.

Theorem 7 (Completeness)
The following is equivalent for every formula A*:*

(a) $\vdash A$,

(b) $e(A) = \mathbf{1}$ *for every IEQ-algebra* $\mathscr{E}$ *and a truth evaluation* $e : F_J \longrightarrow E$.

4.2 EQ(R)-fuzzy logic

This logic seems to be closest to the residuated fuzzy logics. In fact, it already introduces some kind of residuated structure which enables to prove a stronger variant of the completeness theorem.

The language of this logic is that of IEQ-logic extended by two binary connectives: "\" and "/" and the shorts (6) and

$$A \vee B := ((A \Rightarrow B) \Rightarrow B) \wedge ((B \Rightarrow A) \Rightarrow A).$$

Axioms are (EQ1)–(EQ14) plus the following:

(EQ16) $(A \Rightarrow B) \vee (D \Rightarrow (D \,\&\, (C \Rightarrow (B \Rightarrow A) \,\&\, C)))$

(EQ17) $C/(A \otimes B) \equiv (C/(B/A))$

(EQ18) $(A \otimes B)\backslash C \equiv ((A\backslash B)\backslash C)$

Semantics of this logic is formed by good non-commutative prelinear EQ(R)-algebras.

Using Theorem 3, the following can be proved:

Theorem 8 (Completeness)
For every formula $A \in F_J$ *the following is equivalent:*

(a) $\vdash A$.

(b) $e(A) = \mathbf{1}$ *for every truth evaluation* $e : F_J \longrightarrow E$ *and every linearly ordered good non-commutative EQ(R)-algebra* $\mathscr{E}$.

(c) $e(A) = \mathbf{1}$ *for every truth evaluation* $e : F_J \longrightarrow E$ *and every good non-commutative prelinear EQ(R)-algebra* $\mathscr{E}$.

ISBN: 978-989-95079-6-8

4.3 Residuated EQ-logic

This is stronger logic close to MTL-logic because of introducing the residuation property.

The language J is the same as that of basic EQ-logic. Its axioms are (EQ1)–(EQ12) plus the following:

(EQ19) $((A \,\&\, B) \equiv C) \,\&\, (D \equiv A) \Rightarrow (C \equiv (D \,\&\, B))$

Lemma 10
The following is provable residuated EQ-logic:

(a) $\vdash A \,\&\, B \equiv B \,\&\, A$,

(b) $(A \,\&\, B) \,\&\, C \equiv A \,\&\, (B \,\&\, C)$,

(c) $\vdash ((A \,\&\, B) \Rightarrow C) \equiv (A \Rightarrow (B \Rightarrow C))$.

By A^n we mean n-times repetition of A w.r.t. $\&$ (as usual).

Theorem 9 (Deduction)
For every theory T and formulas $A, B \in F_J$:

$T \cup \{A\} \vdash B$ *iff* *there is* $n \geq 1$ *such that* $T \vdash A^n \Rightarrow B$

Semantics for this logic is formed by residuated EQ-algebras. If we replace the constant $\top$ in the language J by $\bot$, take the definition (6) and add axiom (EQ14) and the axiom

(EQ20) $((A \Rightarrow B) \Rightarrow C) \Rightarrow (((B \Rightarrow A) \Rightarrow C) \Rightarrow C)$

then the resulting logic is equivalent with the MTL-logic.

5 Conclusion

In this paper we introduced a class of logics based on a new algebra of truth values called EQ-algebra. This algebra is specific by introducing fuzzy equality (equivalence) as the basic operation. Accordingly, the basic connective of EQ-logic is equivalence.

It is easy to show that BCK-axioms are provable in basic EQ-logic and so, basic EQ-logic is also a BCK-one (cf. [13]). However, since modus-ponens is a derived rule in the former, translation of properties of BCK-logic to EQ-logic is not straightforward. It is an open question whether there is a formal system equivalent with basic EQ-logic with modus ponens as its sole rule.

Acknowledgment

The research was supported by the project MSM 6198898701 of the MŠMT ČR.

References

[1] F.P. Ramsey. The foundations of mathematics. *Proceedings of the London Mathematical Society*, 25:338–384, 1926.

[2] L. Henkin. A theory of propositional types. *Fundamenta Math.*, 52:323–344, 1963.

[3] G. Tourlakis. *Mathematical Logic*. New York, J. Wiley & Sons, 2008.

[4] P. Andrews. *An Introduction to Mathematical Logic and Type Theory: To Truth Through Proof*. Kluwer, Dordrecht, 2002.

[5] V. Novák. On fuzzy type theory. *Fuzzy Sets and Systems*, 149:235–273, 2005.

[6] V. Novák. EQ-algebra-based fuzzy type theory and its extensions. *Logic Journal of the IGPL*, page (submitted).

[7] V. Novák. EQ-algebras in progress. In O. Castillo, editor, *Theoretical Advances and Applications of Fuzzy Logic and Soft Computing*, pages 876–884. Springer, Berlin, 2007.

[8] V. Novák and B. de Baets. EQ-algebras. *Fuzzy Sets and Systems*, page (to appear), 2009.

[9] M. El-Zekey, V. Novák, and R. Mesiar. Semicopula-based EQ-algebras. *Fuzzy Sets and Systems*, page (submitted).

[10] H.N. Castañeda. Leibniz's syllogistico-propositional calculus. *Notre Dame Journal of Formal Logic*, XVII(4):338–384, 1976.

[11] M. El-Zekey. Representable EQ(R)-algebras. page (submitted).

[12] P. Hájek and P. Cintula. On theories and models in fuzzy predicate logics. *Journal of Symbolic Logic*, 71(3):863—-880, 2006.

[13] P. Cintula, P. Hájek, and R. Horčík. Formal systems of fuzzy logic and their fragments. *Annals of Pure and Applied Logic*, 150:40–65, 2007.

ISBN: 978-989-95079-6-8

Discriminant Analysis for fuzzy random variables based on nonparametric regression

Ana Colubi, M.A. Gil, [1] Gil González-Rodríguez, Wolfgang Trutschnig [2]

1.Department of Statistics and OR, University of Oviedo
Oviedo, Spain
2.Unit of Intelligent Data Analysis and Graphical Models, European Centre for Soft Computing
Mieres, Spain
Email: {colubi, magil}@uniovi.es, {gil.gonzalez, wolfgang.trutschnig}@softcomputing.es

Abstract*— *This communication is concerned with the problem of supervised classification of fuzzy data obtained from a random experiment. The data generation process is modelled through fuzzy random variables which, from a formal point of view, can be identified with a kind of functional random element. We propose to adapt one of the most versatile discriminant approaches in the context of functional data analysis to the specific case we handle. The discriminant analysis is based on the kernel estimation of the nonparametric regression. The results are applied to an experiment concerning fuzzy perceptions and linguistic labels

***Keywords*—** fuzzy data, random experiments, supervised classification, kernel estimation, nonparametric regression.

1 Introduction

In many random experiments, as for instance sociological surveys, ecological studies, etc., some characteristics of interest can be assessed in a more meaningful scale if the respondents or the experts are allowed to indicate the degree of precision/imprecision of their answers/judgment/perceptions by means of fuzzy sets (see, for instance, [3] and Section 4 for more details). The results of those experiments can be soundly modelled by means of fuzzy random variables in Puri & Ralescu's sense [16]. It should be remarked that fuzzy random variables in this sense are used to model fuzzy-set-valued random elements and that our goal regards the fuzzy random variables 'per-se'. However, there are other experiments in which the quantity of interest is an underlying real-valued random variables that cannot be precisely observed. In this cases, different approaches may be considered (see, for instance, [5]).

From a formal point of view fuzzy random variables can be identified with a special case of functional random variables, although with some particular features concerning the natural arithmetic and metric structure (see [7] for a deeper discussion). Functional Data Analysis has become an important area of research in recent years (see, for instance, [12, 17, 8, 19]), and as it was suggested in [7], it is possible to take advantage of some of the results developed for those elements to analyze fuzzy data.

As a first step in classification problems concerning fuzzy random variables, some unsupervised approaches have been considered in the literature (see, for instance, [10]). In this communication we deal with the supervised classification problem. That is, given a set of possible groups and a training random sample of fuzzy data of each group, the goal is to predict the group membership of a new fuzzy datum.

As it is indicated in [6], there are different approaches to this problem in the functional context. Most of them are based on modifications of the linear discriminant analysis, however to avoid the inconveniences that frequently arises from the presence of nonlinear class boundaries, we propose a nonparametric method inspired by [6].

The rest of the paper is organized as follows. In Section 2 we introduce the notation and the basic concepts. In Section 3 the classification approach is discussed. Section 4 is devoted to the empirical results, and finally in Section 5 we conclude with some remarks and open problems.

2 Preliminaries

We will denote by $\mathcal{F}_c(\mathbb{R}^p)$ the *class of fuzzy sets* $U : \mathbb{R}^p \to [0,1]$ whose α-levels U_α are nonempty compact convex subsets of $\mathbb{R}^p$ for all $\alpha \in [0,1]$, where $U_\alpha = \{x \in \mathbb{R}^p \,|\, U(x) \geq \alpha\}$ for all $\alpha \in (0,1]$, and $U_0 = \mathrm{cl}(\{x \in \mathbb{R}^p \,|\, U(x) > 0\})$.

Recently, a new class of metrics based on the generalization of the mid-point and the spread of an interval has been defined. This class is very intuitive and versatile and has good properties for the statistical analysis (see [18]).

The generalized mid-point and spread of a fuzzy set A is a way of identifying levelwise the center (location) and the extent (imprecision) by considering each direction in the multidimensional case through the unit sphere $\mathbb{S}^{p-1}$.

Formally, let $\alpha \in [0,1]$ and $u \in \mathbb{S}^{p-1}$, and calculate the lengths $\pi_u(A_\alpha)$ of all orthogonal projections of A_α on this direction, i.e.

$$\pi_u(A_\alpha) = \big[\underline{\pi}_u(A_\alpha)\,,\, \overline{\pi}_u(A_\alpha)\big] = \big[-s_{A_\alpha}(-u)\,,\, s_{A_\alpha}(u)\big]$$

where s stands for the support function of a nonempty convex compact set, that is, $s_{A_\alpha}(u) = \sup_{a \in A_\alpha} \langle u, a\rangle$ ($\langle\cdot,\cdot\rangle$ denoting the usual inner product in $\mathbb{R}^p$). Thus, the generalized mid-point and spread of A are defined as the functions $\mathrm{mid}_A, \mathrm{spr}_A : \mathbb{S}^{p-1} \times [0,1] \to \mathbb{R}$ so that

$$\mathrm{mid}_A(u,\alpha) = \mathrm{mid}_{A_\alpha}(u) = \frac{1}{2}\big(s_{A_\alpha}(u) - s_{A_\alpha}(-u)\big),$$

$$\mathrm{spr}_A(u,\alpha) = \mathrm{spr}_{A_\alpha}(u) = \frac{1}{2}\big(s_{A_\alpha}(u) + s_{A_\alpha}(-u)\big).$$

Note that in the interval case, $\mathbb{S}^{p-1} = \{-1,1\}$ and $\mathrm{mid}_A(-1,\alpha) = -\mathrm{mid}\, A$, $\mathrm{mid}_A(1,\alpha) = \mathrm{mid}\, A$, $\mathrm{spr}_A(-1,\alpha) = \mathrm{spr}_A(1,\alpha) = \mathrm{spr}\, A$ holds for all $\alpha \in [0,1]$.

ISBN: 978-989-95079-6-8

The generalized mid-point and spread are not defined as numbers summarizing the 'central tendency' and the 'imprecision', but as functions identifying the 'central points' and the 'imprecision' in the different directions of the Euclidean space. In this way, we get a meaningful characterization of the fuzzy sets alternative to the classical support functions.

The class of distances in [18] is defined from the distances between the level sets as a meaningful generalization of the Bertoluzza et al. metric [1], by considering L_2-type distances between the mid-points and the spreads. Specifically, for each level set $\alpha \in [0,1]$, we define

$$d_\theta^2\big(A_\alpha, B_\alpha\big) = \|\mathrm{mid}\, A_\alpha - \mathrm{mid}\, B_\alpha\|^2 + \theta\|\mathrm{spr}\, A_\alpha - \mathrm{spr}\, B_\alpha\|^2$$

where $\|\cdot\|$ is the usual L_2-norm in the space of the square-integrable functions $L^2(\mathbb{S}^{p-1})$, and $0 < \theta \leq 1$ determines the relative importance of the squared distance between the spreads in relationship with the squared distance between the mids.

The metric between fuzzy sets D_θ^φ is defined as a weighting mean (w.r.t. a probability measure φ with support $[0,1]$) of the distances between the level sets by

$$D_\theta^\varphi(A,B) = \left(\int_{[0,1]} d_\theta^2(A_\alpha, B_\alpha) d\varphi(\alpha)\right)^{1/2}$$

The weight measure φ reflects the intuitive (or subjective) interpretation of fuzzy sets - for instance one may treat every α-level as equally important (and therefore use the Lebesgue measure as weight measure φ), or give more mass to α-levels close to 1 or to α-levels close to 0.

If we denote by $\|\cdot\|_{\tilde{2}}$ the usual L_2-norm on the Hilbert space $\mathcal{H} = L^2(\mathbb{S}^{p-1} \times [0,1])$, we have the following alternative intuitive expression for the metric:

$$\big(D_\theta^\varphi(A,B)\big)^2 = \big\|\mathrm{mid}_{\,A} - \mathrm{mid}_{\,B}\big\|_{\tilde{2}}^2 + \theta\big\|\mathrm{spr}_{\,A} - \mathrm{spr}_{\,B}\big\|_{\tilde{2}}^2.$$

Let $(\Omega, \mathcal{A}, P)$ be a probability space. A *Fuzzy Random Variable* (FRV) can be identified with a Borel measurable mapping $\mathcal{X} : \Omega \to \mathcal{F}_c(\mathbb{R}^p)$ (see [2, 16]). The concepts of induced probability distribution and independence are the same as for general metric-space valued random elements.

3 Nonparametric discriminant approach for fuzzy data

Assume that we have a population $(\Omega, \mathcal{A}, P)$, and for each individual we observe a fuzzy datum. In addition, each individual may belong to one of k different categories $g_1, \ldots, g_k$, and as learning sampling we have the fuzzy data and the group of n independent individuals. The goal is to find a rule that allows us to classify a new individual in one of the k groups from the fuzzy datum. For this purpose we suggest to use a nonparametric approach.

Formally, let $(\mathcal{X}, G) : \Omega \to \mathcal{F}_c(\mathbb{R}^p) \times \{g_1, \ldots, g_k\}$ be a random element in such a way that $\mathcal{X}(\omega)$ is a fuzzy datum and $G(\omega)$ is the membership group (g_1,... or g_k) of each individual $\omega \in \Omega$. Assume that we have n independent copies of $(\mathcal{X}, G)$ as training sample, that is, we have a random sample $\{\mathcal{X}_i, G_i\}_{i=1}^n$. As in a more general functional context (see [6]), we propose to estimate nonparametrically

$$P(G = g_j | \mathcal{X} = \widetilde{x})$$

for $j = 1, \ldots, k$, $\widetilde{x} \in \mathcal{F}_c(\mathbb{R}^p)$, and then to assign the new data to the class of higher estimated probability.

In order to find a reasonable estimator of the preceding probability consider first $\delta > 0$ and a closed ball $B(\widetilde{x}; \delta)$ defined in terms of the metric introduced in Section 2. From Bayes Theorem we have that

$$P(G = g_j | \mathcal{X} \in B(\widetilde{x}; \delta))$$

$$= \frac{P(X \in B(\widetilde{x}; \delta) | G = g_j) P(G = g_j)}{\sum_{l=1}^k P(X \in B(\widetilde{x}; \delta) | G = g_l) P(G = g_l)}$$

$$= \frac{P(D_\theta^\varphi(\mathcal{X}, \widetilde{x}) \leq \delta | G = g_j)) P(G = g_j)}{\sum_{l=1}^k P(D_\theta^\varphi(\mathcal{X}, \widetilde{x}) \leq \delta | G = g_l) P(G = g_l)}$$

For each $\widetilde{x}$, obviously $D_\theta^\varphi(\mathcal{X}, \widetilde{x})$ is a (real-valued) random variable. Let $F_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}$ denote the distribution function of this variable in the group g_j. Consequently,

$$P(G = g_j | \mathcal{X} \in B(\widetilde{x}; \delta))$$

$$= \frac{F_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}(\delta) P(G = g_j)}{\sum_{l=1}^k F_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_l}(\delta) P(G = g_l)}$$

If we assume (as it is natural in the nonparametric setting) that $F_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}$ is uniformly continuous for all $j = 1, \ldots k$, then we have that there exists an underlying density function $f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}$ and

$$\lim_{\delta \to 0} \frac{F_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}(\delta)}{\delta} = f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}(0)$$

As a result,

$$P(G = g_j | \mathcal{X} = \widetilde{x})$$

$$= \frac{f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}(0) P(G = g_j)}{\sum_{l=1}^k f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_l}(0) P(G = g_l)}.$$

Note that the denominator is just $f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})}(0)$, and for this reason according to [6] we could estimate $P(G = g_j | \mathcal{X} = \widetilde{x})$ by means of

$$\widehat{P}(G = g_j | \mathcal{X} = \widetilde{x})$$

$$\frac{\sum_{i=1}^n I_{G=g_j} K(h^{-1} D_\theta^\varphi(\mathcal{X}_i, \widetilde{x}))}{\sum_{i=1}^n K(h^{-1} D_\theta^\varphi(\mathcal{X}_i, \widetilde{x}))}$$

$$= \frac{(nh)^{-1} \sum_{i=1}^n I_{G=g_j} K(h^{-1} D_\theta^\varphi(\mathcal{X}_i, \widetilde{x}))}{(nh)^{-1} \sum_{i=1}^n K(h^{-1} D_\theta^\varphi(\mathcal{X}_i, \widetilde{x}))}$$

ISBN: 978-989-95079-6-8

$$= \frac{\left[(n_j h)^{-1} \sum_{i \in N_j} K(h^{-1} D_\theta^\varphi(\mathcal{X}_i, \widetilde{x}))\right] \cdot (n_j n^{-1})}{(nh)^{-1} \sum_{i=1}^{n} K(h^{-1} D_\theta^\varphi(\mathcal{X}_i, \widetilde{x}))}$$

where $N_j = \{i \in \{1, \ldots, n\} | G_i = g_j\}$, n_j is the cardinality of N_j, K is a kernel and h the bandwidth.

The last expression shows clearly that the estimator depends on an estimate of $f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}(0)$ (the densities within each group at the point 0), the empirical estimate of $P(G = g_j)$ and an estimate of $f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})}(0)$ (the overall density at the point 0). In this approach the bandwidth is the same for all the estimates, and it is determined by employing the whole sample. In this way, it is verified that

$$\sum_{j=1}^{k} \widehat{P}(G = g_j | \mathcal{X} = \widetilde{x}) = 1.$$

However, since that $f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})}(0)$ is composed by a mixture of the variables $\{D_\theta^\varphi(\mathcal{X}, \widetilde{x}) | G = g_j\}_{j=1}^k$ whose distributions are expected to be different (recall that the aim is to discriminate among them), it seems more convenient to estimate separately each distribution (each one of them with its own bandwidth) and then to combine them to estimate $f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})}(0)$. Thus, we propose to estimate $f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}(0)$ by means of

$$\widehat{f}_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}(0) = (n_j h_j)^{-1} \sum_{i \in N_j} K(h^{-1} D_\theta^\varphi(\mathcal{X}, \widetilde{x}))$$

for all $j = 1, \ldots, k$. In this way,

$$\widehat{P}(G = g_j | \mathcal{X} = \widetilde{x}) = \frac{n_j n^{-1} \widehat{f}_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}(0)}{\sum_{l=1}^{k} (n_j n^{-1}) \widehat{f}_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_l}(0)},$$

and it is also verified that

$$\sum_{j=1}^{k} \widehat{P}(G = g_j | \mathcal{X} = \widetilde{x}) = 1.$$

The considered estimators are statistically consistent under the unique assumption that the corresponding densities exist for each $\widetilde{x} \in \mathcal{F}_c(\mathbb{R}^p)$ (and the usual regularity conditions are satisfied), however there is no assumption about the changes in the distributions as $\widetilde{x}$ varies.

4 Empirical results

In an experiment about the visual perception of the length of a rule relatively to a maximum, we have shown images like those in Figure 1 to different people. The aim has been to know the relative measure of the dark line (below)in contrast with the light line (above) for these people. They have been allowed to express the lack of precision that they may have by using a fuzzy scale. After a trial with the software, people have seen different rules with random length and they have indicated their perception about the length in the fuzzy scale and the linguistic label (very small, small, medium, large, very large) that they consider suitable for such a length.

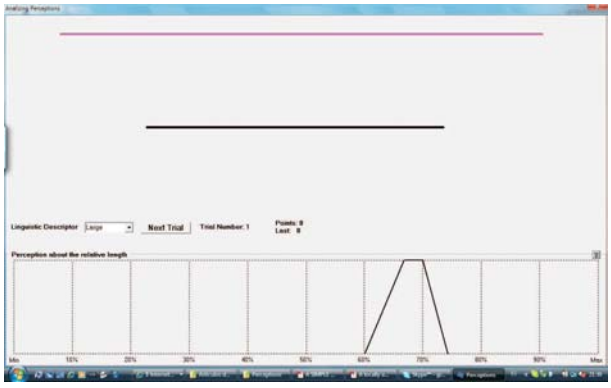


Figure 1: Software to evaluate the visual perception of a rule

The way of using fuzzy scale as explained in the software is as follows:

"*This experiment regards your perception about the relative length of different lines.*

On the top of the screen, we have plotted in light color the longest line that we could show to you. This line will remain visible in the current position during all the experiment, so that you can always have a reference of the maximum length.

At each trial of the experiment we will show you a dark line and you will be asked about its relative length (in comparison with the length of the reference light line):

- *Firstly you will be asked for a linguistic descriptor of the relative length. We have consider five descriptors (Very Small; Small; Medium; Large; Very Large). The aim is to select one of these descriptors at first sign (you can change it later if you want to).*
- *Secondly you will be asked for your own estimate or perception (without physically measuring it) of the relative length (in percentage) by means of a Fuzzy Set (the information about the design and interpretation of the Fuzzy Set will be shown to you at this time).*
- *Finally, in case your initial perception had been changed during the process you can readjust again the linguistic descriptor of its relative length.*"

Concerning the design and interpretation of the fuzzy set, when the image of Figure 1 is shown, the respondents have been asked to draw a fuzzy set representing their perceptions. In accordance with the usual interpretation, the respondents have to choose the 0-level (set of all those points with a positive degree of membership) as the set of all values that they consider compatible with the length to a greater or lesser extent. In the same way, the 1-level (set of all those points with total degree of membership) has to be fixed as the set of values that they consider completely compatible with their perception about the measure of the line. Although it is possible to change the shape of the resulting fuzzy sets, by default the trapezoidal fuzzy set formed by the interpolation of both intervals is fixed (as shown in Figure 1).

In Table 1 we show some of the data of a person who made 220 trials.

The goal here is to predict the category (very small, small, medium, large or very large) that this person would consider

ISBN: 978-989-95079-6-8

Table 1: Perceptions about the relative length of the rule.

Trial	$\inf P_0$	$\inf P_1$	$\sup P_1$	$\sup P_0$	Ling. descrip.
1	78.27	80.94	84.41	87.40	large
2	54.93	58.00	62.20	65.67	large
3	47.25	49.43	50.89	53.31	medium
4	92.65	95.72	97.58	99.11	very large
5	12.92	15.51	17.77	20.03	very small
6	32.55	36.03	39.90	42.89	small
7	2.50	4.44	6.22	9.21	very small
8	24.80	28.19	30.45	33.28	small
9	55.17	58.40	61.79	65.75	large
10	2.26	3.63	5.57	8.08	very small

suitable from the fuzzy perception that he/she has about the length of the line. It should be remarked that the categories are treated here simply as different classes, which may be also labelled as 1, 2, 3, 4 and 5, irrespectively of the fuzzy representation that they may have. The consideration of fuzzy labels would lead to a different approach.

The line showed at each trial has been chosen at random, although to obtain also a validation sample we have proceeded as follows:

- 166 lengths were generated by means of random numbers between 0 and 100. They will be used as training sample.
- The 9 lengths in the equally spaced discrete set $\{100/27 + (i/8)100(1 - 2/27)\}_{i=0,\ldots,8}$ have been repeated 6 times. Thus, we have 54 lengths that are representative of the different situations that may arise, and that can be used as validation sample.
- The 166 random lengths of the training sample and the 54 lengths of the validation sample are interspersed and shown at random.

As it was explained in Section 3, we need to estimate the density of the random variable $D_\theta^\varphi(\mathcal{X}, \widetilde{x})$ at the point 0 for any fuzzy response $\widetilde{x}$ we obtain. The problem of estimating a density in the boundary of its support is not easy. Several approaches can be found in the literature (see for instance, [15], [13] and [11]), however none of them have solved all the inconveniences that may arise. We have considered different options to evaluate its performance in this case study:

M1: Firsty a kernel density estimation with Quartic Kernel

$$K_1(u) = \frac{15}{16}(1-u^2)^2 I_{(|u|\leq 1)}.$$

and bandwidth chosen by cross-validation were considered. It is well-known that this estimator is non consistent in the boundaries. In particular the estimation at 0 in this case is not good in general (see for instance, [15]). As we have mentioned, several boundary corrections to overcome this problem are available. Many of these corrections (basically based on Jacknife) lead to consistent estimators at the boundaries, although taking negative values in many practical situations (which makes them useless in our case). There are some exceptions in the literature that yield to both consistent and non-negative estimates.

M2: In particular, as a second scenario we have considered the proposal in [6] of using an asymmetric kernel. To be precise, we consider the absolute value of the Quartic kernel K_1, that is,

$$K_2(u) = \frac{15}{8}(1-u^2)^2 I_{(u\in[0,1])}.$$

This approach leads to consistent estimates at the boundary (but with a low convergence speed ratio). There are no studies about the optimal way of choosing the bandwidth for this asymmetric kernel, so we have used the common one provided by cross-validation techniques applied to this particular case.

M3: Finally, as a third strategy we have considered the estimator proposed by [14] which is a kernel density estimation with boundary correction that provides with non-negative estimates, that is,

$$f_P(0) = \overline{f}(0)\exp\left\{\frac{\hat{f}(0)}{\overline{f}(0)} - 1\right\}$$

where $\hat{f}(0)$ denotes the basic kernel density estimator divided by $a_0(0) = \int_{-1}^{0} K(u)du$ and $\overline{f}$ is the boundary corrected kernel density estimator based on generalised jackknifing by combining $\hat{f}$ with the analogous estimator based on the kernel $L(u) = 0.5K(2u)$. In this case and optimal (local) bandwidth is the solution of a complex variational problem (see [15] for a similar situation), that in general is not possible to obtain. For this reason we have considered the global cross-validation bandwidth for the kernel $K = K_1$.

In Table 2 we show the percentage of right classification corresponding to each one of the considered methods. We observe a better behaviour for **M2** than for **M3**. In addition **M1** is quite similar to **M3** in spite of being theoretically inconsistent. This seems to indicate that the selection of a right bandwidth is more important than the theoretical accuracy of the considered estimator (**M1** and **M2** are worse from a theoretical point of view than **M3**, provided that the optimal bandwidth is considered).

Table 2: Percentage of right classification with different methods.

	M1	**M2**	**M3**
% of right classification	75.92	83.33	77.77

5 Concluding remarks

This works is a preliminary study concerning the problem of supervised classification of fuzzy random variables. We propose to use a nonparametric approach to better capture nonlinear boundaries. However, other interesting viewpoints may be used (either by extending those in functional data analysis, as the penalized or flexible discriminant analyses, or by being developed ad-hoc for this case). Further theoretical and empirical comparative studies should be developed.

As it is mentioned in Section 4, the nonparametric estimation problem to be solved here is not easy, and there is no

ISBN: 978-989-95079-6-8

general optimal method to do it. Our empirical results seem to indicate the importance of the bandwidth selection. As it is well-known, the optimal bandwidth for the estimation of a density function is not appropriate in general for the estimation of its curvature. In the same way, the optimal bandwidth for the estimation of a density function is non necessarily appropriated for the classification problem. As an open problem, we propose to optimize the bandwidth selection for the classification problem by choosing, for instance, the values maximizing the proportion of right classifications.

On the other hand, it seems also very interesting to consider the case in which the group membership of the training data is imprecise, as it is done in [4] for real learning data.

Acknowledgment

The research in this communication has been partially supported by the Grant from the Spanish Ministry of Education and Science MTM2006-07501. This financial support is gratefully acknowledged.

References

[1] Bertoluzza, C., Corral, N., Salas A., 1995. On a new class of distances between fuzzy numbers. Mathware & Soft Computing 2, 71–84.

[2] Colubi, A., Domínguez-Menchero, J.S., López-Díaz M., Ralescu, D. A. On the formalization of fuzzy random variables, Inform. Sci. 133 (2001) 3–6.

[3] Colubi, A. 2009. Statistical inference about the means of fuzzy random variables: Applications to the analysis of fuzzy- and real-valued data. Fuzzy Sets and Systems. 160 (3), 344-356.

[4] Côme, E., Oukhellou, L., Denœux, T., Aknin, P. 2009 Learning from partially supervised data using mixture models and belief functions. Pattern Recognition 42 (3), pp. 334-348

[5] Denœux, T., Masson, M.-H., Hébert, P.-A. 2005 Nonparametric rank-based statistics and significance tests for fuzzy data Fuzzy Sets and Systems 153 (1), 1-28.

[6] Ferraty, F., Vieu, P. (2003). Curves discrimination: a nonparametric functional approach. *Comput. Statist. Data Anal.* **44**, 161–173.

[7] Flórez, J.L., Colubi, A., González-Rodríguez, G., Gil, M.A. 2006. Nonparametric regression with fuzzy random variables through the support functions. In Proceedings of the 11th International Conference on Information Processing and Management of Uncertainty (IPMU'2006, Paris), pp. 724-730.

[8] González-Manteiga, W.G., Vieu, P. Statistics for Functional Data. Computational Statistics and Data Analysis 51 (10), pp. 4788-4792. 2007.

[9] González-Rodríguez, G., Colubi, A., Trutschnig, W. 2009. Simulation of fuzzy random variables. Information Sciences. 179 (5), 642–653.

[10] González-Rodríguez, G., Colubi, A., D'Urso, P., Montenegro, M. 2009. Multi-sample test-based clustering for fuzzy random variables. International Journal of Approximate Reasoning.

[11] Hall, P., Park, B.U. 2002. New methods for bias correction at endpoints and boundaries. The Annals of Statistics 30 (5), 14601479.

[12] Hall, P., Müller, H.-G., Wang, J.-L. 2006. Properties of principal component methods for functional and longitudinal data analysis. Annals of Statistics 34 (3), pp. 1493-1517

[13] Jones, M. C., Foster, P. J. 1993. Generalized jackknifing and higher order kernels. J.Nonparametric Statist. 3, 81-94.

[14] Jones, M. C., Foster, P. J. 1996. A simple nonnegative boundary correction method for kernel density estimation. Statistica Sinica 6(1996), 1005-1013.

[15] Müller, H.-G. 1991. Smooth optimum kernel estimators near endpoints. Biometrika 78, 521-530.

[16] Puri, M. L., Ralescu, D. A. 1986. Fuzzy random variables. J. Math. Anal. Appl. 114, 409–422.

[17] Ramsay, J. O., Silverman, B. W. (2002). *Applied functional data analysis. Methods and case studies.* Springer Series in Statistics. Springer-Verlag, New York.

[18] Trutschnig, W. González-Rodríguez, G., Colubi, A., Gil, M.A. 2009. A new family of metrics for compact, convex (fuzzy) sets based on a generalized concept of mid and spread. Submitted.

[19] Zhang, J.-T., Chen, J. Statistical inferences for functional data. Annals of Statistics 35 (3), pp. 1052-1079. 2007

ISBN: 978-989-95079-6-8

Confidence regions for the mean of a fuzzy random variable

Gil González-Rodríguez, Wolfgang Trutschnig[1], Ana Colubi[2]

1. European Centre for Soft Computing
Mieres, Spain
2. Department of Statistics and OR, University of Oviedo
Oviedo, Spain
Email: {gil.gonzalez,wolfgang.trutschnig}@softcomputing.es, colubi@uniovi.es

Abstract— *The aim of this paper is to extend the classical problem of confidence interval estimation for the mean of a random variable to the case of a fuzzy random variable. The key idea consists in considering a confidence region defined as a ball w.r.t. a given metric, which is centered in the sample mean and whose radius is determined via bootstrapping. The developed approach is illustrated both by means of simulated examples using C and R and by means of a practical example consisting of courses evaluations.*

Keywords— Fuzzy random variables, confidence interval, confidence region, expectation, bootstrap.

1 Introduction

Inferential procedures for handling fuzzy information have gained in importance in the literature during the last years. Hypothesis testing problems concerning fuzzy random variables (hereafter FRV for short) in different situations (see, for instance, [9, 13, 14, 5, 6, 3]) as well as consistency properties of the considered estimators, i.e. the associated point-wise estimation problems, have been studied extensively (see, for instance, [12, 10, 16, 7]). Nevertheless, the problem of confidence interval estimation for FRVs has not received too much attention. In this communication we will focus on the the problem of estimating by a confidence interval the expected value of an FRV in Puri & Ralescu's sense [15].

Initially, there are two main difficulties in connection with this estimation problem: the well-known lack of ordering and the lack of linearity of the space of fuzzy sets. The lack of ordering is not crucial - indeed this difficulty has been faced previously in some classical situations (as for instance in the determination of confidence bands for the linear regression). On the other hand the lack of linearity makes it senseless to attempt to express the confidence interval in the standard way (sample mean plus/minus a given quantity depending on the variability and the sample size). Consequently, we will speak about confidence regions instead of confidence intervals in the following, keeping in mind that these regions cannot be specified simply by means of an upper and lower bound.

It is well-known that if X is a real-valued random variable with mean μ and finite variance, then, based on a random sample $X_1, \ldots, X_n$ of n independent random variables having the same distribution as X, a confidence interval with confidence level $100(1-\beta)\%$ for μ, can be determined by

$$IC_n = [\overline{X} - \delta, \overline{X} + \delta],$$

whereby $\overline{X}$ is the sample mean and $\delta = \delta(X_1, \ldots, X_n)$ is so that $P(\mu \in IC_n) = 1 - \beta$. The keys are, firstly, that this confidence interval can be seen as a ball with respect to the Euclidian distance which is centered in the sample mean $\overline{X}$ and has radius δ and, secondly, that in practice δ can be computed via bootstrapping. In other words: classical confidence intervals for the mean μ can equivalently be seen as balls, having the sample mean as center and a suitable radius.

Having this is in mind, our goal is to consider a *good* metric (that is, both easy to handle and easy to interpret) on the space of all fuzzy sets and, on the basis of this metric, to define confidence regions as balls centered in the sample mean and with a given radius which is empirically determined by applying bootstrap procedures. By the comments made before this technique surely is a natural extension of the classical procedure for real random variables.

The rest of the manuscript is organized as follows. In Section 2 we will recall some preliminary concepts concerning FRVs. At the beginning of Section 3 we propose an initial algorithm to determine a confidence ball for the mean of an FRV by using bootstrap. After checking its capability by means of various simulations made in C and R we finally present an improvement of the algorithm. In Section 4 we will apply the developed procedure to a concrete fuzzy sample consisting of courses evaluations. Finally, some concluding remarks, next developments and open problems are gathered in Section 5.

2 Preliminaries

Fuzzy random variables in Puri and Ralescu's sense [15] are an extension of both random variables and random (convex compact) sets. They allow us to model situations in which an imprecise (fuzzy) value is associated with the outcome of a stochastic experiment. In order to give an exact definition of FRVs we need some preliminary notions.

$\mathcal{K}_c(\mathbb{R})$ will denote the family of all non-empty compact intervals. $\mathcal{F}_c(\mathbb{R})$ denotes the class of fuzzy numbers, i.e. functions $U : \mathbb{R} \to [0,1]$, whose α-levels U_α fulfill $U_\alpha \in \mathcal{K}_c(\mathbb{R})$ for all $\alpha \in [0,1]$ whereby, as usual, $U_\alpha := [\underline{U}_\alpha, \overline{U}_\alpha] := \{x \in \mathbb{R} | U(x) \geq \alpha\}$ for every $\alpha \in (0,1]$ and $U_0 := [\underline{U}_0, \overline{U}_0]$ is defined as the topological closure of $\cup_{\alpha \in (0,1]} U_\alpha$.

The class $\mathcal{F}_c(\mathbb{R})$ is usually endowed with a semilinear structure that is compatible with Zadeh's Extension Principle (see [17]) - i.e. with a *sum* (extending the Minkowsky addition of intervals), defined by $(U+V)_\alpha = U_\alpha + V_\alpha$ for all $\alpha \in [0,1]$, and with a *product by a scalar*, $(aU)_\alpha = aU_\alpha$ for all $\alpha \in [0,1]$ and $a \in \mathbb{R}$. For every $U \in \mathcal{F}_c(\mathbb{R})$ mid_U and spr_U are defined to be real-valued functions on $[0,1]$ such that $mid_U(\alpha) := (\underline{U}_\alpha + \overline{U}_\alpha)/2$ and $spr_U(\alpha) := (\overline{U}_\alpha - \underline{U}_\alpha)/2$

ISBN: 978-989-95079-6-8

for every $\alpha \in [0,1]$.
A very useful metric on $\mathcal{F}_c(\mathbb{R})$ from the statistical point of view is the one introduced by Bertoluzza *et al* [1]. It depends on two weighting probability measures W and φ, defined on $([0,1], \mathcal{B}_{[0,1]})$ ($\mathcal{B}_{[0,1]}$ being the Borel σ-field on $[0,1]$). For any two fuzzy numbers $U, V \in \mathcal{F}_c(\mathbb{R})$ the D_W^φ-distance is defined as:

$$\sqrt{\int_{[0,1]^2} \left[t(\underline{U}_\alpha - \underline{V}_\alpha) + (1+t)(\overline{U}_\alpha - \overline{V}_\alpha)\right]^2 \, dW(t)d\varphi(\alpha)}$$

Thereby φ is assumed to have a strictly increasing distribution function on $[0,1]$ (and therefore has the whole interval $[0,1]$ as support) and W is non-degenerate (i.e. it is not a Dirac measure). A especially useful subclass of choices for W, on that we will concentrate in the sequel is that for which $\int_0^1 t dW(t) = 1/2$ (which in fact is equivalent to assuming D_W^φ to be invariant to rigid motions on $\mathbb{R}$). In this case it is easy to see that $(D_W^\varphi(U,V))^2$ can be rewritten as follows:

$$(D_W^\varphi(U,V))^2 = \int_{[0,1]} (\mathrm{mid}_U(\alpha) - \mathrm{mid}_V(\alpha))^2 d\varphi(\alpha)$$

$$+\theta_W \int_{[0,1]} (\mathrm{spr}_U(\alpha) - \mathrm{spr}_V(\alpha)^2 d\varphi(\alpha)$$

That is, the squared D_W^φ-distance of $U, V \in \mathcal{F}_c(\mathbb{R})$ is the sum of the squared φ-weighted L_2-distances between the mids and between the spreads, where the parameter $\theta_W \in (0,1]$ allows us to weight the relative importance of the distance between mids w.r.t. the distance between spreads (in particular, when W is chosen as the Lebesgue measure λ then $\theta_W = 1/3$).

Having this, we can now return to the definition of an FRV mentioned before. Let $(\Omega, \mathcal{A}, P)$ be an arbitrary probability space. an FRV $\mathcal{X}$ is a mapping $\mathcal{X} : \Omega \to \mathcal{F}_c(\mathbb{R})$, which is Borel-measurable with respect to the D_W^φ-metric. This definition is different from the original one by Puri and Ralescu, however in many cases and in particular in the cases we will consider in the sequel coincides with the original one (see [15, 11, 2]).
The FRV $\mathcal{X}$ is said to be integrably bounded whenever $\max\{|\min X_0|, |\max X_0|\} \in L^1(\Omega, \mathcal{A}, P)$. If $\mathcal{X}$ is an integrably bounded FRV then its *expected value (or mean)* is defined as the unique fuzzy number $\mathbb{E}(\mathcal{X}) \in \mathcal{F}_c(\mathbb{R})$ such that $\big(\mathbb{E}(\mathcal{X})\big)_\alpha$ is the Aumman integral of the random set $\mathcal{X}_\alpha$ for all $\alpha \in [0,1]$ (see [15]). In particular,

$$\big(\mathbb{E}(\mathcal{X})\big)_\alpha = [\mathbb{E}(\min \mathcal{X}_\alpha), \mathbb{E}(\max \mathcal{X}_\alpha)]$$

holds for all $\alpha \in [0,1]$ if $\mathcal{X}$ is integrably bounded.
If, in addition, $\max\{|\min \mathcal{X}_0|, |\max \mathcal{X}_0|\} \in L^2(\Omega, \mathcal{A}, P)$, then the (W,φ)-*variance* of $\mathcal{X}$ (see [12, 10]) is defined as

$$\mathrm{Var}(\mathcal{X}) = \mathbb{E}([D_W^\varphi(\mathcal{X}, \mathbb{E}(\mathcal{X}))]^2). \quad (1)$$

Remark: The main reasons for using the above mentioned D_W^φ-metric are that (1) it has more good intuitive properties (see [1]) than the usual L_p metrics d_p (see [4]) do not have, and (2) it can be expressed as an inner product on the Hilbert space of all square integrable functions on $\mathcal{H} = L^2([0,1]\times\{-1,1\})$ (via using the concept of support function).

3 Confidence regions for the mean of an FRV

Let $(\Omega, \mathcal{A}, P)$ be an arbitrary probability space, and consider an FRV $\mathcal{X} : \Omega \to \mathcal{F}_c(\mathbb{R})$ verifying $\max\{|\min \mathcal{X}_0|, |\max \mathcal{X}_0|\} \in L^2(\Omega, \mathcal{A}, P)$. As mentioned in the introduction, our aim is to calculate a confidence ball for the expectation $\mu = \mathbb{E}(\mathcal{X}) \in \mathcal{F}_c(\mathbb{R})$ on the basis of n independent observations $(\mathcal{X}_1, \ldots, \mathcal{X}_n)$ of $\mathcal{X}$.
We will denote by $\overline{\mathcal{X}}$ and $\widehat{S}$ the associated sample fuzzy mean and sample (W,φ)-deviation respectively, that is,

$$\overline{\mathcal{X}} = \frac{1}{n}\sum_{i=1}^n \overline{\mathcal{X}}_i, \quad \widehat{S} = \sqrt{\frac{1}{n}\sum_{i=1}^n [D_W^\varphi(\mathcal{X}_i, \overline{\mathcal{X}})]^2}. \quad (2)$$

Following the approach mentioned in the introduction, for a given significance level $\beta \in (0,1)$ we define the confidence ball CR_β with respect to the D_W^φ-metric as

$$CR_\beta = B(\overline{\mathcal{X}}, \delta) = \{U \in \mathcal{F}_c(\mathbb{R}) | D_W^\varphi(U, \overline{\mathcal{X}}) \le \delta\}, \quad (3)$$

whereby the radius δ has to verify the coverage condition

$$P(\mu \in B(\overline{\mathcal{X}}, \delta)) = P(D_W^\varphi(\mu, \overline{\mathcal{X}}) \le \delta) = 1 - \beta. \quad (4)$$

Due to the non-existence of sufficiently general parametric models for FRVs in general it is not possible to find δ fulfilling the coverage condition. Nevertheless we may choose δ as the $(1-\beta)$-quantile of the distribution of $D_W^\varphi(\mu, \overline{\mathcal{X}})$, which in turn can be approximated by the corresponding bootstrap $(1-\beta)$-quantile. In fact, given a sample $(\mathcal{X}_1, \ldots, \mathcal{X}_n)$ of $\mathcal{X}$ we propose to proceed as follows:

Algorithm 3.1

Step 1. *Fix the significance level $\beta \in (0,1)$ and the number B of bootstrap replications.*

Step 2. *Obtain B bootstrap samples $(\mathcal{X}_1^{b\star}, \ldots, \mathcal{X}_n^{b\star})$ $(b = 1, \ldots, B)$ from $(\mathcal{X}_1, \ldots, \mathcal{X}_n)$, and for each one compute its corresponding sample mean $\overline{\mathcal{X}}^{b\star}$.*

Step 3. *Compute the distance between the sample mean and each bootstrap sample mean, $d_b^\star = D_W^\varphi(\overline{\mathcal{X}}, \overline{\mathcal{X}}^{b\star})$, for each $b = 1, \ldots, B$.*

Step 4. *Choose δ as one of the $(1-\beta)$-quantiles of the sample $(d_1^\star, \ldots, d_B^\star)$ (that is, choose δ so that at least the $100(1-\beta)\%$ of the computed distances are smaller or equal than δ and at least the $100\beta\%$ of the computed distances are greater or equal than δ).*

In order to analyze the quality and accuracy of the confidence regions obtained by means of this algorithm we have made simulations with different sample sizes n. We have chosen φ as the Lebesgue measure on $[0,1]$ and used two different weight distributions W (in fact $\theta_W = 1/3$, which corresponds to the Lebesgue measure, and $\theta_W = 9/10$ that gives more weight to the spreads, were used). Each simulation corresponds to 10000 iterations with 1000 bootstrap replications and a significance level β of 0.05.
For the simulation of general FRVs the approach introduced in [8] has been applied and $n_0 = 101$ equally spaced alpha-levels $(\alpha_i = (i-1)/(n_0-1),\ i = 1, \ldots, n_0)$ were considered. In short, this approach consists essentially of the following three steps (for more details see [8]):

ISBN: 978-989-95079-6-8

SIM1 **Decomposition:**
Fix the (future) expectation $V \in \mathcal{F}_c(\mathbb{R})$, an index $n_0 \in \mathbb{N}$ and α-levels $0 = \alpha_1 < \alpha_2 < \cdots < \alpha_{n_0} = 1$. Let V^c be the mid-point of the 1-level of V, $V^l := V^c - \min(V_0)$ the total left spread and $V^r := \max(V_0) - V^c$ the total right spread. Furthermore define

$$L_\alpha = \begin{cases} \{0\} & \text{if } V^l = 0, \\ [(\min(V_\alpha) - V^c)/V^l, 0] & \text{if } V^l \neq 0 \end{cases}$$

$$R_\alpha = \begin{cases} \{0\} & \text{if } V^r = 0, \\ [0, (\max(V_\alpha) - V^c)/V^r] & \text{if } V^r \neq 0 \end{cases}$$

for every $\alpha \in [0,1]$ and set $F^l(t) := \max(-L_{1-t})$ and $F^r(t) := \max(R_{1-t})$ for every $t \in [0,1]$ (in fact F^l and F^r are distribution function but we do not use their stochastic interpretation).

SIM 2 **Discretization:**
Define $p_1^l = F^l(\alpha_1)$ and $p_i^l = F^l(\alpha_i) - F^l(\alpha_{i-1})$ for all $i = 2, \ldots, n_0$, and $p_1^r = F^r(\alpha_1)$ and $p_i^r = F^r(\alpha_i) - F^r(\alpha_{i-1})$ for all $i = 2, \ldots, n_0$ respectively. Let $B_x \in \mathcal{F}_c(\mathbb{R})$ denote the fuzzy set fulfilling $(B_x)_\alpha = [0, \mathbf{1}_{[\alpha,1]}(x)]$ for every $\alpha \in [0,1]$ and set

$$V_{n_0} = V^c + \sum_{i=1}^{n_0} -B_{1-\alpha_i} V^l p_i^l + \sum_{i=1}^{n_0} B_{1-\alpha_i} V^r p_i^r$$

SIM 3 **Stochastic Perturbation**
Consider a $(2n_0+1)$-dimensional random vector

$$\mathcal{Y} = (C_0, C_1^l, \ldots, C_{n_0}^l, C_1^r, \ldots, C_{n_0}^r) : \Omega \to \mathbb{R} \times [0,\infty)^{2n_0}$$

of coefficients for the 'approximating' simulated FRV $\mathcal{X}^\star$ as random perturbation of $(V^c, V^l p_1^l, \ldots, V^l p_{n_0}^l, V^r p_1^r, \ldots, V^r p_{n_0}^r)$ in such a way that

$$\mathbb{E}(C_0, C_1^l, \ldots, C_{n_0}^l, C_1^r, \ldots, C_{n_0}^r) = \\ = (V^c, V^l p_1^l, \ldots, V^l p_{n_0}^l, V^r p_1^r, \ldots, V^r p_{n_0}^r)$$

holds and set

$$\mathcal{X}^\star := C^0 + \sum_{i=1}^{n_0} -B_{1-\alpha_i} C_i^l + \sum_{i=1}^{n_0} B_{1-\alpha_i} C_i^r \qquad (5)$$

This procedure generates a sample of size 1 of the FRV $\mathcal{X}$ - for a sample of size n SIM 3 simply has to be repeated n times.

Remark: To be precise, in the simulations we have made we have used perturbations from the same distributions, i.e. $C_1^l, \ldots, C_{n_0}^l$ are independent and distributed as a fixed D^l, and $C_1^r, \ldots, C_{n_0}^r$ are independent and distributed as a fixed D^r (these distributions are also mentioned in the tables below).

In particular we made simulations for three FRVs $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ with different expectations - the above-mentioned parameters defining these FRVs are listed in Table 1. Moreover for the FRV $\mathcal{Z}$ (1) the mean, (2) examples of the simulated samples of $\mathcal{Z}$, (3) the distance between the mean V and one sample mean

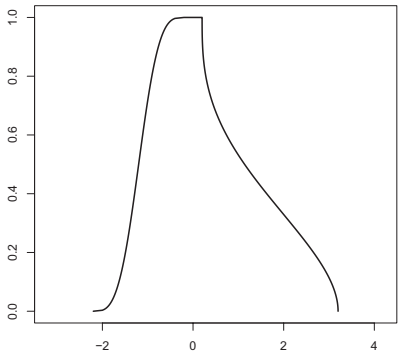


Figure 1: Expectation V of the FRV $\mathcal{Z}$

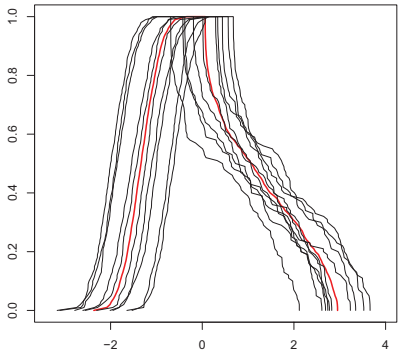


Figure 2: Simulated sample and sample mean (in red)

Table 1: Parameters defining the FRVs for the simulation

	$V = [V^c, V^l, V^r, F^l, F^r]$	$C = [C^0, D^l, D^r]$
$\mathcal{X}$	[0,1,1,Beta(1,1),Beta(1,1)]	$[U_{(-\frac{1}{2},\frac{1}{2})}, U_{(0,2)}, U_{(0,2)}]$
$\mathcal{Y}$	[4,2,2,Beta(0.5,1),Beta(3,2)]	$[U(0,8), \chi_1, \chi_2/2]$
$\mathcal{Z}$	[0,2.2,3.2,MixL,MixR)]	$[U(-1,1), U(0,2), Ex(1)]$

and (4) the empirical distribution function of the quantity $d_b^\star$ are depicted in Figure 1 - Figure 4.

Thereby MixL denotes a mixture distribution consisting of a point mass in 0 with weight 0.083 and a Beta(0.3,0.3)-distribution with weight 0.9167, and MixR denotes a mixture distribution consisting of a point mass in 0 with weight 0.067 and a Beta(3,2)-distribution with weight 0.933.

In Table 2 and Table 3 the percentage of confidence regions (among the 10,000 simulated ones) containing the populational mean in each situation is collected.

Table 2: Percentage of accurate confidence regions using Algorithm 3.1 - Part 1

	$n = 10$			$n = 30$		
θ_W	$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{Z}$	$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{Z}$
1/3	89.86	90.03	90.15	93.58	93.25	93.80
9/10	89.89	89.69	0.00	93.07	93.64	93.64

ISBN: 978-989-95079-6-8

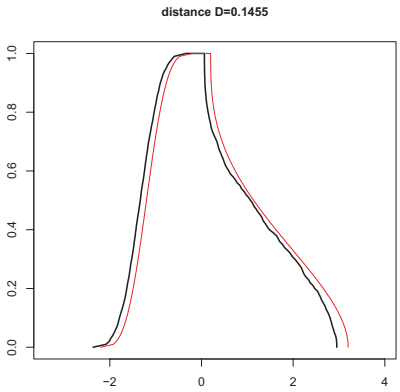


Figure 3: D_W^{φ}-distance of mean and sample mean ($\theta_W = 1/3$)

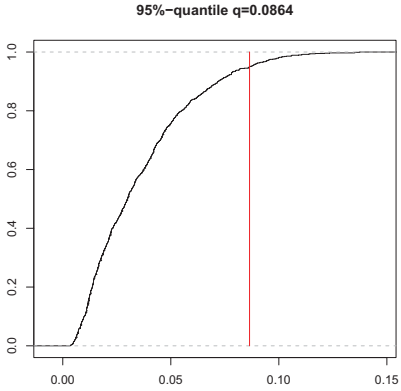


Figure 4: Empirical distribution function of $d_b^{\star}$ and 95%-quantile

Table 3: Percentage of accurate confidence regions using Algorithm 3.1 - Part 2

	$n = 100$		
θ_W	$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{Z}$
$1/3$	94.51	94.83	0.00
$9/10$	94.54	94.65	0.00

Taking a look at the results in Table 2 and Table 3 it seems that (irrespective of the underlying distribution and the selected value for the parameter θ_W), the accuracy of the confidence regions generated by means of Algorithm 3.1 is sufficient for either moderate or large sample sizes (30-100 observations), whereas the results are not satisfactory for small sample sizes (10 observations). As usual, the bigger the sample size, the closer the empirical accuracy to the theoretical one.

In order to improve the results, especially for small sample sizes, we are going to modify the proposed method. The modification we make is the usual one in classical statistics - we define the radius of the confidence ball to be directly proportional to the (estimated) population variability (standardization). For this purpose, define a new confidence ball w.r.t. the D_W^{φ}-metric as follows

$$CR_\beta = B(\overline{\mathcal{X}}, \widehat{S}\delta) = \{U \in \mathcal{F}_c(\mathbb{R}) | D_W^{\varphi}(U, \overline{\mathcal{X}})/\widehat{S} \leq \delta\}, \quad (6)$$

whereby the radius δ has to verify the coverage condition

$$P(\mu \in B(\overline{\mathcal{X}}, \widehat{S}\delta)) = P(D_W^{\varphi}(\mu, \overline{\mathcal{X}})/\widehat{S} \leq \delta) = 1 - \beta. \quad (7)$$

For the same reasons mentioned in the previous case we again use bootstrapping. In this case the variability of the population, from which the bootstrap sample is taken, is completely known (denoted by $\widehat{S}$). Consequently, one possibility consists in approximating the distribution of $D_W^{\varphi}(\mu, \overline{\mathcal{X}})/\widehat{S}$ by means of $D_W^{\varphi}(\overline{\mathcal{X}}, \overline{\mathcal{X}^{\star}})/\widehat{S}$, which, however, leads exactly to the procedure proposed in Algorithm 3.1. Another possibility is to reestimate the variability of the bootstrap population instead of using the exact known value $\widehat{S}$. We will see that this alternative approach leads to better results.
In fact, the proposed procedure works as follows:

Algorithm 3.2

Step 1. *Fix the significance level $\beta \in (0, 1)$ and the number B of bootstrap replications.*

Step 2. *Obtain B bootstrap samples $(\mathcal{X}_1^{b\star}, \ldots, \mathcal{X}_n^{b\star})$ $(b = 1, \ldots, B)$ from $(\mathcal{X}_1, \ldots, \mathcal{X}_n)$, and for each one compute its corresponding sample mean $\overline{\mathcal{X}}^{b\star}$ and its sample deviation $\widehat{S}^{b\star}$.*

Step 3. *Compute the distance between the sample mean and each bootstrap sample mean, and calculate $d_b^{\star} = D_W^{\varphi}(\overline{\mathcal{X}}, \overline{\mathcal{X}}^{b\star})/\widehat{S}^{b\star}$ for each $b = 1, \ldots, B$.*

Step 4. *Choose δ as one of the $(1 - \beta)$ quantiles of the sample $(d_1^{\star}, \ldots, d_B^{\star})$.*

In Table 4 the percentage of accurate confidence regions using Algorithm 3.2 in the same situations as in the preceding simulating study is collected. As it was expected the accuracy of the confidence regions obtained by using the second approach is much better than the one of the first approach. Indeed, the procedure can be applied also for small sample sizes ($n = 10$) and seems to be a little bit conservative (with an empirical confidence level greater than the theoretical one).

Table 4: Percentage of accurate confidence regions using Algorithm 3.2 - Part 1

	$n = 10$			$n = 30$		
θ_W	$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{Z}$	$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{Z}$
$1/3$	96.31	95.18	95.70	95.38	95.17	95.42
$9/10$	96.17	95.77	95.57	94.89	95.26	95.27

Table 5: Percentage of accurate confidence regions using Algorithm 3.2 - Part 2

	$n = 100$		
θ_W	$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{Z}$
$1/3$	94.93	95.06	0.00
$9/10$	94.94	94.92	0.00

ISBN: 978-989-95079-6-8

4 Case Study

In this section we will apply the proposed methodology in order to obtain a confidence ball for the mean overall rating of the II Summer Course organized by the European Centre for Soft Computing in July 2008. A survey regarding different aspects of this summer course was presented to the students at the end of the course. Among other questions, the students where asked about the overall rating of the course. The answers where collected directly as fuzzy sets. In fact, a scale ranging from 0 (minimum rating) to 100 (maximum rating) was presented (see Figure 5). The students were asked to draw a trapezoidal fuzzy set R representing their rating by fixing two intervals, the 0-level and the 1-level. Firstly the 0-level was chosen in such a way that their perception about the overall rating was surely not outside of this interval. Finally the 1-level was chosen as an interval in which they thought their overall rating would be contained.

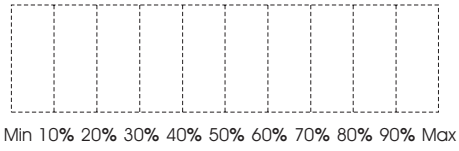


Figure 5: Scale for the made questionary

The overall rating of 33 students were collected, edited in Table 6 and the sample mean was calculated.

Applying Algorithm 3.1 and Algorithm 3.2 with $B = 10000$ bootstrap iterations and a significance level $\beta = 0.05$ we obtained the following results for the confidence balls listed in Table 7. The centre of the balls in each case is the sample mean $\overline{\mathcal{X}}$ (see Figure 6).

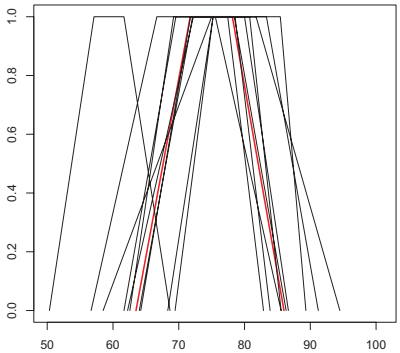


Figure 6: Thn elements of the fuzzy sample and the sample mean (red)

Figures 7 and Figures 8 depict the empirical distribution functions for the quantities $d_b^\star$ in Algorithm 3.1 and Algorithm 3.2 as well as the corresponding 95%-quantiles (for the case $\theta_W = 1/3$).

Remark: Having calculated the 95%-confidence balls for the mean overall rating, one question that naturally arises is, how this confidence balls really look like. Going back to the definition of the D_W^φ-metric it is clear that $U \in B(V, \delta)$ (ball with respect to D_W^φ), for instance, does not imply any restrictions on the support of U - in fact, the support can be arbitrarily big as long as (loosely speaking) not too many[1] α-levels are arbitrarily big. The reason for this behaviour lies in the fact that the D_W^φ-metric (because of the integration involved) averages over the distances of the α-levels (the same holds for most metrics on $\mathcal{F}_c(\mathbb{R})$). Nevertheless the condition $U \in B(V, \delta)$ (ball with respect to D_W^φ) surely implies strong restrictions on U - if for instance the spread of V is small for all α, then the spread of U can not be too big for many α's. As an example, the trapezoidal fuzzy number $(73, 73, 78, 78)$ (representation as in Table 6) is not contained in the calculated confidence Ball $B(\overline{\mathcal{X}}, 2.8239)$ since the mean spread is

Table 6: Overall ratings of the II Summer Course.

Student	$\min R_0$	$\min R_1$	$\max R_1$	$\max R_0$
1	58,50	75,00	81,75	94,50
2	64,22	71,78	78,44	86,33
3	69,44	75,22	80,78	85,56
4	64,00	72,11	85,44	89,33
5	61,67	69,56	83,33	91,22
6	56,67	66,67	75,56	85,56
7	62,22	72,22	80,00	86,67
8	50,33	57,11	61,67	68,67
9	68,33	75,22	77,44	82,89
10	62,56	69,22	78,56	83,89
11	82,22	87,33	88,11	93,44
12	50,33	60,00	64,44	74,44
13	45,56	56,67	70,00	81,11
14	69,13	75,00	89,00	94,00
15	58,89	68,89	73,11	83,11
16	51,11	62,22	63,33	77,78
17	65,00	72,50	78,50	86,25
18	51,89	61,22	68,22	86,56
19	45,89	51,67	59,44	66,33
20	74,11	80,56	88,78	95,11
21	67,89	76,22	78,33	85,56
22	74,17	85,00	88,33	97,33
23	72,22	75,00	80,44	85,33
24	51,11	60,00	71,11	78,89
25	55,56	65,56	68,89	77,22
26	76,11	80,89	90,11	93,44
27	69,89	75,11	78,00	83,22
28	80,00	90,00	90,00	100
29	71,29	79,14	82,86	89,57
30	63,78	70,89	79,44	86,67
31	63,33	74,44	85,56	95,56
32	60,00	72,11	72,11	79,89
33	78,00	83,00	87,80	92,60
Mean	63,50	71,74	78,15	86,00

Table 7: Results of the case study - calculated radius

		Radius	
θ_W	$\hat{S}$	Algorithm 3.1	Algorithm 3.2
$1/3$	8.41	2.8239	0.3587
$9/10$	8.61	2.8224	0.3476

[1] "many" in this context has to be understood as "for a set of big Lebesgue measure"

ISBN: 978-989-95079-6-8

too small.

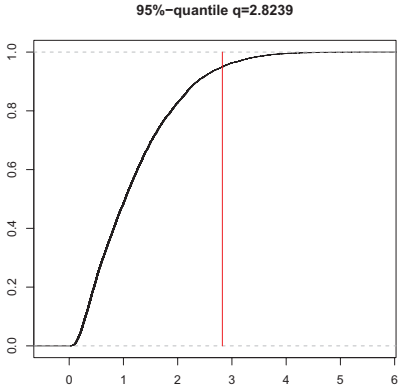


Figure 7: Empirical distribution function of $d_b^\star$ and 95%-quantile (Algoritm 3.1)

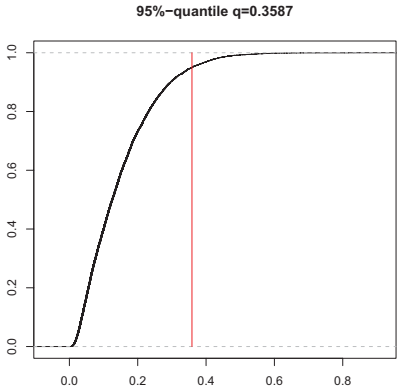


Figure 8: Empirical distribution function of $d_b^\star$ and 95%-quantile (Algoritm 3.1)

5 Conclusions and open points

In this paper we have proposed two algorithms for calculating confidence balls for the mean of an FRV - we used the sample mean as centre of the ball and calculated the radius of the ball via bootstrapping. Since we already have implemented the developed procedures (and others) in C and R our aim is to write a general R-package for statistics with fuzzy data, which could then be freely used by everybody working with R.

The simulations indicate that the second method is quite accurate, in the sense of achieving the nominal significance level for moderate and even small samples. Nevertheless, it is essential to analyze theoretically this approach in order to soundly justify its usefulness.

We have focused on a kind of percentile bootstrapping for the estimations, however it can be interesting to consider other kinds of bootstrap approaches.

Finally, for a future work we propose to analyze a possible modification of the metrics to make the estimation as reasonable as possible, by avoiding to include in the confidence region fuzzy sets which are close to the sample mean, but are not too realistic.

Acknowledgment

The research in this communication has been partially supported by the Grant from the Spanish Ministry of Education and Science MTM2006-07501. This financial support is gratefully acknowledged.

References

[1] Bertoluzza, C., Corral, N., Salas A., On a new class of distances between fuzzy numbers, Mathware & Soft Computing 2, 71–84 (1995)

[2] Colubi, A., Domínguez-Menchero, J.S., López-Díaz M., Ralescu, D. A., On the formalization of fuzzy random variables, Inform. Sci. 133, 3–6 (2001)

[3] Colubi, A., Statistical inference about the means of fuzzy random variables: Applications to the analysis of fuzzy- and real-valued data, Fuzzy Sets and Systems, 160 (3), 344-356 (2009) (doi:10.1016/j.fss.2007.12.019)

[4] Diamond, P., Kloeden, P.: Metric spaces of fuzzy sets, World Scientific, Singapore, 1994

[5] Gil, M.A., Montenegro, M., González-Rodríguez, G., Colubi, A., Casals, M.R., Bootstrap approach to the classic Oneway ANOVA test with imprecise data, Comp. Stat. Data Anal, (submitted).

[6] G. González-Rodríguez, M. Montenegro, A. Colubi, M.A. Gil, Bootstrap techniques and fuzzy random variables: Synergy in hypothesis testing with fuzzy data, Fuzzy Sets and Systems 157, 2608-2613 (2006)

[7] González-Rodríguez, G., Blanco, A., Colubi, A., Lubiano, M.A., Estimation of a simple linear regression model for fuzzy random variables, Fuzzy Sets and Systems 160 (3), 357–370 (doi:10.1016/j.fss.2008.07.007)

[8] González-Rodríguez, G., Colubi, A., Trutschnig, W., Simulation of fuzzy random variables. Information Sciences. 179 (5), 642–653 (2009) (doi:10.1016/j.ins.2008.10.018)

[9] Körner, R., An asymptotic α-test for the expectation of random fuzzy variables, J. Stat. Plann. Inference 83, 331–346 (2000)

[10] Körner, R., Näther, W., On the variance of random fuzzy variables, In: Bertoluzza, C., Gil, M.A., Ralescu, D.A. (Eds.) Statistical Modeling, Analysis and Management of Fuzzy Data, Physica-Verlag, Heidelberg, pp. 22–39 (2002).

[11] H. Kwakernaak, Fuzzy random variables, definition and theorems, Inform. Sci. 15 (1989) 1–29.

[12] M. A. Lubiano, M. A. Gil, M. López-Díaz, M. López-García, The $\overrightarrow{\lambda}$-mean squared dispersion associated with a fuzzy random variable, Fuzzy Sets and Systems 111 (2000) 307–317.

[13] Montenegro, M., Casals, M. R., Lubiano, M. A., Gil, M. A., Two-sample hypothesis tests of means of a fuzzy random variable, Inform. Sci 133, 89–100 (2001).

[14] Montenegro, M., Colubi, A., Casals, M. R., Gil, M. A., Asymptotic and Bootstrap techniques for testing the expected value of a fuzzy random variable, Metrika 59, 31–49 (2004)

[15] Puri, M. L., Ralescu, D. A., Fuzzy random variables, J. Math. Anal. Appl. 114, 409–422 (1983)

[16] Ramos-Guajardo, A.B., Colubi, A., González-Rodríguez, G., Gil, M.A., One sample tests for a generalized Fréchet variance of a fuzzy random variable, Metrika, Article in Press (doi:10.1007/s00184-008-0225-0)

[17] L.A. Zadeh, L.A., The concept of a linguistic variable and its application to approximate reasoning, Part 1, Inform. Sci. 8 (1975) 199–249; Part 2, Inform. Sci. 8 (1975) 301–353; Part 3, Inform. Sci. 9 (1975) 43–80.

ISBN: 978-989-95079-6-8

A Fuzzy Multi-Criteria Approach to WEEE Treatment Strategy Selection*

İlke Bereketli[1] Tuncay Gürbüz[2] Müjde Erol Genevois[3]

1, 2, 3 Department of Industrial Engineering, Galatasaray University
İstanbul, Turkey
Email: ibereketli@gsu.edu.tr, tgurbuz@gsu.edu.tr, merol@gsu.edu.tr

Abstract— *Today's one of the biggest environmental problem is the alarming increase of the wastes electrical and electronic equipments (WEEE). This type of waste threats the environment and the human health by contaminating air, soil and water, because of toxic materials including in the electrical and electronic equipments (EEE). That's why we should look for exploring the environmental friendly ways to dispose of these wastes and the producers should select an appropriate strategy. Waste treatment strategies contribute also to either local or global economies by creating a new sector and employment, and by reducing use of scarce resources. In this paper, we use a fuzzy hierarchical TOPSIS method for solving multi attribute group decision making (MAGDM) problems with preference information on alternatives in fuzzy environment. Our aim is to develop a fuzzy hierarchical TOPSIS model to evaluate and to select of a waste treatment strategy for electrical and electronic equipments (EEE). In this direction, three treatment strategies alternatives and six criteria are determined. The best strategy is selected and the key criterion is found.*

Keywords— Fuzzy AHP, Fuzzy Hierarchical TOPSIS, Fuzzy TOPSIS, WEEE treatment strategies.

1 Introduction

Both technological innovation and market expansion continue to accelerate the replacement of equipment leading to a significant increase of waste electric and electronic equipment (WEEE). Hence, electrical and electronic equipments, which are a subset of technological equipments, have already begun to accumulate at the garbage dumps.

Many countries have drafted legislation to improve the reuse, recycling and other forms of recovery of such wastes so as to reduce disposal [1]. The European Commission identified the need for legislation to address the escalating problem of WEEE at the Community level, and this has taken the form of the WEEE Directive. The EU's Directive 2002/96/EC about WEEE has extended producer responsibility as its main principal and aims to increase reuse and recycling ratios [2]. This Directive obligates electronic and electrical equipment producers to take back old equipment from customers free of charge and to dispose of these wastes environmentally [3]. Turkey is also on the way to prepare a new regulation about WEEE.

In this paper, a Fuzzy multi-criteria decision making (MCDM) approach is proposed for multiple attribute group decision making (MAGDM) problems with preference information on alternatives. This approach is used to evaluate and select a waste treatment strategy for electrical and electronic equipments (EEE). In this study, Fuzzy Hierarchical Technique for Order Preference by Similarity to Ideal Solution (FH-TOPSIS) is used to reflect the decision maker's subjective preference information and to determine the weight vector of criteria, and to get the rankings of different alternatives.

In the remaining part of the paper, waste treatment strategies and the case study are mentioned. Section 2 includes some definitions, alternatives and criteria required for the evaluation of the strategies. The proposed methodology is represented in Section 3. Section 4 contains the scope of the application and the results of the study. Finally in Section 5, conclusions driven from the research are provided.

2 Selection of Waste Treatment Strategies and a Case Study

WEEE, or e-waste, can be described in general as all electrical and electronic equipments that have terminated their useful life. Widmer *et al.* defines WEEE as a generic term embracing various forms of electric and electronic equipment that have ceased to be of any value to their owners, or a waste type consisting of any broken or unwanted electrical or electronic appliance [4]. In respect of EU's Directive 2002/96/EC, WEEE's definition is as follows: WEEE means electrical or electronic equipment including all components, subassemblies and consumables which are part of the product at the time of discarding [5]. So many official and non-official foundations give its own definition, therefore there is, as yet, no standard definition.

There are generally three different ways of treating WEEE: reuse, recycling and disposal (such as incineration, landfill). The EEE producers prefer at least one of these treatment strategies in order to discard their e-wastes. In this paper, it is examined that which one should be selected considering related criteria.

A case study was already realized by Bereketli *et al.*, in a telecommunication enterprise which produces and sells cell phones in Turkey [6]. A linear programming technique for multidimensional analysis of preference (LINMAP) method for solving MAGDM problems with preference information on alternatives in fuzzy environment was proposed. The aim was to develop a fuzzy LINMAP model to evaluate and to select of a waste treatment strategy for electrical and electronic equipments (EEE). At the end of the study, The

* This work is financially supported by Galatasaray University Research Fund.

ISBN: 978-989-95079-6-8

Fuzzy LINMAP method gave the ranking of the alternatives under the problems criteria as $A1 \phi\ A3 \phi\ A2$.

The objective of this study is to evaluate the alternatives considering all criteria with a different approach, Fuzzy Hierarchical TOPSIS method, and to observe the differences between the results. As we need the independence between the problem's criteria with the proposed method, some of the criteria between which there may be dependences are grouped under one criterion.

There are three alternatives: treating WEEE by reuse and recycling methods (2R), by insourced disposal methods, and by outsourced disposal methods. In the first alternative, the enterprise reuses or recycles its own WEEE within the facility. In the second alternative, the enterprise carries its own WEEE from its own waste collection store to an area which is purchased in order to dispose of the e-wastes by landfill or incineration. In the third and last alternative, the enterprise outsources its disposal process of its own WEEE to a specialized company. This company goes to the waste collection center periodically, gets there e-waste of the telecommunication enterprise, and brings them to its own facility. In its facility, it realizes disassembly of the e-waste and dissects some valuable integrated circuits. The telecommunication enterprise takes back these circuits periodically and reuses them in its assembly line. The outsourced company environment friendly disposes of the other parts of the e-waste.

It has been determined 6 criteria, 3 of them are quantitative (Type-1: T1) and 3 of them are qualitative (Type-2: T2). The criteria based on three alternatives are as follows:

- C1 - Waste release period: It is the period in which e-waste returns to the production line by passing reuse or recycling processes, or ends its life cycle by being disposed of. Small number of periods is preferred. Its unit is period time. (T1)
- C2 - Cost: This criterion includes the first investment, stock and process costs. It is the cost for creating new lines, implementing hardware, buying new machines, and new trucks for transporting e-waste, or purchasing an area where the telecommunication enterprise disposes of its e-waste, or contracting with the outsourced enterprise, stocking e-waste transported from waste collection center, every operation for each process during the period of e-waste treatment. Its unit is thousands of Turkish Liras (TL). (T1)
- C3 - Ratio of resource conservation: It is the ratio of the returned products (e-waste) quantity over total electrical and electronic equipment produced. Its unit is percentage. (T1)
- C4 - Capacity need: It is the need of labor-hour. (T2)
- C5 - Risk of damage for the nature and the human health: It is the potential of damaging the nature and the human health for each alternative when e-wastes are treated by their own way. (T2)
- C6 - Convenience to the possessed technology: It is the criteria about the degree of sufficiency of the technology that the telecommunication enterprise have. (T2)

3 Methodology

3.1 Basic concepts of Fuzzy Analytic Hierarchy Process (FAHP)

In this paper the FAHP approach is introduced, with the use of Triangular Fuzzy Numbers for pairwise comparison scale of FAHP according to the method of Chang's [7] fuzzy extent analysis by applying correct normalization formula given later by Wang *et al.* [8].

3.1.1 The construction of fuzzy judgment

In the conventional AHP, the pairwise comparisons for each level with respect to the goal are conducted using a nine-point scale proposed by Saaty [9]. According to Zadeh [10], it is very difficult for conventional quantification to define the complex situations, so the notion of a linguistic variable, whose values are words or sentences, is necessary. To assess the relative importance of the criteria and to evaluate the alternatives respecting the problems criteria an assumed weighting set has been developed.

The decision makers can use the linguistic rating set. The triangle fuzzy conversion scale of the linguistic values in the weighting set is shown in Table 1.

Table 1: The triangular fuzzy conversion scale

Very Low (VL)	(0, 3/6,1)
Low (L)	(2/3, 7/6, 5/3)
Medium (M)	(4/3, 11/6, 7/3)
High (H)	(2, 15/6, 3)
Very High (VH)	(8/3, 19/6, 11/3)

3.1.2 Value of fuzzy synthetic extent

The steps of Chang's extent analysis can be given as in the following [7]:

- First, by fuzzy arithmetic operations, take the sum of each row of the fuzzy comparison matrix.

$$RS_i = \sum_{j=1}^{n} \widetilde{a}_{ij} = \left(\sum_{j=1}^{n} l_{ij},\ \sum_{j=1}^{n} m_{ij},\ \sum_{j=1}^{n} u_{ij} \right),\quad i = 1,\mathrm{K}\ ,n \tag{1}$$

- Using fuzzy synthetic extent analysis, the value of fuzzy synthetic extent with respect to the ith object x_i, $i = 1,2,\mathrm{K}\ ,n$ that represents the overall performance of the object across all goals can be determined by the following normalization formula given by Wang et al. [8]:

$$\widetilde{S}_i = \frac{RS_i}{\sum_{j=1}^{n} RS_j} = \left(\frac{\sum_{j=1}^{n} l_{ij}}{\sum_{j=1}^{n} l_{ij} + \sum_{\substack{k=1 \\ k \neq i}}^{n} \sum_{j=1}^{n} u_{kj}},\ \frac{\sum_{j=1}^{n} m_{ij}}{\sum_{k=1}^{n} \sum_{j=1}^{n} m_{ij}},\ \frac{\sum_{j=1}^{n} u_{ij}}{\sum_{j=1}^{n} u_{ij} + \sum_{\substack{k=1 \\ k \neq i}}^{n} \sum_{j=1}^{n} l_{kj}} \right),\quad i = 1,\mathrm{K}\ ,n \tag{2}$$

ISBN: 978-989-95079-6-8

- Compute the degree of possibility of $\widetilde{S}_i \geq \widetilde{S}_j$ by the following equation where $\widetilde{S}_i = (l_i, m_i, u_i)$ and $\widetilde{S}_j = (l_j, m_j, u_j)$:

$$V(\widetilde{S}_i \geq \widetilde{S}_j) = \begin{cases} 1, & \text{if } m_i \geq m_j, \\ \dfrac{u_i - l_j}{(u_i - m_i) + (m_j - l_j)}, & \text{if } l_j \leq u_i, \quad i, j = 1, \ldots, n, i \neq j \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

- To compare $\widetilde{S}_i = (l_i, m_i, u_i)$ and $\widetilde{S}_j = (l_j, m_j, u_j)$, we need both the values of $V(\widetilde{S}_i \geq \widetilde{S}_j)$ and $V(\widetilde{S}_j \geq \widetilde{S}_i)$. The degree of possibility for a convex fuzzy number $\widetilde{S}$ to be greater than k convex fuzzy numbers $\widetilde{S}_i$, i=1,…,k can be defined by:

$$V(M \geq M_1, \ldots, M_k) = V[(M \geq M_1) \wedge \ldots \wedge (M \geq M_k)] = \min V(M \geq M_i) \quad (4)$$
$$\text{for } i = 1, 2, \ldots, k$$

- Assume that for the alternative Ai,

$$d'(A_i) = \min V(S_i \geq S_j), \text{ for } j = 1, 2, \ldots, n; j \neq i \quad (5)$$

- Then the weight vector is given by,

$$W' = (d'(A_1), d'(A_2), \ldots, d'(A_n))^T \quad (6)$$

- Via normalization, the normalized weight vectors are,

$$W = (d(A_1), d(A_2), \ldots, d(A_n))^T \quad (7)$$

3.2 Fuzzy TOPSIS

3.2.1 Basic concepts of Fuzzy TOPSIS

TOPSIS, developed by Yoon [11] and Hwang and Yoon [12] uses the following intuitive principle: The chosen alternative has to be the one having the smallest distance to the ideal solution and the biggest distance to the negative-ideal solution. But, the alternative which has the smallest distance to the ideal solution doesn't have necessarily the biggest distance to the nadir solution and vice versa. Hence, TOPSIS considers both distances to the ideal and nadir solutions simultaneously by measuring the relative closeness, RC, of the alternatives to the alternative called the « ideal solution ».

We can convert the decision matrix into a fuzzy decision matrix and construct a weighted normalized fuzzy decision matrix once the decision makers' fuzzy ratings have been collected. According to the TOPSIS model, we define the fuzzy ideal solution and fuzzy negative-ideal solution. And then a vertex method, proposed by Chen [13] is used in this paper to calculate the distance between two triangular fuzzy ratings.

Finally, the relative closeness of each alternative is defined to determine the final ranking of the alternatives.

3.2.2 Selection Procedure with Fuzzy TOPSIS

- Construct the Decision Matrix where all the data for the alternatives according the problems criteria is collected.
- Normalize the DM: To avoid the complicated normalization formula used in classical TOPSIS, the linear scale transformation is used here to transform the various criteria scales into a comparable scale.

$$\widetilde{x}_{ij} = \begin{cases} \left(\dfrac{a_{ij} - a_j^-}{c_j^* - a_j^-}, \dfrac{b_{ij} - a_j^-}{c_j^* - a_j^-}, \dfrac{c_{ij} - a_j^-}{c_j^* - a_j^-} \right) & \text{for j of a criteria of output nature} \\ \left(\dfrac{c_j^* - c_{ij}}{c_j^* - a_j^-}, \dfrac{c_j^* - b_{ij}}{c_j^* - a_j^-}, \dfrac{c_j^* - a_{ij}}{c_j^* - a_j^-}, \right) & \text{for j of a criteria of input nature} \end{cases} \quad (8)$$

- Find the weighted DM and define the ideal and the nadir alternatives namely A^+ and A^-. Note that after this normalization, the ideal (resp. nadir) alternative will be found by taking the maximum (resp. minimum) value for each criterion.
- Define distances of A_i from A^+ and A^- as follows:

$$d_i^+ = \sum_{j=1}^{n} d(\widetilde{x}_{ij}, \widetilde{x}_j^+) \text{ where } i = 1, 2, \ldots, M \quad (9)$$

$$d_i^- = \sum_{j=1}^{n} d(\widetilde{x}_{ij}, \widetilde{x}_j^-) \text{ where } i = 1, 2, \ldots, M \quad (10)$$

with

$$d(\widetilde{x}_i, \widetilde{x}_j) = \sqrt{\frac{1}{3}\left[(l_i - l_j)^2 + (m_i - m_j)^2 + (u_i - u_j)^2\right]} \quad (11)$$

- Rank the preference order of m alternatives by their relative closeness (RC) to the ideal alternative (the smaller the value of RC for an alternative, the more that alternative is preferred) which is given for the ith alternative as:

$$RC_i = \frac{d_i^+}{(d_i^+ + d_i^-)} \quad (12)$$

3.2.3 Proposed Method

In this study, the proposed method works as follows:

- Relative priorities for the criteria and the alternatives for the subjective criteria are obtained from the pairwise comparison matrices of the FAHP method using the extension principle.
- Subjective criteria ratings for the alternatives obtained in the previous step and objective criteria values are collected in a decision matrix.
- Finally Fuzzy TOPSIS is applied in order to find the ranking between the alternatives.

ISBN: 978-989-95079-6-8

4 Application

For the application of the proposed method, we asked their opinions to three decision makers. You can see in Table 2 one of the comparison matrices.

Table 2: Comparison matrices

C6	*A1*	*A2*	*A3*
A1	*	MH, H, H	VH, VH, VH
A2		*	MH, H, MH
A3			*

According to the ratings they have given, the pairwise comparison matrices are formed in order to weigh the problem's criteria and to obtain the relative priorities of the subjective criteria.

From these comparison matrices, using Fuzzy AHP method, the following results presented in Table 3 and Table 4 are obtained:

Table 3: Criteria weights

Criteria	C1	C2	C3	C4	C5	C6
Weights	.125	.198	.20	.11	.213	.15

Table 4: Relative priorities of the alternatives for the subjective criteria

	C4	C5	C6
A1	0.49	0.047	0.273
A2	0.415	0.735	0.281
A3	0.095	0.218	0.446

For the objective criteria, the values obtained by the alternatives are provided by the experts. For C1, the values were given by crisp values. On the other hand, for C2 and C3 they were agreed on giving by using the fuzzy values presented in Table 5.

As we defuzzified the fuzzy linguistic values given by the experts on the subjective criteria using Fuzzy AHP, while we apply Fuzzy TOPSIS in order to rank our alternatives, for these criteria, we will use the relative priorities given in Table 3 as crisp values. For example, the value obtained by A1 for the criteria C4, will be noted as (.49,.49,.49).

Therefore, we constructed the final decision matrix and normalized decision matrix as follows:

Table 5: Final decision matrix

	C1	C2	C3
A1	(4,4,4)	(887.5, 1095, 1207)	(30, 50, 80)
A2	(1,1,1)	(354.5, 460, 516)	(0,0,0)
A3	(2,2,2)	(174.5, 205, 231)	(3, 6, 10)
	C4	**C5**	**C6**
A1	(.49,.49,.49)	(.087,.087,.087)	(.273,.273,.273)
A2	(.415,.415,.415)	(.689,.689,.689)	(.281,.281,.281)
A3	(.095,.095,.095)	(.224,.224,.224)	(.446,.446,.446)

Table 6: Normalized decision matrix

	C1	C2	C3
A1	(0,0,0)	(0,.11,.31)	(.38,.63,1)
A2	(1,1,1)	(.67,.72,.83)	(0,0,0)
A3	(.67,.67,.67)	(.95,.97,1)	(.04,.08,.13)
	C4	**C5**	**C6**
A1	(0,0,0)	(1,1,1)	(0,0,0)
A2	(.19,.19,.19)	(0,0,0)	(.05,.05,.05)
A3	(1,1,1)	(.75,.75,.75)	(1,1,1)

From this matrix, using the weights given in Table 3, the weighted normalized decision matrix is calculated and ideal and nadir alternatives are deducted as follows:

$$A^{+} = [(.13,.13,.13)\ \ (.19,.19,.2)\ \ (.08,.,16,.,2)\ \ (.11,.11,.11)\ \ (.21,.21,.21)\ \ (.15,.15,.15)]$$

$$A^{+} = [(0,0,0)\ \ (0,.02,.06)\ \ (0,0,0)\ \ (0,0,0)\ \ (0,0,0)\ \ (0,0,0)]$$

Using the distance calculation formula given in (3.2.2) we calculated the distances of the alternatives from ideal and nadir alternatives and finally found the RC values as shown in the following table:

Table 6: Relative closeness values

	d^{-}	d^{+}	RC
A1	0.357	0.553	0.607
A2	0.272	0.639	0.701
A3	0.689	0.221	0.243

Hence the final ranking:

$$A3 \phi\ A1 \phi\ A2$$

5 Conclusions

Nowadays, the depletion of resources and the pollution are two serious problems in the world. WEEE, or e-waste for short, is one of the most influential factors on these problems. This type of waste threats the environment and the human health by contaminating air, soil and water, because of toxic materials including in the EEE. While the hazardous substances in e-waste pollute the environment, their shorter life cycle causes the fast consumption and the continuous overproduction; hence the depletion of resources increases. Therefore, a strong interest is arising in the field improving to recovery of e-waste. To cope with these problems, governments make laws about e-waste treatment and the producers begin to take the responsibility of their e-wastes. There are generally three different ways of treating WEEE: reuse, recycling and disposal.

In this paper, a Fuzzy MCDM approach is proposed for MAGDM problems with preference information on alternatives. This approach is used to evaluate and select a waste treatment strategy for EEE.

In this direction, three treatment alternatives are determined and six criteria based on these alternatives are also stated. The Fuzzy Hierarchical TOPSIS method gives the ranking

ISBN: 978-989-95079-6-8

of the alternatives under the problems criteria as $A3 \phi A1 \phi A2$. So, the best alternative is treating e-wastes by outsourcing its disposal process of its own WEEE to a specialized company. And the worst alternative is to carry its own WEEE from its own waste collection store to an area which is purchased in order to dispose of the e-wastes by landfill or incineration.

The difference of ranking between the LINMAP method and the method used in this paper may be due to the following reasons:

- Major difference between the first and third alternatives' cost. Relatively small difference for C5 and the difference in C3 for the advantage of the first alternative are not enough to compensate this huge difference of cost as well as capacity need and the convenience to the possessed technology.
- Assumption of independence between criteria caused us to unite some of the problem's criteria and as the LINMAP model doesn't require this assumption, this situation possibly caused a different ranking of the alternatives.
- Not enough consistency in expert judgments.

References

[1] J. Cui, and E. Forssberg, Mechanical recycling of waste electric and electronic equipment: a review, *Journal of Hazardous Materials*, B99, 243–263, 2003.

[2] L. Darby, and L. Obara, Household recycling behavior and attitudes towards the disposal of small electrical and electronic equipment, *Resources, Conservation and Recycling*, Vol.No.44, 17–35, 2005.

[3] M. Schmidt, A production-theory-based framework for analyzing recycling systems in the e-waste sector, *Environmental Impact Assessment Review*, Vol. No. 25, 505-524, 2005.

[4] R. Widmer, H. Oswald-Krapf, D. Sinha-Khetriwal, M. Schnellmann, and H. Böni, Global perspectives on e-waste, *Environmental Impact Assessment Review*, 25, 436-458, 2005.

[5] Directive 2002/96/Ec of the European Parliament and of the Council of 27 January 2003 on waste electrical and electronic equipment (WEEE), *Official Journal of the European Union*, L 37/24-38, 13.2.2003.

[6] İ. Bereketli, M. Erol Genevois, Y. E. Albayrak, M. Özyol, WEEE Treatment Strategies' Evaluation Using Fuzzy LINMAP Method, *Proceedings of 8th International FLINS Conference*, 647-652, 2008.

[7] D. Y. Chang, Applications of the extent analysis method on fuzzy AHP, *European Journal of Operational Research* 95, 649–655, 1996.

[8] Y.-M. Wang, Y. Luo, Z. Hua, On the extent analysis method for fuzzy AHP and its applications, *European Journal of Operational Research* 186, 735 – 747, 2008.

[9] T. L. Saaty, *The Analytic hierarchy process*, Mc. Graw-Hill, New York, 1980.

[10] Lotfi A. Zadeh. Fuzzy sets. *Information and Control*, 8:338-353, 1965.

[11] K. Yoon, *Systems selection by multiple attribute decision making*. Ph.D. dissertation. KS: Kansas State University Press, Manhattan, 1980.

[12] CL Hwang, K. Yoon, *Multiple attribute decision making: methods and applications, a state-of-the-art survey*. Berlin: Springer-Verlag, 1981.

[13] C.-T. Chen, Extensions of the TOPSIS for group decision-making under fuzzy environment, *Fuzzy Sets And Systems*, 114, 1-9, 2000.

ISBN: 978-989-95079-6-8

IFSA-EUSFLAT 2009

HIERARCHIC APPROACH IN THE ANALYSIS OF TOMOGRAPHIC EYE IMAGE

Robert Koprowski, Zygmunt Wróbel

University of Silesia, Faculty of Computer Science and Materials Science
Institute of Computer Science, Department of Biomedical Computer Systems
ul. Będzińska 39, 41-200 Sosnowiec
e-mail: koprow@us.edu.pl, wrobel@us.edu.pl.

Abstract *- The paper presents an algorithm designed to detect layers of eye's retina using an analysis in a hierarchic approach. This type approach has been implemented and tested in images obtained by means of Copernicus OCT (Optical Coherence Tomography). The algorithm created is an original approach to detect contours, layers and their thicknesses. The approach presented is an expansion of approaches described in [1], [2] and [4] and enables identification and recognition of external limiting membranes, retina and others in a very short time. The algorithm has been implemented in Matlab and C environment.*

Keywords - tomographic eye image, algorithm, recognition

1. Introduction

Images originating from a Copernicus tomograph due to its specific nature of operation are obtained in sequences of a few, a few dozen 2D images within approx. 1s, which provide the basis for 3D reconstruction. Because of their number, the analysis of a single 2D image should proceed within a time not exceeding 10 ms, so that the time of operator's waiting for the result would not be onerous (as it is easy to calculate for the above value, for a few dozen images of resolution usually MxN = 740x820 in a sequence, this time will be shorter than 1 s).

At the stage of image preprocessing (like in [1], [2] and [3]) the input image L_{GRAY} is initially subject to filtration using a median filter of (M_h x N_h) size of mask h equal to M_h x N_h=3x3 (in final version of the software this mask may be set also at M_h x N_h=5 x 5 so as to obtain better precision of algorithm operation for a certain specified group of images). Image L_M obtained this way is subject consecutively to decomposition to an image of lower resolution and analysed in terms of layers detection.

2. Image decomposition

It is assumed that the algorithm described should give satisfactory results considering mainly the criterion of operating speed. Although methods (algorithms) described in [1], [2] and [3] feature high precision of computations, however, they are not fast enough (it is difficult to obtain the speed of single 2D image analysis on a PII 1.33 GHz processor in a time not exceeding 10 ms). Therefore a reduction of image L_M resolution by approx. a half was proposed to such value of pixels number in lines and columns, which is a power of '2', i.e.: MxN=256x512 (L_{M2}) applying further on its decompositions to image L_{D16} (where symbol 'D' means decompositions, while '16' the size of block, for which it was obtained). Each pixel of the input image after decomposition has a value equal to a median of the area (block) of 16x16 size of the input image, acc. to Figure. 1.

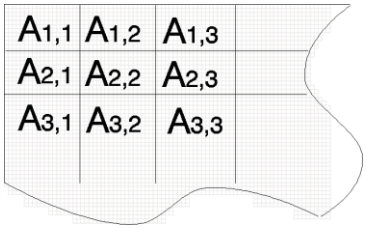


Figure. 1. Blocks arrangement in L_M image

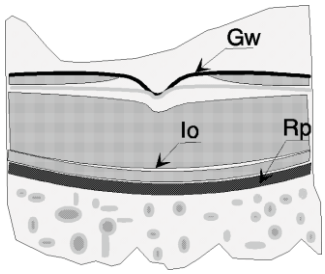


Figure. 2. Pictorial diagram of layers sought arrangement in a tomographic image

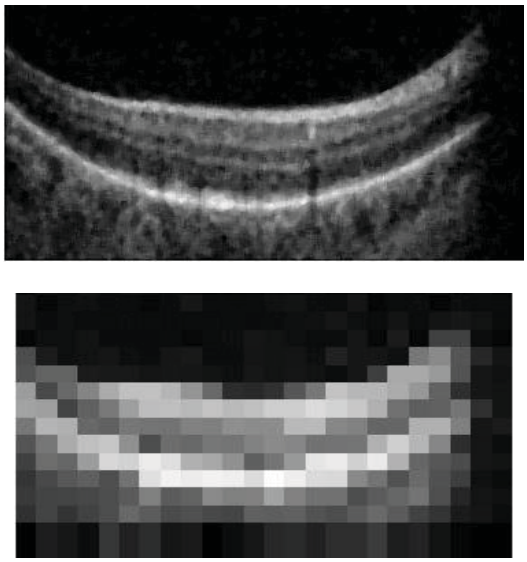


Figure. 3. Image before and after decomposition – L_{M2} and L_{D16}, respectively

ISBN: 978-989-95079-6-8

An example of L_{D16} result and the input image L_{M2} is shown in Figureure. 3. Image L_{D16} is then subject to determination of pixels position of maximum value for each column, i.e.:

$$L_{DM16}(m,n)=\begin{cases}1 & \text{if} \quad L_{D16}(m,n)=\max\limits_{m}(L_{D16}(m,n))\\ 0 & \text{others}\end{cases} \tag{1}$$

where:

m – means a line numbered from one;

n – means a column numbered from one.

Using the described method of threshold setting for the maximum value in lines, in 99 percent of cases only one maximum value in a column is obtained (Figureure. 4).

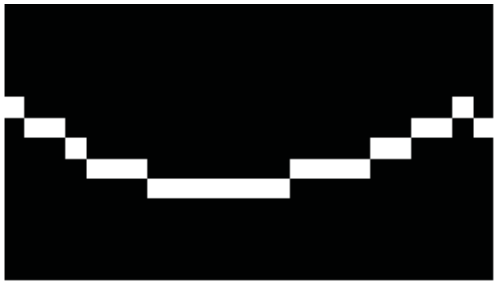


Figure. 4. Example of L_{DM16} image



Figure. 5. Example of L_{DB16} image

To determine precise position of Gw and Rp limits (Figure. 2) it turned out necessary to use one more L_{DB16} image, i.e.:

$$L_{DB16}(m,n)=\begin{cases}1 & \text{if} \quad |L_{D16}(m,n)-L_{D16}(m+1,n)|>p_r\\ 0 & \text{others}\end{cases} \tag{2}$$

for $m\in(1,M-1)$, $n\in(1,N)$,

where:

pr – threshold assumed within the range (0, 0.2).

As a result, coordinates of Gw(n) and Rp(n) limits position points are obtained as such positions of values '1' in L_{DB16} image, for which $Gw(n) \le Rp(n)$ and Rp(n) is obtained from L_{DB16} image in the same way.

This method for pr threshold selection at the level of 0.01 gives satisfactory results in around 70 percent of cases of not composed images (i.e. such, which are not images with a visible pathology). Unfortunately for the other 30 percent cases the selection of pr threshold in the adopted limits does not reduce the originated errors (Figure. 5).

The correction on this level of erroneous recognitions of Gw(n) and Rp(n) layers is that important, that for this approach these errors will not be duplicated (in the hierarchic approach presented below) for the subsequent more precise approximations.

3. Correction of erroneous recognitions

In L_{DB16} image (Figure. 5) white pixels are visible in an excess number for most columns. Two largest objects arranged along 'maxima' in columns entirely coincide with Gw and Rp limits position. Based on that and having carried out the above analysis for a few hundred images, the following limitations were adopted:

- for coordinates Rp(n) found in L_{DM16} image there must be at the same time $L_{DM16}(m,n)=1$ in other cases this point is considered as disturbance or as a point of Gw(n) layer,
- if only one pixel of value '1' occurs in image L_{DM16} and L_{DB16} for the same position, i.e. for the analysed n there is $L_{DM16}(m,n) = L_{DB16}(m,n)$ the history is analysed for $n>1$ and it is checked, whether $|Gw(n-1) - Gw(n)| > |Rp(n-1) - Rp(n)|$, i.e.:

$$Rp(n)=\begin{cases}m & \text{if} \quad L_{DB16}(m,n)=L_{DM16}(m,n)=1 \wedge \\ & \wedge |Gw(n-1)-Gw(n)|>|Rp(n-1)-Rp(n)|\\ 0 & \text{others}\end{cases} \tag{3}$$

for $m\in(1,M-1)$, $n\in(2,N)$,

- if $|Gw(n-1) - Gw(n)| \le |Rp(n-1) - Rp(n)|$, the condition $Gw(n-1) - Gw(n) = \pm 1$ is checked (giving thereby up fluctuations against history n-1 within the range ± 1 of area A (Figure. 1). If so, then this point is the next Gw(n) point. In the other cases the point is considered as a disturbance. It is assumed that lines coincide Gw(n) = Rp(n) if $Rp(n-1) - Rp(n)=\pm 1$ and only one pixel occurs of value '1' in L_{DM16} image.
- in the case of occurrence in specific column of larger number of pixels than 2, i.e. if $\operatorname{sum}_{m}(L_{DB16}(m,n))>2$ a pair is matched (if occurs) Gw(n-1), Rp(n-1) so that $|Gw(n-1) - Gw(n-1)| - |Rp(n) - Rp(n)| = \pm 1$ would occur. In this case it may happen that lines Gw(n) and Rp(n) will coincide. However, in the case of finding more than one solution, that one is adopted, for which $L_{D16}(Gw(n),n) + L_{D16}(Rp(n),n)$ assumes the maximum value (the maximum sum of weights in L_{D16} occurs).

The presented correction gives for the above class of images the effectiveness of around 99% of cases. Despite adopted limitations the method gives erroneous results for the initial value n=1, unfortunately these errors continue to be duplicated.

ISBN: 978-989-95079-6-8

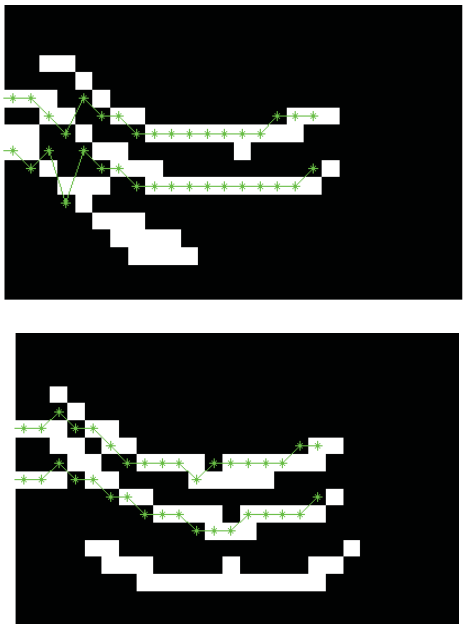


Figure. 6. Examples of L_{DB16} images for p_r=0.01 with incorrectly marked Gw(n), Rp(n) points (layers)

Unfortunately, the adopted relatively rigid conditions of acceptable difference |Gw(n-1) - Gw(n-1)| or |Rp(n)-Rp(n)| cause origination of large errors for another class of tomographic images, in which a pathology occurs in any form (Figure. 7).

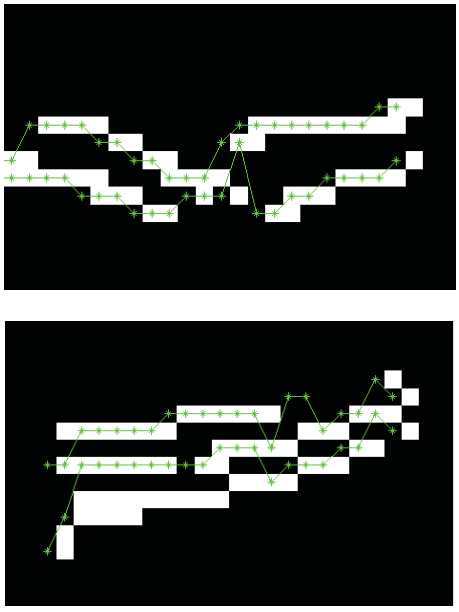


Figure. 7. Examples of L_{DB16} images for p_r=0.01 with incorrectly marked Gw(n), Rp(n) points (layers)

As it may be seen in Figure. 6 and Figure. 7 problems occur not only for the initial n values, but also for the remaining points. The reason of erroneous recognitions of layers positions consists of difficulty in distinguishing proper layers in the case of discovering three 'lines', three points in a specific column, which position changes in acceptable range for individual n.

These errors cannot be eliminated at this stage of decomposition into 16x16 pixels areas (or 16x32 image resolution). They will be the subject of further considerations in the next sections.

4. Reducing the decomposition area

The increasing of accuracy and thereby reducing the $A_{m,n}$ area size (Figure. 1) – block in L_M image – is a relatively simple stage of tomographic image processing with particular focus on the operating speed. It has been assumed that $A_{m,n}$ areas will be sequentially reducing by half in each iteration – down to 1x1 size. The reduction of $A_{m,n}$ area is equivalent to performance of the next stage of lines Gw and Rp position approximation.

The increasing of accuracy (precision) of Gw and Rp lines position determined in the previous iteration is connected with two stages:

- concentration of (m,n) coordinates in the sense of determining intermediate ((m,n) points situated exactly in the centre) values by means of linear interpolation method;
- change of concentrated points position so that they would better approximate the limits sought.

If the first part is intuitive and results only in resampling, the second requires more precise clarifications. The second stage consists in matching individual points to the layer sought. As in the ox axis the image by definition is already decomposed and pixel's brightness in the image analysed corresponds to the median value of the original image in window A (Figure. 1), the modification of points Rp and Gw position occurs only in the vertical axis. The analysis of individual Rp and Gw points is independent in the sense of dependence on n-1 point position, as was the case in the previous section.

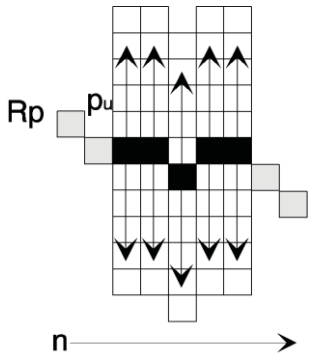


Figure. 8. Pictorial diagram of the process of Rp course matching to the edge of the layer sought. Individual pixels independent of each other may change the position within the $\pm p_u$ range

Each of Rp points, left from the previous iteration, and newly created from interpolation, in the consecutive algorithm stages is matched with increasingly high precision to the RPE layer. Point's Rp(n) position changes within the range of $\pm p_u$ (Figure. 8), where the variation range does not depend on the scale of considerations (size of A area) and strictly results from the distance between Gw and Rp (Figure. 2). For blocks A of size from 16 x 16 to 1 x 1 p_u is constant and equal 2. This value has been taken based on typical average, for analysed a few hundred L_{GRAY} images, distance between

ISBN: 978-989-95079-6-8

Gw and Rp equal to around 32 pixels, what means that after decomposition into blocks A of size 16x16 these are two pixels, i.e. p_u=2. In this ±2 range a maximum is sought in L_{DM} image and a new position of point Rp or Gw assumed for it. Thus the course of Rp or Gw is closer to the actual course of the layer analysed.

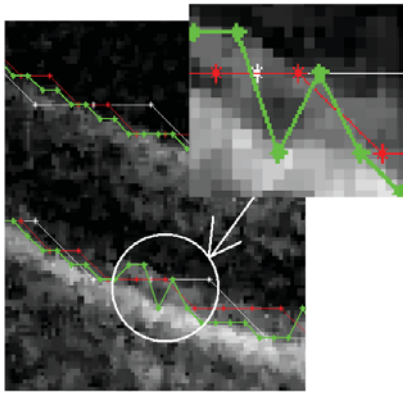


Figure. 9. Results of matching for two iterations White colour marks input Rp points and red and green – consecutive approximations

The obtained results of matching are presented in Figure. 9. White colour shows input Rp values as input data for this stage of algorithm and decomposition into A blocks of size 16 x 16 (L_{DM16} and L_{DB16} images), red colour – results of matching for A blocks of size 8 x 8 (L_{DM8} and L_{DB8} images), and green colour – results of matching for A blocks of size 4 x 4 (L_{DM4} and L_{DB4} images). As may be seen from Figure. 9 the next decompositions into consecutive smaller and smaller A areas and thus image of higher resolution, a higher precision is obtained at the cost of time (because the number of analysed Rp(n), Gw(n) points and their neighbourhoods ±p_u increases).

This method for A of 16 x 16 size has that high properties of global approach to pixels brightness that there is no need to introduce at this stage additional actions aimed at distinguishing layers situated close to each other (which have not been visible so far due to image resolution). While at A areas of 4x4 size other layers are already visible, which should be further properly analysed. At increased precision, Io layer is visible, situated close to Rp layer (Figure. 9). Thereby in the area marked with a circle there is a high position fluctuation within the oy axis of Rp layer. Because of that the next step of algorithm has been developed, taking into account separation into Rp and Io layers for appropriately high resolution.

5. Analysis of Rp and Io layers

The analysis of layers consists in separating line Io from line Rp originating from previously executed stages of the algorithm. The case is facilitated by the fact that on average around 80, 90% of pixels in each tomographic image has a maximum value in each column exactly in point Rp (this property has been already used in the first section). So the only problem is to detect the position of Io line. One of possible approaches consists of an attempt to detect the contour of the layer sought in L_{IR} image. This image originated from L_M image thanks to widening of Rp(n) layer range within oy axis within the range of ±p_I=20 pixels. L_{IR} image has been obtained with the number of columns consistent with the number of L_M image columns and with the number of lines $2 \cdot p_I+1$.



Figure. 10. Image $L_{IR} = L_M(m\text{-}R_p(n),n)$

Figure. 10 shows image $L_{IR} = L_M(m\text{-}Rp(n),n)$ originating from L_M image from Figure. 9.

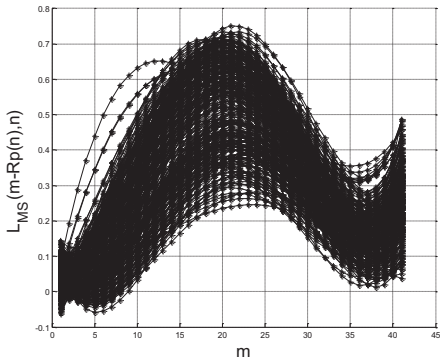


Figure. 11. Courses $L_{IRS} = L_{MS}(m\text{-}Rp(n),n)$ versus m

The upper layer visible in Figure. 10 as a pretty sharp contour is the sought course of Io. Unfortunately, because of a pretty high individual variation within the Io layer position relative to Rp the selected p_I range in further stages of the algorithm may be increased even twice (that will be described later). To determine consecutive points of Io layer position interpolations with 4th order polynomial of grey level degree for individual columns of L_{IR} image obtaining this way L_{IRS}, which changes of grey levels in individual columns are shown in Figure. 11. The position of point Io(n) occurs in the place of the highest gradient occurring within the range (Rp(n)- p_I) ÷ Rp(n) relative to L_{MS} image or 1÷ p_I relative to L_{IRS} image.

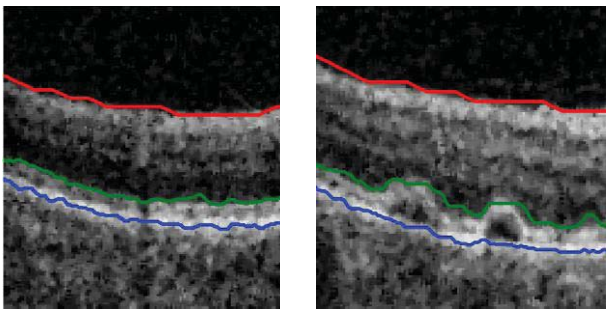


Figure. 12. Parts of L_M images with marked courses Gw – red, Rp – blue, and Io – green

As may be seen in Figure. 12, the method presented perfectly copes with detecting Gw, Rp and Io layers marked in red, blue and green, respectively.

ISBN: 978-989-95079-6-8

6. Layers thickness map and 3D reconstruction

The analysis of L_M images sequence and precisely the acquiring of Gw, Rp and Io layers allows performing 3D reconstruction and layers thickness measurement. A designation for an image sequence with an upper index (i) has been adopted, where i = {1,2,3,...,k-1,k) i.e. $L_M^{(1)}$, $L_M^{(2)}$, $L_M^{(3)}$,.., $L_M^{(k-1)}$, $L_M^{(k)}$. For a sequence of 50 images the position of Gw layers (Figure. 13), Rp (Figure. 14) and Io (Figure. 15) was measured as well as Io - Rp layer thickness (Figure. 16).

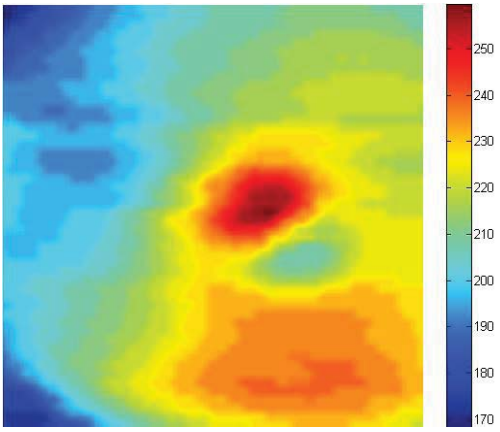


Figure. 13. Spatial position of Gw

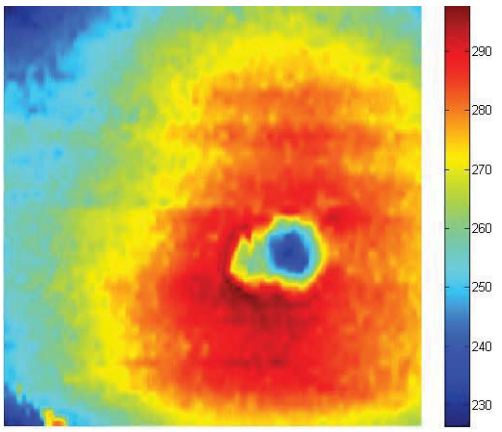


Figure. 14. Spatial position of Rp

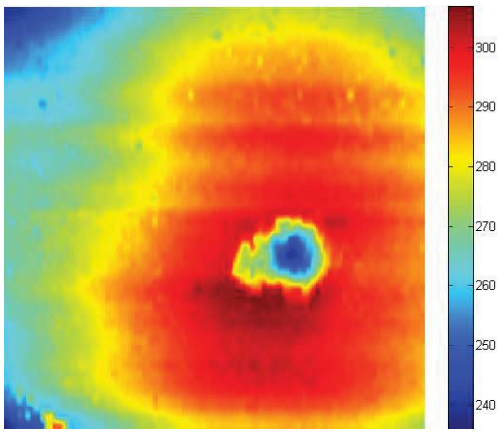


Figure. 15. Spatial position of Io

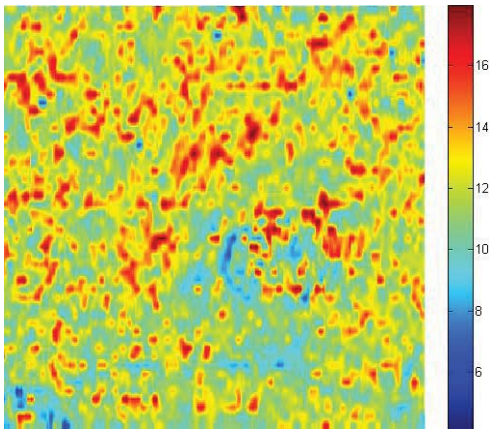


Figure. 16. Thickness of Io - Rp layer

3D reconstruction performed based on $L_M^{(i)}$ images sequence is the key element crowning the results obtained from the algorithm suggested. The sequence of images, and more precisely the sequence of $Gw^{(i)}(n)$, $Rp^{(i)}(n)$ and $Io^{(i)}(n)$ layers position provides the basis for 3D reconstruction of a tomographic image. For an example sequence of 50 images and one $L_M^{(i)}$ image resolution of MxN= 256 x 512 a 3D image is obtained, composed of three Gw, Rp and Io layers of 50 x 512 size.

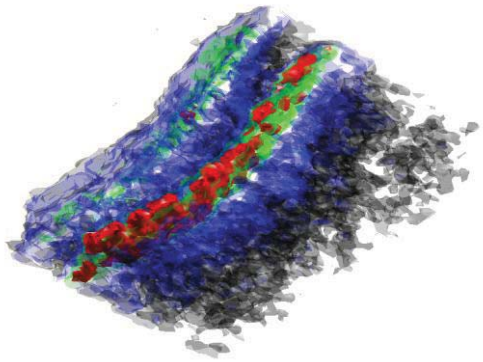


Figure. 17. Example 3D reconstruction of layers: Gw – blue, Rp – red, and Io - green

In an obvious way a possibility of automatic determination of the thickest or the thinnest places between any points results from layers presented in Figure. 18.

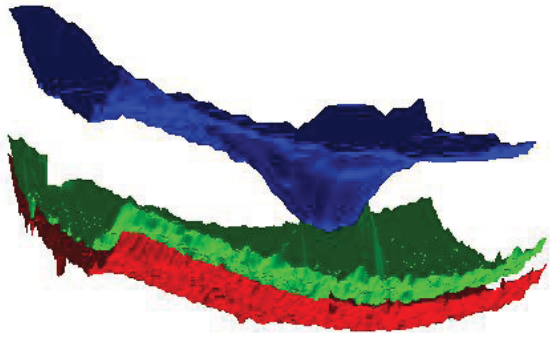


Figure. 18. Example 3D reconstruction of layers: Gw – blue, Rp – red, and Io - green

ISBN: 978-989-95079-6-8

Results are shown in Figure. 17 for an example reconstruction of original images (without processing described above) based on pixels brightness and in Figure. 18 – the reconstruction performed using the algorithm described above was carried out based on $Gw^{(i)}(n)$, $Rp^{(i)}(n)$ and $Io^{(i)}(n)$ information.

7. Summary

The algorithm presented detects Gw, Rp and Io layers within up to 50 ms time on a PC with a 2.5 GHz Intel Core 2 Quad processor. The time was measured as a mean value of 700 images analysis dividing individual images into A blocks (Figure. 1) consecutively of sizes 16 x 16, 8 x 8, 4 x 4, 2 x 2. This time may be reduced modifying the number of approximation blocks and at the same times increasing the layer position identification error – results are presented in the table below.

Table 1. Execution time of algorithm to find Gw, Rp and Io layers and 3D reconstruction

Processing stage	Total time since processing start [ms]	Time of individual stages computations [ms]
Preprocessing	10.10	10.10
Preliminary breakdown into Gw and Rp+Io	13.20	3.20
Gw and Rp+Io approximation for A – 16x16	15.90	2.74
Gw and Rp+Io approximation for A – 8x8	23.70	7.63
Precise Rp and Io breakdown	44.60	20.85

The specification of individual algorithm stages' analysis times presented in the table above clearly shows the longest execution of the first stage of image preprocessing, where filtration with a median filter is of prevailing importance (in terms of execution time) as well as of the last stage of precise determination of Rp and Io layers position. Because precise Rp and Io breakdown is related to the analysis and mainly to the correction of Rp and Io points position in all columns of the image for the most precise approximation (because of a small distance between Rp and Io it is not possible to perform this breakdown in earlier approximations). So the reduction of computation times may occur only at increasing the error of layers thickness measurement. And so for example for the analysis in the first approximation for A of 32 x 32 size and then for 16 x 16 gross errors are obtained generated in the first stage and duplicated in the next ones. For approximations for A of 16 x 16 and then 8 x 8, 4 x 4, 2 x 2 and 1 x 1 sizes the highest accuracy is obtained, however the computation time increases approximately twice.

References

1 Koprowski R., Wróbel Z.: Layer Analysis in a Tomographic Eye Image (in Polish), Proceedings Conference Decision Support Systems, Zakopane Poland (2007).

2 Koprowski R., Wróbel Z.: Identification of layers in a tomographic image of an eye based on the canny edge detection., Advances in Soft Computing 47, 2008, pp. 232-239

3 Koprowski R., Wróbel Z.: Layers Recognition in Tomographic Eye Image Based on Random Contour Analysis – sent to conference 2009 IFSA World Congress - 2009 EUSFLAT Conference

4 Koprowski R., Wróbel Z.: Determining the contour of cylindrical biological objects using the directional field. Proceedings 5th International Conference on Computer Recognition Systems CORES-07, Advances in Soft Computing, (2007)

5 Wróbel Z,. Koprowski R.: Automatic methods of microtubules orientation analysis (In Polish), Wydawnictwo UŚ, (2007).

ISBN: 978-989-95079-6-8

A Two Step Solution Procedure to a Fuzzy Medical Waste Disposal Facility Location Problem

H. Ziya Ulukan Yeşim Kop

Industrial Engineering Department, Galatasaray University

Ciragan Cad. No:36 Ortakoy, Istanbul, Turkey
Email: zulukan@gsu.edu.tr, ykop@gsu.edu.tr

Abstract— *This paper handles the medical waste disposal facility location problem in Istanbul by using fuzzy TOPSIS (FETOPSIS) to select the adequate place between some candidate points that are obtained from the Ishii's undesirable facility location algorithm. Medical waste must be disposed without damaging environment and human health, complying with new regulations. The aim is to provide an alternative method to the medical waste disposal facility location problem in Istanbul, able to handle the fuzziness of the real world. People attempt to minimize the undesirable effects introduced by the new facility by maximizing its minimum Euclidean distance with respect to all demand points and also to minimize the total transportation costs. This study proposes a modified model of Ishii's algorithm to obtain a list of candidate points. Then a selection is made between these candidate points considering other criteria as earthquake risk, distance to natural water sources and level of air pollution.*

Keywords— Undesirable Facility Location, Maximin, FETOPSIS, Medical Waste, Euclidean Distances.

1 Introduction

The infectious, pathological, cutter-driller wastes emanating from medical institutions are known as medical wastes. Recently, the controlled collection and safe disposal of medical waste has become an important field of environment protection in Turkey as in whole world. The 50% of environmental pollution has occurred in the last 35 years. As a result, the scientists work on eliminating all harmful factors on air, earth, water and human life. The safe disposal and the regain of waste are some of the important fields that they work on. Even if the medical waste consist a small part of waste amount, when we consider the threat they constitute on human life, their safe disposal is vital.

The present position is that medical wastes are stored at special hazardous waste storage areas or medical waste disposal areas if they cannot be disposed by incineration. The technical criteria concerning storage are included in the By-Law on the Control of Medical Wastes (2005).

Throwing away without control or common disposal with household waste poses a serious threat for all livings. That is why medical waste must be disposed without damaging the human mental and physical health, the animal health, the flora, the water and the welfare of the society, complying with new regulations.

The aim of this study is to provide an alternative method to solve the medical facility location problem in Istanbul, able to handle the fuzziness of transportation costs and people preferences. Nuclear plant, oil refining plant, waste disposal plant must not be close to residential area. People attempt to minimize the undesirable effects introduced by the new facility by maximizing its minimum Euclidean distance with respect to all demand points and also to minimize the total transportation costs. These are semi-obnoxious facilities and to locate them, facility planners determined two objectives [1]. The first objective aims to maximize the minimum distance from the new facility to the demand points; this is the maximin problem. The second aim is to minimize the total distance from the facility to the demand points in order to minimize the transportation costs. In this study, we propose the following two-step solution procedure where we obtain all efficient solutions at the first step and then we choose the best solution at the second step. Considering the fuzzy nature of the people attitude towards the location of this kind of facilities we will try to find the site of the facility which maximizes the minimal satisfaction degree among all demand points and maximizes the preference of the site by using a method based on the Ishii's where the attitudes of people are expressed by a trapezoidal membership function [2]. The function represents the satisfaction degree of demand points with respect to the distance from these points to the facility site. We reformulate the Ishii's model to obtain a list of candidate points. Then, we make a selection between these candidate points by solving a fuzzy TOPSIS (FETOPSIS). The use of this method allows us to consider other criteria as earthquake risk, distance to natural water sources and level of air pollution. These criteria will be evaluated by experts with the first two criteria: distance to districts and transportation costs.

2 Disposal of medical waste

There are many proven techniques for the safe disposal of medical waste all around the world. Nowadays, new contagious diseases that appeared in diverse countries induced all medical authorities to take severe precautions. In 2003, with the outbreak of severe acute respiratory syndrome (SARS) the authorities take more serious steps in managing medical waste. The procedures for handling, treatment and disposal of this waste were required to comply with the most stringent standards [3, 4]. The health authority in Taiwan had to handle the current status of waste

ISBN: 978-989-95079-6-8

production for further management planning. The quantity of waste generated by hospitals varies by changes in local legislation according to the studies [5, 6, 7, 8].

In Turkey, the medical waste production increases due to the economic and social changes and the population increase. The amount of waste is too huge to dispose in the dumping ground and requires an integrated management concept involving collection, transportation and disposal. By 2005 the number of hospital in Turkey has become 1198 and due to development of medical technologies and hygienic precautions, total waste amount has been increasing. According to MEF's researches, total waste amount in Turkey is 238.26 tons per day and 86 968 tons per year. 23 000 tons of this amount is collected in Istanbul by ISTAC (sub institution of Istanbul Metropolitan Municipality) and district municipalities.

The mortal contagious diseases that appeared recently in Turkey as Avian Influenza, as well-known name bird flu, Crimean-Congo hemorrhagic fever (CCHF) and the holocausts related to these diseases, furthermore the increase of other contagious diseases as aids, hepatitis and tuberculosis accentuate the vitality of the collection and the safe disposal of medical waste.

The regulations of medical waste control inure on 20.07.2005 in Turkey. The regulations are based on the 8th article of the environment code where it is denoted that inappropriate collection, transportation and disposal are forbidden. In the 35th article of regulations there is a legal decision on the medical waste disposal facility location: The distance of disposal facility can't be closer than 1000 meters to the residential area. As it is mentioned clearly in the regulations these following criteria must be considered in the decision making process:

1. Distance between disposal facility and districts. (C_1)
2. Transportation costs. (C_2)
3. Earthquake risk. (C_3)
4. Distance to the water sources. (C_4)
5. Existing air quality. (C_5)

The first two quantitative criteria will be handled by using a fuzzy facility location method so as to obtain worthwhile candidate points. Then these points will be handled by a fuzzy TOPSIS method considering the five criteria above all together to select a single point.

3 Methodology

3.1 Fuzzy Facility Location with Preference of Candidate Sites

Facilities as power plants, chemical plants, dumping grounds, airports are undesired close to residential area but they must also be at a reasonable distance easy to reach in order to minimize transportation costs. By using different decision making methods one must compromise two objectives. People don't want to live near a dumping ground but want also to get rid of the waste they produce as fast as possible. So the facility can not be placed to an inaccessible distance.

The first step of this study is based on the [2]'s fuzzy facility location problem, where the attitudes of the demand points (people, clients) towards the location of new facility are expressed by a trapezoidal membership function. Ishii's problem is reformulated to obtain a list of candidate points. Reference 2 categorizes people attitudes in three categories. In this study we have two categories of attitude: 1. People don't want the facility near. 2. People don't want high costs of transportation. As it is mentioned in the medical waste regulations, the burning and storage facility can not be close more than 1000 meters. Anyway people prefer the facility to be more than 5000 meters far from their residential area.

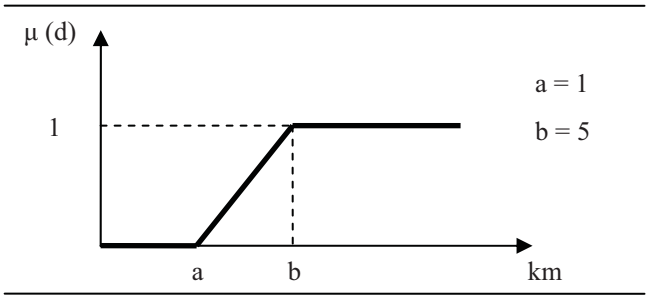


Figure 1: Membership function related to distances.

$$d_i(X) = \begin{cases} E(x-p_i)+N(y-q_i) & (x \ge p_i, y \ge q_i) \\ W(x-p_i)+N(y-q_i) & (x \le p_i, y \ge q_i) \\ E(x-p_i)+S(y-q_i) & (x \ge p_i, y \le q_i) \\ W(x-p_i)+S(y-q_i) & (x \le p_i, y \le q_i) \end{cases} \quad (1)$$

$$\mu(d_i) = \begin{cases} 0 & (d_i \le a) \\ \dfrac{d_i - a}{b-a} & (a \le d_i \le b) \\ 1 & (b \le d_i) \end{cases} \quad (2)$$

N: North = 1; *S*: South = –1; *E*: East = 1; *W*: West = –1 are constants used to provide nonnegative differences between coordinates. Let be n demand points in the rectangular area. These points are represented as (p_i, q_i), $i = 1, 2, \ldots, n$. The facility that we will build is represented as $\chi = (x, y)$ and $X = \{(x, y) | p_L \le x \le p_U,\ q_L \le y \le q_U\}$ [2]. Site $\chi \in X$.

Define α: α = $\mu(d_i)$. Then, we model the problem as [2]:

Maximize α

such that

$$\alpha \le \frac{d_i(x) - a}{b - a}$$
$$0 < \alpha \le 1$$
$$p_L \le x \le p_U$$
$$q_L \le y \le q_U \quad (3)$$

But the way that we will pursue to solve this problem is to partition the rectangular area to small rectangles by tracing the horizontal and vertical lines passing by each demand

ISBN: 978-989-95079-6-8

points. For each line intersection point called x_{jm} we will solve the sub-problem (4). The optimums of each sub problem can be handled by another decision making tool in order to select the best location for the facility.

$$\begin{aligned} Q:&\ \text{Maximize}\ \alpha \\ &\text{such that}\ \ \alpha(b-a)+a \le E(x-p_i)+N(y-q_i) \\ &\alpha(b-a)+a \le W(x-p_i)+N(y-q_i) \\ &\alpha(b-a)+a \le E(x-p_i)+S(y-q_i) \\ &\alpha(b-a)+a \le W(x-p_i)+S(y-q_i) \\ &0<\alpha\le 1 \qquad x\in X \qquad i=1,2,...,\ n \end{aligned} \tag{4}$$

3.2 Selection of the Best Candidate Site with Fuzzy TOPSIS

Method TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) was developed by Hwang and Yoon (1981) as an alternative to the ELECTRE method. The basic concept of the method is that the selected alternative should have the shortest distance to the ideal solution and the farthest distance to the negative-ideal solution. The Euclidian distance approach was proposed to evaluate the relative closeness of the alternatives to the ideal solution [4]. It solves the dilemma of the choice between ideal and anti-ideal by using an idea that Dasarathy (1976) applies to the data analysis. The TOPSIS method evaluates the decision matrix which refers to m alternatives which are evaluated in terms of n criteria [9].

In real-world situation, because of incomplete or non-obtainable information, (for example, human judgments including preferences are often vague and cannot estimate his preference with an exact numerical data), the data are not deterministic; therefore they usually are fuzzy / imprecise, so, TOPSIS for fuzzy data is used [10]. The main advantage of fuzzy formulation compared to the crisp formulation is that the decision maker is not obligated to give a precise formulation, for the sake of mathematical reasons, even though he or she might be able or willing to describe the problem in fuzzy terms [11].

The extension of TOPSIS to fuzzy TOPSIS provides a new multi-criteria decision making method compatible with the real world decisions. There are diverse applications of this method in the literature as the evaluation of airline service quality of [12], selection of expatriate host country of [13], bridge risk assessment of [14], new product introduction of [15], industrial robotic system selection of [16] etc.

Let $A_1, A_2,..., A_m$ be m alternatives among which we will make the selection, $C_1, C_2,..., C_n$ be n the criteria that are under consideration during the decision making process. $\tilde{x}_{ij}$ is the fuzzy rating of alternative A_i according to the criterion C_j. Fuzzy data used here is triangular fuzzy number. We can express this fuzzy multi criteria decision making problem in matrix format [10]:

$$\begin{array}{cccccc} & C_1 & C_2 & \cdots & C_n \\ A_1 & \tilde{x}_{11} & \tilde{x}_{12} & \cdots & \tilde{x}_{1n} \\ A_2 & \tilde{x}_{21} & \tilde{x}_{22} & \cdots & \tilde{x}_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ A_m & \tilde{x}_{m1} & \tilde{x}_{m2} & \ldots & \tilde{x}_{mn} \end{array} \tag{5}$$

The purpose of linear scales transformed with normalization function used in this study is to preserve the property that the ranges of normalized triangular fuzzy numbers to be included in (0,1). [17]:

$$\tilde{r}_{ij}=\left(\frac{a_{ij}}{c_j^*};\frac{b_{ij}}{c_j^*};\frac{c_{ij}}{c_j^*}\right) \tag{6}$$

Here, $c_j^*=\max_i c_{ij}$

Thereby, the normalized fuzzy decision matrix $\tilde{R}=\left[\tilde{r}_{ij}\right]_{m\times n}$ is generated. In decision making process each criterion can have a different importance for the decision maker who will assign different weights to each criterion. But in our study we assume that all criteria are equally important.

We define the positive ideal solution A^* and the negative ideal solution A^- [17]:

$$A^*=\left(\tilde{r}_1^*,\tilde{r}_2^*,\ldots,\tilde{r}_n^*\right) \tag{7}$$

$$A^-=\left(\tilde{r}_1^-,\tilde{r}_2^-,\ldots,\tilde{r}_n^-\right) \tag{8}$$

Then, we have to calculate the distance of each alternative to A^* and A^-, by using the formula of distance between two fuzzy numbers.

$$d_i^*=\sum_{j=1}^{n} d\left(\tilde{r}_{ij},\tilde{r}_j^*\right), i=1,\ldots,m \tag{9}$$

$$d_i^-=\sum_{j=1}^{n} d\left(\tilde{r}_{ij},\tilde{r}_j^-\right), i=1,\ldots,m \tag{10}$$

The distance between two fuzzy numbers is calculated as:

$$\tilde{x}=\left(a_1,b_1,c_1\right),\ \ \tilde{y}=\left(a_2,b_2,c_2\right) \tag{11}$$

$$d\left(\tilde{x},\tilde{y}\right)=\sqrt{\frac{1}{3}\left[\left(a_1-a_2\right)^2+\left(b_1-b_2\right)^2+\left(c_1-c_2\right)^2\right]} \tag{12}$$

A relative closeness index between d_i^* and d_i^- is calculated to determine the ranking order of alternatives [17,10]:

$$R_i=\frac{d_i^-}{d_i^*+d_i^-},\ \ i=1,\ldots,m \tag{13}$$

ISBN: 978-989-95079-6-8

4 Application

This two-step model is applied to the medical waste facility selection for the most populated region of Istanbul including the districts Şişli, Kağıthane, Beşiktaş and Sarıyer. The figure 2 is the Google Earth map showing the application region. The distances and coordinates are provided by Google Earth.



Figure 2: Selection of a medical waste disposal facility in Istanbul.

In rectangular area there are four demand points and all lines passing through form 25 points which means that we will solve the sub problem (4) 25 times by using LINDO software. The optimum membership values and optimum candidate facility sites are in the Table 1. As our aim is to maximize membership value α, we can select the biggest α value as new facility site but there are many α values closer to each other. So, the most reasonable attitude is to handle some candidate points having big α value, considering other criteria. In this study, we handle seven points as alternatives having α value bigger than 0.600 in terms of an observation made on the map.

Table 1: Optimum α values and optimum candidate facility sites

SUB PROBLEMS	OPTIMUM POINT	OPTIMUM α VALUE
1	-	-
2	(1.790; 3.450)	0.110
3	(6.350; 0.000)	0.140
4	(5.330; 7.110)	0.167
5	(5.080; 2.160)	0.217
6	(5.080; 3.645)	0.261
7	(5.330; 3.645)	0.324
8	(4.445; 2.160)	0.376
9	(5.080; 2.795)	0.376
10	(5.080; 2.795)	0.376
11	(3.230; 0.000)	0.410
12	(6.350; 2.160)	0.410
13	(2.675; 0.540)	0.413
14	(6.350; 7.110)	0.422
15	(2.675; 0.000)	0.549
16	(0.000; 0.565)	0.565
17	(3.230; 7.110)	0.567
18	(6.350; 3.645)	0.579
19	(1.640; 5.645)	0.621
20	(6.350; 3.045)	0.631
21	(0.000; 3.100)	0.645
22	(0.000; 0.000)	0,700
23	(1.640; 7.110)	0.965
24	(0.000; 5.520)	1.000
25	(0.080; 5.690)	1.000

Seven candidate waste disposal facilities are highlighted in the Table 1. As there are several candidate points, we need a second decision process where we will introduce new decision criteria that are mentioned in the regulations to find out the best place for a medical waste disposal facility. The five criteria mentioned in the second section are evaluated for each candidate site by experts. The fuzzy TOPSIS method of [17] was used to handle these evaluations.

For the evaluation of the first criterion, the distance between the facility and the districts, experts used the α values of seven candidate points. Bigger α values improve our model.

To figure out the transportation costs, the distances between districts and candidate points are measured and given in the Table 2. The waste amount generated by each district is in the Table 3. These amounts help us to determine the weights of each district that are used in the calculation of transportation costs.

Table 2: Distances of candidate sites to each demand point

SITE	SISLI	BESIKTAS	KAGITHANE	SARIYER
1	6.680	5.870	3.800	10.770
2	3.580	7.890	2.580	7.670
3	3.525	3.525	5.595	3.915
4	5.300	10.310	5.000	5.250
5	3.785	8.795	3.485	3.485
6	5.390	10.400	5.090	5.000
7	5.250	10.260	4.950	4.860

Table 3: Waste amounts of each district and weights (Istanbul Metropolitan Municipality).

DISTRICTS	MEDİCAL WASTE AMOUNT (KG/YEAR)	WEIGHT
SISLI	87 600	0.362
BESİKTAS	64 800	0.268
KAGITHANE	64 800	0.268
SARIYER	24 600	0.102

To evaluate the earthquake risk criteria, the experts used the earthquake map of Istanbul published by earthquake observatory of Istanbul. There are five different risk levels in the map. They evaluated each candidate point according to their location on the map.

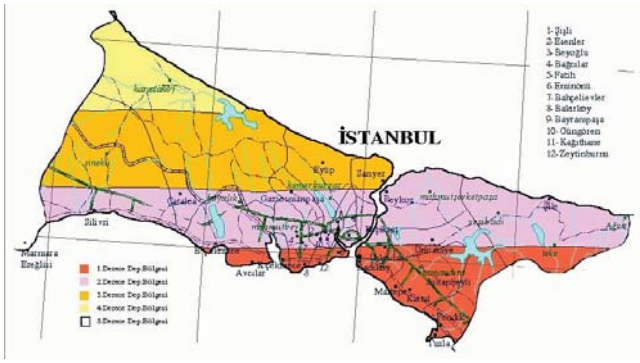


Figure 3: Earthquake Risk Map of Istanbul.

Distances of each candidate point to the water sources were measured using Google Earth. The results were evaluated by experts.

For the evaluation of air quality criterion, the experts used the data of Istanbul Technical University pollution reports.

ISBN: 978-989-95079-6-8

The figure 4 is the regional pollution map. The amount of SO_2 differs for each color in the map. It is denoted in the regulations that the waste disposal facility must be located in a region where it will not increase the pollution to high levels. So it is convenient to locate the facility in a less polluted area.

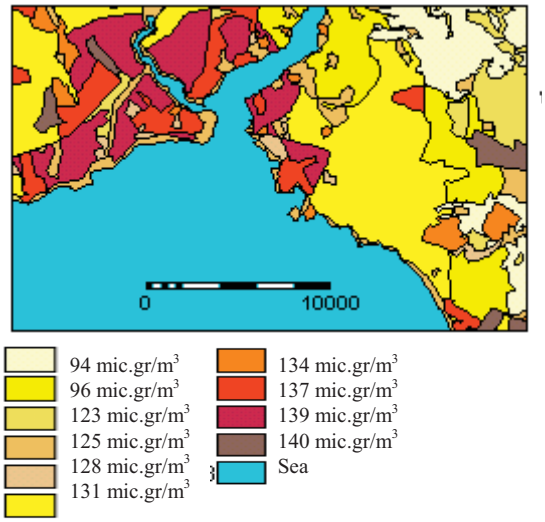


Figure 4: SO_2 amount for different places in Istanbul

The linguistic variables used by experts are in the Table 4. In our study we used a scale of [17]. The evaluation matrix and the normalized decision matrix of fuzzy TOPSIS decision process of this study are given in the Table 5 and Table 6.

Table 4: Linguistic Variables

Very poor (VP)	(0; 1; 3)
Poor (P)	(1; 3; 5)
Fair (F)	(3; 5; 7)
Good (G)	(5; 7; 9)
Very Good (VG)	(7; 9; 10)

Table 5: Expert Evaluation

	C1	C2	C3	C4	C5
A1	VP	G	F	VP	VG
A2	VP	VG	P	F	VP
A3	P	G	P	G	F
A4	F	P	VP	VG	P
A5	VG	VP	F	P	VG
A6	VG	VP	F	P	VG
A7	VG	VP	F	P	VG

Table 6: Normalized Decision Matrix

	C_1	C_2	C_3	C_4	C_5
	0.00	0.50	0.43	0.00	0.70
A_1	0.10	0.70	0.71	0.10	0.90
	0.30	0.90	1.00	0.30	1.00
	0.00	0.70	0.14	0.30	0.00
A_2	0.10	0.90	0.43	0.50	0.10
	0.30	1.00	0.71	0.70	0.30
	0.10	0.50	0.14	0.50	0.30
A_3	0.30	0.70	0.43	0.70	0.50
	0.50	0.90	0.71	0.90	0.70
	0.30	0.10	0.00	0.70	0.10
A_4	0.50	0.30	0.14	0.90	0.30
	0.70	0.50	0.43	1.00	0.50
	0.70	0.00	0.43	0.10	0.70
A_5	0.90	0.10	0.71	0.30	0.90
	1.00	0.30	1.00	0.50	1.00
	0.70	0.00	0.43	0.10	0.70
A_6	0.90	0.10	0.71	0.30	0.90
	1.00	0.30	1.00	0.50	1.00
	0.70	0.00	0.43	0.10	0.70
A_7	0.90	0.10	0.71	0.30	0.90
	1.00	0.30	1.00	0.50	1.00

The results are given in the Table 7. According to these results, three areas, A_5, A_6, and A_7 are equally adequate for the medical waste disposal facility. Selected point is sufficiently far from residential area and not too far to increase the transportation costs. So one can conclude that this study is meaningful for this kind of problems. But this model does not take into account other logistical constraints.

Table 7: Range of alternatives

	d_i^*	Range	d_i^-	Range	R_i	Range
A_1	1.64	3	1.83	3	0.53	3
A_2	2.13	5	1.35	5	0.39	5
A_3	1.57	2	1.93	2	0.55	2
A_4	2.04	4	1.45	4	0.42	4
A_5	1.30	1	2.17	1	0.62	1
A_6	1.30	1	2.17	1	0.62	1
A_7	1.30	1	2.17	1	0.62	1

5 Conclusions

This paper deals with the problem of determining the undesirable facility location site in Istanbul. We tried first to find the site of the facility which maximizes the minimal satisfaction degree among all demand points and maximizes the preference of the site by using a method based on the Ishii's.

There are four drawbacks in this kind of solution procedure. The first is that this solution procedure suggests that residential areas are discrete points. This problem can be handled by revising the intervals of membership function as in this study. Second is that more than one sub-problem can have the same optimum solution. To handle this problem we use the optimum of each sub-problem having nearly same α values, bigger than a fixed value, in a new decision making process. The inaptitude to introduce qualitative criteria to the decision process is another drawback. A fuzzy TOPSIS

ISBN: 978-989-95079-6-8

method is introduced to the decision making process as a second step to improve the model. The last drawback is not handled in this study but can be subject to a further study. When the number of districts increases, the number of sub problems increases hugely and it complicates the decision making process. For a limited number of districts, the apparent advantage of our solution procedure is that all problems are converted to linear problems and we solve them easily using LINDO software.

Acknowledgment

This research has been financially supported by Galatasaray University Research Fund.

References

[1] E. Melachrinoudis, Z. Xanthopulos, A Semi-obnoxious single location in Euclidean space, *Computers&Operations Research*, 30, 2191-2209, 2002.

[2] H. Ishii, Y.L. Lee, K.Y. Yeh, Fuzzy facility location problem with preference of candidate sites, *Fuzzy Sets and Systems*, 158, 1922-1930, 2007.

[3] C.F. Chiang, F.C. Sung, F.H. Chang, C.T. Tsai, Hospital waste generation during the SARS outbreak in Taiwan, *Infection Control and Hospital Epidemiology*, 27, 519-522,2006.

[4] Y.W. Cheng, F.C. Sung, Y. Yang, Y.H. Lo, Y.T. Chung, K.-C. Li, Medical waste production at hospitals and associated factors, *Waste Management*, 29, 440-444, 2009.

[5] M. Askarian, M. Vakili, G. Kabir, Results of a hospital waste survey in private hospitals in Fars Province, Iran, *Waste Management*, 24, 347-352, 2004.

[6] S. Fisher, Healthcare waste management in the UK: the challenges facing healthcare waste producers in light of changes in legislation and increased pressures to manage waste more efficiently, *Waste Management*, 25, 572-574, 2005.

[7] R. Mohee, Medical wastes characterization in healthcare institutions in Mauritius, *Waste Management*, 25, 575-581, 2005.

[8] A. Bdour, B. Altrabsheh, N. Hadadin, M. Al-Shareif, Assessment of medical wastes management practice: a case study of Northern Part of Jordan, *Waste Management*, 27, 746-759, 2007.

[9] J-C. Pomero, S. Barba-Romero, *Choix multicritère dans l'entreprise, principe et pratique*, Editions Hermes, Paris, (1993).

[10] G.R. Jhanshahloo, F. H. Lotfi, M. Izadikhah, Extension of the TOPSIS metod for decision-making problems with fuzzy data, *Applied Mathematics and Computation*, 181(2), 1544-1551, 2006.

[11] H.-J. Zimmermann, Fuzzy set theory and mathematical programming in: A. Jones et al., *Fuzzy Set Theory and Applications*, 99-114, Reidel, Dordrecht, 1986.

[12] S-H. Tsaur, T-Y. Chang and C-H. Yen, The evaluation of airline service quality by fuzzy MCDM, *Tourism Management*, 23(2), 107-115, 2002.

[13] M-F. Chen and G-H. Tzeng, Combining grey relation and TOPSIS concepts for selecting an expatriate host country, *Mathematical and Computer Modelling*, 40(13), 1473-1490, 2004.

[14] Y-M. Wang and T. M.S. Elhag, Fuzzy TOPSIS method based on alpha level sets with an application to bridge risk assessment, *Expert Systems with Applications*, 31(2), 309-319, 2006.

[15] C. Kahraman, G. Büyüközkan and N. Y. Ates, A two phase multi-attribute decision-making approach for new product introduction, *Information Sciences*, 177(7), 1567-1582, 2007.

[16] C. Kahraman, S. Çevik, N. Y. Ates and M. Gülbay, Fuzzy multi-criteria evaluation of industrial robotic systems, *Computers & Industrial Engineering*, 52(4), 414-433, 2007.

[17] T-C. Wang, T-H. Chang, Application of TOPSIS in evaluating initial training aircraft under a fuzzy environment, *Expert Systems with Applications*, 33, 870-880, 2007.

ISBN: 978-989-95079-6-8

Online Classification of Machine Operation Modes Based on Information Compression and Fuzzy Similarity Analysis

Gancho Vachkov

Faculty of Engineering, Kagawa University
Takamatsu City, Kagawa 761-0396, Japan
Email: vachkov@eng.kagawa-u.ac.jp

Abstract—*This paper proposes a computational scheme for online classification of process data sets that represent different operation modes of machines and other systems by use of fuzzy similarity analysis. The classification procedure starts with a preliminary given small number of known operation modes (data sets) which constitute the initial size of the Knowledge Base (KB). During the online classification, the dissimilarity (difference) degree between the newly submitted operation mode and each of the current modes stored in the KB is computed. As a result, depending on the preliminary given threshold for classification, the newly submitted mode is classified as belonging to a certain class (operation mode) from the KB or is considered as a new (unknown and different) mode. In the latter case, this mode is added as a new entry of the KB and used for the further online classification.*

An unsupervised learning algorithm (a modification of the Neural-Gas leaning algorithm) is used in the paper for compression of the original "raw" data from each operation mode into Compressed Information Model (CIM) with a smaller number of neurons. Then the similarity analysis is performed as a two-input fuzzy inference procedure that uses the Center-of-Gravity Distance and the Weighted Average Size Difference between two CIMs. The membership functions and the singletons of the fuzzy inference procedure are tuned in the paper by using a given set of test operation modes. The whole computational scheme is illustrated and discussed on a real example for classification of 5 main operation modes of a hydraulic excavator.

Keywords— Fuzzy Similarity, Information Compression, Online Classification, Operation Modes, Unsupervised Learning

1 Introduction

Classification of different available process data taken from the real operation of machines and other systems is a problem of utmost importance for the proper monitoring and performance evaluation of the machines. Generally speaking, the classification and pattern recognition problems, as well as various methods for their solution are very well presented and discussed in the literature [1-6]. However in the most often cases the problem is viewed as off-line classification of preliminary given set of data (patterns) with fixed size. Then the task is to classify every single data (pattern) from the given data set as belonging to one or another class.

When dealing with industrial systems and machines, the problem is sometimes different and more complicated, since one single data (i.e one measurement from a group of sensors at a given instant) is not sufficient to reveal clearly the current machine status (known as machine *operation mode*). We need a data set containing multiple measurements during a limited time of machine operation, in order to classify the current operation mode as *belonging* (or *not belonging*) to a certain operation mode.

During long time operation of the machine, *large* (or even *endless*) sequence of limited time operation data sets is obtained. Therefore a flexible *online* and *incremental classification scheme* is needed that should be able to *1)* properly classify the data sets into respective operation modes and *2)* to discover *new* operation modes that are stored in the KB in order to be used in the further online classification.

In this paper we present a special *two-stage* computation scheme for solving the problem of online classification, The *first stage* is information compression of the "raw operation data" representing a certain operation mode into a respective compressed information model (CIM), which consists of a small number of neurons. As a tool for information compression, a modified version of the off-line Neural-Gas unsupervised learning algorithm [3,4] is used in the paper. In the *second stage* a special fuzzy inference procedure for similarity analysis is proposed that uses two parameters, extracted from the CIM, namely the C*enter-of-Gravity* and the *Weighted Average Size* of the CIM. The differences between the respective parameters for a given pair of operation modes are used as two input *features* in the fuzzy inference procedure for similarity analysis.

The proposed classification scheme has the ability to perform *incremental classification*, which is very useful in the typical scenario of an unlimited sequence of operation data sets that represent different (n*ew*, unknown) *operation modes*. The flexibility and applicability of this scheme is shown in the paper on a real example for classification of five main operation modes of a hydraulic excavator.

2 The proposed unsupervised classification scheme

The Block-Diagram of the proposed scheme is shown in Fig. 1. This scheme is a further development of our previous idea for similarity analysis, discussed in [7].

The procedure starts with a small number of preliminary known *core operation modes* (core CIMs), which build the initial size of the Knowledge Base. During classification, the *difference (dissimilarity) degree* between the newly submitted (unknown) operation mode (as CIM) and all the core CIMs in the KB is computed. As a result, depending on a preliminary given *threshold Th* for classification, the new operation mode could be classified as *belonging* to a certain class (mode) from the KB or is regarded as a *quite different* (new) mode thus creating a *new class* in the KB. In such way, the proposed general concept for classification is incremental one, allowing the Knowledge Base to gradually grow, when a new operation mode with a very high dissimilarity degree to any of the core modes in the KB is discovered.

ISBN: 978-989-95079-6-8

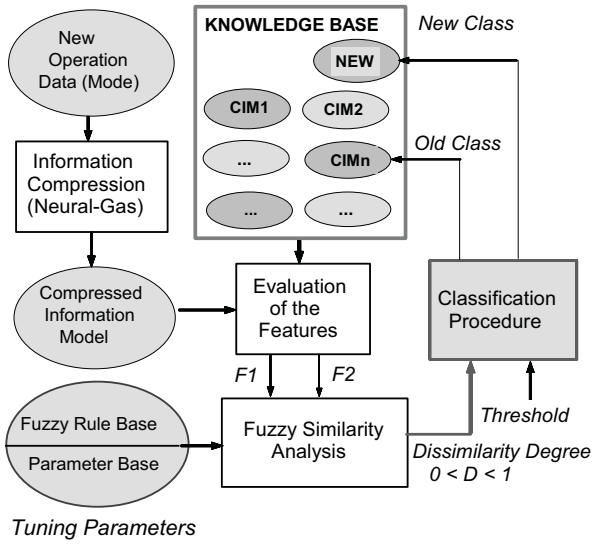


Figure 1: Block diagram of the proposed Online Classification Scheme based on fuzzy similarity analysis.

3 Unsupervised learning algorithm for information compression

The first step before the actual similarity analysis and classification of the operation modes is to find a way to reduce the large amount of the "raw data" information contained in the data set. Further on we refer to this computation step as *Information Compression*. From a computational viewpoint the information compression can be viewed as a kind of *transformation* of the original large data set: $\mathbf{x}_i = [x_{i1}, x_{i2}, ..., x_{iK}], i = 1,2,..., M$, consisting of *M* data in the *K*-dimensional input space into a respective *Neural Model* consisting of *N neurons* in the same space. Here $N << M$. and *CR = M/N* is the so called *Compression Ratio.*

The *information compression* of the original large data set (e.g. pixels or process data from sensors) can be perform by using different unsupervised competitive learning algorithms, such as clustering algorithms [1,2], Neural-Gas and its modifications [3,4,7], Support Vector Machines and Self-Organizing Maps [5,6] etc. All these algorithms try to find the most appropriate positions of the preliminary fixed number of *N* neurons (clusters) in the *K*-dimensional data space so that the resulting group of neurons resembles as much as possible the density distribution of the original data in the same space.

The essential part of any unsupervised learning algorithm is the so called *updating rule* for the neuron *centers* $\mathbf{c}_i, i = 1,2,...,N$ in the *K*-dimensional space. The algorithm updates the neuron centers iteratively with preliminary fixed number of iterations *T* (*t = 0,1,2,...,T*), as follows:

$$\mathbf{c}_i(t) = \mathbf{c}_i(t-1) + \Delta\mathbf{c}_i(t),\ i = 1,2,...,N. \qquad (1)$$

Here the computation of the update $\Delta\mathbf{c}_i(t)$ varies depending on the type of the unsupervised algorithm.

In this paper we use a modified version of the original *Neural-Gas* unsupervised learning algorithm, first presented in [3]. At every iteration the update is computed as:

$$\Delta\mathbf{c}_i(t) = R(t) H_s(t, r_i)\left[\mathbf{x}_s - \mathbf{c}_i(t-1)\right], \quad i = 1,2,...,N;\ s = 1,2,...,M \qquad (2)$$

Here $R(t),\ 0 \le R(t) \le 1,\ t = 0,1,2,...,T$ is a monotonically decreasing *Learning Rate,* which guarantees the convergence and stability of the learning process:

$$R(t) = R_0 \exp\left(-t/T_C\right),\ t = 0,1,...,T \qquad (3)$$

The so called *Neighborhood Function* in (2) $0 \le H_s(t, r_i) \le 1$ also decreases exponentially with the iterations. It computes the dynamically changing (decreasing) *activity area* for each neuron during the iterations, as follows:

$$H_s(t, r_i) = \exp\left[-(r_i - 1)/B(t)\right], \quad t = 0,1,...,T;\ s = 1,2,...,M;\ i = 1,2,...,N \qquad (4)$$

$$\text{where} \quad B(t) = \exp\left(-t/T_W\right),\ t = 0,1,...,T \qquad (5)$$

Here $r_i \in [1,2,...,N]$ is an integer number representing the so called *ranking position* of the *i*-th neuron (*i = 1,2,...,N*) to the *s*-th data point (*s = 1,2,...M*). This position is determined by the distance between the *i*-th neuron and the *s*-th data point. The closest neuron (in a sense of a minimal *Euclidean* distance) is called "winning neuron" and gets ranking position *r = 1*. The second closest neuron gets *r = 2* and so on. Example for information compression of a given data set by using the above algorithm is presented in the next Section.

4 Example of machine operation data for online classification and similarity analysis

The hydraulic excavator (HE) is a typical example of a complex machine. It is equipped with a turbo-diesel engine, which powers a special hydraulic system for performing different kinds of movements and working operations. HE normally works in a dynamical sequence of several repetitive *operation modes.* The following *five operation mode*s are typical for a normal long time operation of the excavator, when it does not move on the ground, as follows:

Mode 1. Loading the bucket with the raw material (sand, soil, stones etc);

Mode 2. Transporting the load in the bucket to the nearby truck by moving the arm (arrow) with the full bucket;

Mode 3. Unloading the bucket material into the truck;

Mode 4. Returning the arm with the empty bucket to the initial position for the next loading;

Mode 5. Short repetitive movements of the arm (*up* and *down*) for *pressing* the raw material in the truck;

All these modes differ from one to another in terms of the required engine power and the amount of the load in the bucket. Then the real practical problem is to properly *recognize* all these five operation *modes* and to *discover* possibly *new* ones (if any), by analyzing the available *online sequence* of data sets, each of them corresponding to *one operation mode*.

Such recognition and classification of the modes is important for detecting a slow trend of *deterioratio*n in the performance of the HE that needs appropriate maintenance or repair.

From an engineering viewpoint, the following six parameters are considered as important for the proper mode recognition, namely: *P1*- Engine Speed [*rpm*]; *P2* – Engine Boost Pressure; *P3* - Engine Oil Pressure; *P4* - Fuel Consumption; *P5* – Left Hydraulic Pump Pressure and *P6* – Right Hydraulic

ISBN: 978-989-95079-6-8

Pump Pressure. During the experiments, online measurements (at every second) were performed and the following Fig. 2 depicts the normalized data collected separately for each of the five typical operation modes.

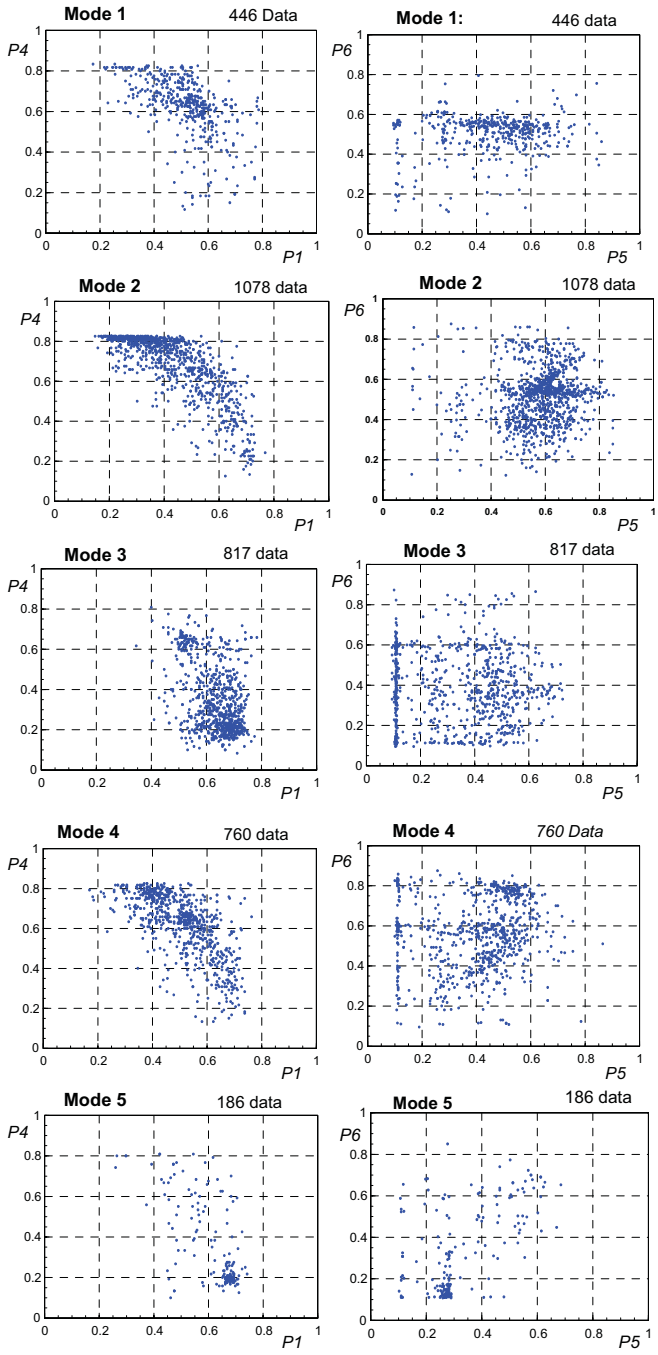


Figure 2: Plot of the normalized "raw data" representing all *5* operating modes of a hydraulic excavator.

As seen above, *two* different 2-dimensional plots are presented in Fig. 2, *P1-P4* and *P5-P6* respectively, for a better understanding of the complexity of the original *6*-dimensional problem.

Fig. 3. serves as illustration of the information compression of the "raw data" set, corresponding to operation *Mode 2*, by the Neural-Gas algorithm explained in Section 3. The following settings of the learning parameters for information compression have been used for all data sets from Fig. 2, as follows: $T = 500; R_0 = 0.16$ and $T_C = T_W = T/5$. The number of the neurons has been fixed to $N = 50$ in all simulations.

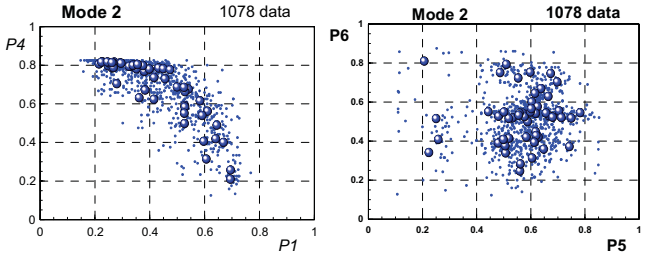


Figure 3: Results from the information compression of the "raw data" for operation *Mode 2* with $N = 50$ neurons.

5 Feature selection and computation for the fuzzy similarity analysis

As seen from Fig. 1 in Section 2, in order to evaluate the similarity between a given pair of operation modes, based on their compressed models (CIMs), we have to evaluate two important *features F1* and *F2* that characterize in an easy-to-understand numerical way the relation (similarity) between each pair of modes. For this purpose, we propose here to extract the following two distinct *parameters P1* and *P2* that characterize the *location* and the *size* of each operation mode in the *K*-dimensional input space. They are further called: *P1 - Center-of-Gravity* **CG** and *P2 - Weighted Average Size WAS* of the given operation mode.

1) The *Center-of-Gravity* $\mathbf{CG} = [CG_1, CG_2, ..., CG_K]$ of a *K*-dimensional operation mode is a *vector* that is computed directly from the respective CIM as follows:

$$CG_j = \sum_{i=1}^{N} c_{ij} g_i \Big/ \sum_{i=1}^{N} g_i, \; j = 1,2,...,K \qquad (6)$$

Here $c_{ij}, j = 1,2,...,K$ denotes the *center* (coordinates) of the *i*-th neuron in the *K*-dimensional input space and $0 < g_i \le 1, i = 1,2,...,N$ are the *normalized weights* of the neurons:

$$g_i = m_i / M; \; i = 1,2,...,N \qquad (7)$$

$m_i \le M, i = 1,2,...,N$ is the number of all data points: $\mathbf{x}_s, s = 1,2,...,m_i$, for which the *i*-th neuron is a *winning neuron* (i.e. the neuron with the shortest *Euclidean* distance to all of these data points). Obviously, the following equation holds: $\sum_{i=1}^{N} m_i = M$ and therefore $\sum_{i=1}^{N} g_i = 1$.

2) The *Weighted Average Size WAS* of the operation mode (and its respective CIM) is a *scalar* value, which takes into account the normalized weights of all neurons and the Euclidean distance ED_{pq} between all pairs of neurons, $\{p,q\}, p = 1,2,...,N; q = 1,2,...,N$, as shown in the next two equations (8) and (9):

$$WAS = \sum_{p=1}^{N-1} \sum_{q=p+1}^{N} ED_{pq} w_{pq} \Big/ \sum_{p=1}^{N-1} \sum_{q=p+1}^{N} w_{pq} \; . \qquad (8)$$

where $w_{pq} = g_p \times g_q, \quad p = 1,2,...,N; q = 1,2,...,N \qquad (9)$

Fig. 4 shows the locations of the centers-of-gravity **CG** for all *5* operation modes, computed by (6) and (7). It is seen from this figures that **CG** of some modes (such as: *Mode 1* and *Mode 4* and also *Mode 3* and *Mode 5*) are quite close to each

ISBN: 978-989-95079-6-8

other in the *K*-dimensional input space, which could result in a wrong classification.

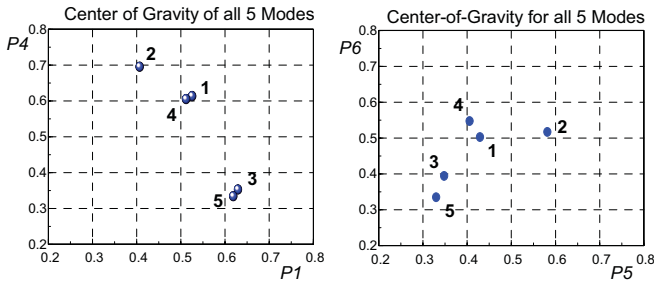


Figure 4:Center-of-Gravities **CG** for all *5* operation modes.

The following Table 1. shows the sizes *WAS* of all *5* Operation Modes, computed by (8) and (9). It also can be noticed here that some sizes are quite similar (such as *WAS* for *Mode 3* and *Mode 4*), which could also lead to wrong classification.

Table 1: Model Size *WAS* for all *5* operation modes.

Mode	Mode**1**	Mode**2**	Mode**3**	Mode**4**	Mode**5**
Model Size	*0.39538*	*0.38647*	*0.42592*	*0.42463*	*0.44486*

The above two parameters *P1* = **CG** and *P2* = *WAS* carry important information that can be used for selection of the two *features F1* and *F2* used as inputs of the Fuzzy Inference procedure for similarity analysis from Fig. 1.

We propose here an easy way to extract the features *F1* and *F2* as follows:

- The feature *F1* is a *scalar* value, computed as the *distance CGD* between the centers-of-gravities **CG** of a given pair $\{A,B\}$ of operation modes:

$$F1 = CGD_{AB} = \sqrt{\sum_{j=1}^{K} [CG_j^A - CG_j^B]^2} \qquad (10)$$

- Similarly the feature *F2* is a *scalar* value computed as the *difference WSD* between the weighted average sizes *WAS* of the same pair $\{A,B\}$ of operation modes, namely:

$$F2 = WSD_{AB} = |WAS_A - WAS_B| \qquad (11)$$

6 Structure of the block for fuzzy rule based similarity analysis

According to the block diagram shown in Fig. 1., we use the features *F1* and *F2*, computed by (10) and (11) as the inputs of the *Fuzzy Rule Based Inference Procedure* for similarity analysis of a given pair $\{A,B\}$ of operation modes.

Thus the Fuzzy Rule Based Procedure becomes a two-input / one output fuzzy system, as follows: $D = \mathbf{F}(F1,F2)$. Here $0.0 \le D \le 1.0$ is the *Difference Degree* (or *Dissimilarity Degree*). A difference degree $D = 0$ means that the operation modes *A* and *B* are *identical* (equal) and difference degree $D = 1$ means that *A* and *B* are *completely different modes*.

As well known [2], the fuzzy decision procedure consists of the following three main computation steps, as follows:

1) *Fuzzyfication* (with triangular Membership Functions);
2) *Fuzzy Inference* (with Product Operation) and
3) *Defiuzzification* (Weighted Mean Average).

For the next simulations in the paper, we assume *five triangular membership functions* that characterize linguistically the two inputs (features), namely *F1* and *F2*. They are used in the *fuzzification step* and have the following linguistic meaning: *VS = Very Small*; *SM = Small*; *MD = Medium*; *BG = Big* and *VB = Very Big*.

The Fuzzy Rule Base for the Fuzzy Inference procedure is shown in Fig. 5. It consists of *25* fuzzy rules, each of them with individual output as one of the *9 crisp* numerical values (*Singletons*): $U_1,U_2,...,U_9$, as seen in the figure. Note that the following inequality is required for achieving meaningful (plausible) results from the fuzzy similarity analysis:

$$U1 < U2 < ... < U9 \qquad (12)$$

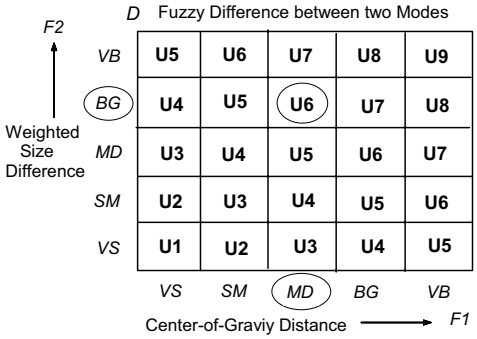


Figure 5: The Fuzzy Rule Base used for fuzzy similarity analysis.

The structure of the fuzzy rule base in Fig. 5 has been generated by using general *human logic and experience* from comparison and evaluation of data sets and other objects.

The well common weighted average method is used for the *Defuzzification* step, as follows:

$$D = \sum_{i=1}^{L} u_i v_i \Big/ \sum_{i=1}^{L} v_i \qquad (13)$$

Here $0 \le v_i \le 1, i=1,2,...,L$ is the *Firing* (*Activation*) *Degree* of the *i*-th fuzzy rule and $L = 25$ is the total number of the fuzzy rules. All rules have their individual crisp values (*Singletons*): $u_i \in [U_1,U_2,...,U_9], i=1,2,...,L$, according to the notations of the Fuzzy Rule Base in Fig. 5. For example, the *crisp* output of the Fuzzy Rule *No. 14* ($i = 14$), marked by circles in Fig. 5 is, as follows:

IF(*P1 is* MD AND *P2 is* BG) THEN $u_{14} = U6$

7 Optimization of the procedure for fuzzy similarity analysis

Once the structure of the Fuzzy Rule base is fixed, according to Fig. 5., then all the remaining parameters in the fuzzy inference procedure should be appropriately tuned. These are the parameters (*locations*) of the triangle membership functions and the *singletons* (consequents of the fuzzy rules). If successful optimization of these parameters is performed (according to a given optimization criterion), then we can expect correct (plausible) classification results.

There are two problems in connection with the optimization, namely: 1) construction of the *optimization criterion* with possible constraints and 2) selection of the *optimization method* (algorithm) to be used. As for the best selection of the

ISBN: 978-989-95079-6-8

optimization method, this is not a topic of interest in this paper, so we concentrate further on in the paper on the first problem, namely the selection of the optimization criterion.

We have collected *three groups* of data sets for all *5* operation modes of the hydraulic excavator, as well as *two* additional data sets representing two *different* (unknown) modes. One data set actually represents an *idling mode* of HE, while the other mode is not quite typical for the excavator (*moving* on the ground).

All the collected data sets have been first compressed into respective compressed information models (CIMs), according to the concept of the block diagram in Fig. 1. Then the respective CIMs have been split into the following *three* groups of models:

1) Five *training* models for each of the preliminary known operation modes, named as Mode**1**, Mode**2**,…,Mode**5**.

2) Five *test* models for the same group of five operation modes, named as: Test**1**, Test**2**, …, Test**5**;

3) A sequence of *16* operation models, used for *validation* of the classification procedure. These validation models are named as: Val**1**, Val**2**, Val**3**, …, Val**16** (or shortly: **1,2,…,16**). Then the sequence **1,2,…,16** is used to simulate the *Online Classification* process.

Table 2 shows the results from the similarity analysis between the *test modes* and the *training modes* for the initially assumed parameters (before optimization). The columns named as *Rank No.1* and *Rank No.2* refer to the *first* and *second* choice for classification (i.e the *least* difference and the *second least* difference). It is easy to notice the *contradiction* in this table arising at operation *Mode 5*. It is wrongly classified as *Mode 3* in this table.

Table 2: Similarity results before optimization.

	Training Modes: Mode **1** – Mode **5**			
Test	Rank **No.1**		Rank **No.2**	
Modes:	*No.*	*D*	*No.*	*D*
Test **1**	**1**	0.092	2	0.146
Test **2**	**2**	0.050	1	0.156
Test **3**	**3**	0.012	5	0.106
Test **4**	**4**	0.069	2	0.174
Test **5**	**3**	0.138	5	0.144

In order to correct this contradiction, we have to tune the parameters of the fuzzy inference procedure. It is done in this paper by assuming a special type of *optimization criterion,* which counts the *discrepancy DIS* between the human decision (or *human preference*) and the result from the computer-based classification. Table 3 shows example of one (plausible) human preference for classification of all *5* operation modes.

Table 3: Construction of the Optimization Criterion: *Preferred Human Decision*

Rank No. 1			**Rank No. 2**		
Test Mode	Train Mode	*D*	Test Mode	Train Mode	*D*
1	**1**	*0.10*	**1**	**4**	*0.20*
2	**2**	*0.10*	**2**	**1**	*0.20*
3	**3**	*0.10*	**3**	**5**	*0.20*
4	**4**	*0.10*	**4**	**1**	*0.20*
5	**5**	*0.10*	**5**	**3**	*0.20*

Now, by summing up the discrepancies between all *10* human preferences listed in Table 3 and the respective computer results from Table 2, we can compute the initial value of the optimization criterion as: *DIS = 0.5052.*

Further on we use a relatively simple (random search) optimization algorithm (further details omitted here) for a *two-stage* tuning of the fuzzy inference parameters:

- *Stage 1* optimizes the three intermediate locations (*SM, MD, BG*) of the membership functions for both inputs *F1* and *F2*, which means *3 + 3 = 6* optimization parameters;
- *Stage 2* optimizes the singleton values: $U_1,U_2,...,U_9$, taking into account the constraints (12).

The two stages were performed once in a consequence: *Stage1* → *Stage 2* with *30000* iteration steps for each stage. As a result the optimization criterion *DIS* was decrease to *0.3011* (after *Stage1*) and to *0.2983* (after *Stage2*). The final classification results are shown in Table 4. It is seen that all the classification results have become *correct* and with much closer degrees to the human preferences from Table 3.

Table 4: Similarity results after optimization.

	Training Modes Mode **1** – Mode **5**			
Test	Rank **No. 1**		Rank **No. 2**	
Modes:	*No.*	*D*	*No.*	*D*
Test **1**	**1**	0.123	2	0.169
Test **2**	**2**	0.096	1	0.192
Test **3**	**3**	0.016	5	0.168
Test **4**	**4**	0.114	2	0.208
Test **5**	**5**	0.141	3	0.189

The next three figures: Fig. 6., Fig. 7. and Fig. 8. are self explanatory. They show how the optimization has changed the parameters of the membership functions, the singletons and the Fuzzy Rule base response surface.

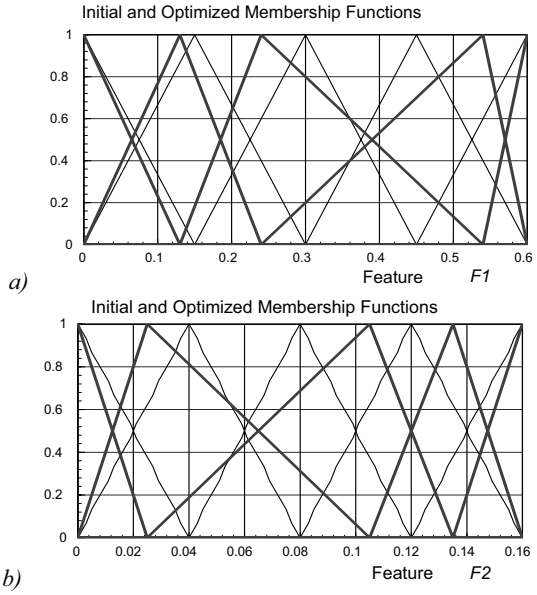


Figure 6: Results from the tuning of the membership functions. Bold lines denote the optimal positions.

8 Online classification results

After the fuzzy inference procedure has been tuned appropriately, then the whole *online classification* procedure has been tested on the sequence **1,2,…,16** of *16 validation* models Val**1**,Val**2**,…, Val**16**, according to the Block Diagram in Fig. 1. The results from the classification are shown in Fig. 9 and Fig. 10.

ISBN: 978-989-95079-6-8

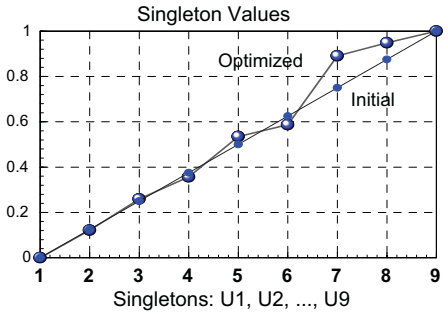


Figure 7: Results from the tuning of the singleton values.

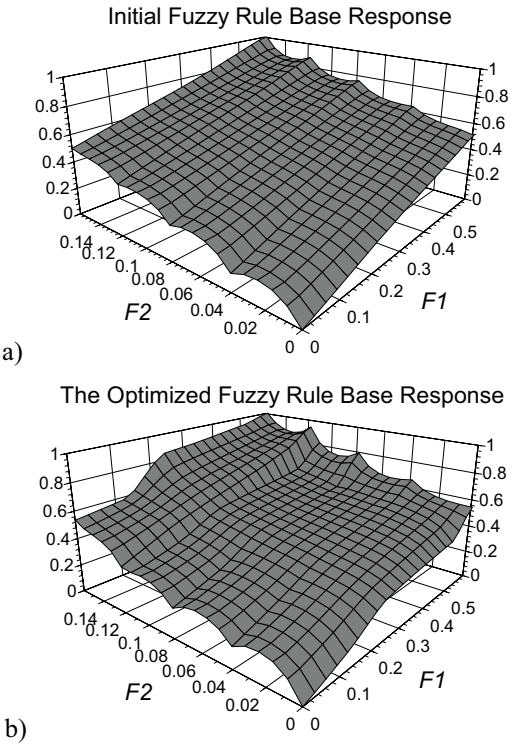


Figure 8: Response Surface of the fuzzy rule base: *a)* before optimization; *b)* after optimization.

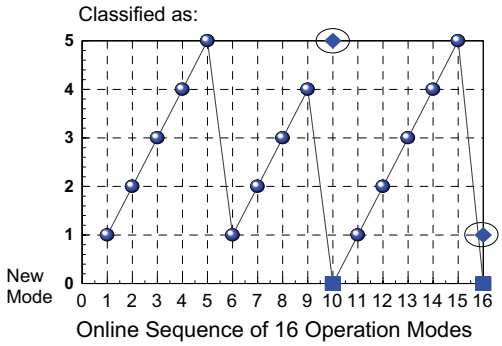


Figure 9: Classification Results for all *16* validation modes 1,2,…,16, presented as online sequence.

The "ball" symbols in Fig. 9. correspond to the correct classified modes. The two "diamond" symbols for operations *10* and *16* represent the computer classification results, while the respective two "square" symbols represent the actual "new" (unknown) operation modes.

It is seen from the similarity results in Fig. 9 that the operation sequences *10* and *16* were classified as *Mode 5* and *Mode 1* respectively. However, this decision was later *rejected,* as shown in Fig. 10, because their dissimilarity degrees exceed the predetermined threshold: *Th = 0.25*.

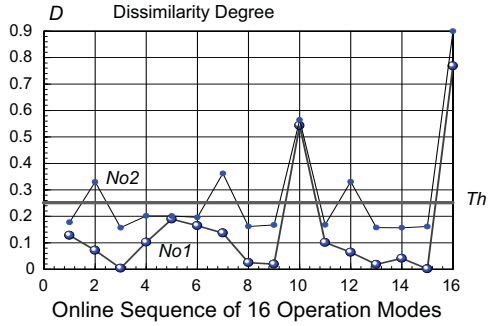


Figure 10: Confirmation results from the online classification with the predetermined threshold *Th = 0.25*.

Therefore finally the two operations *10* and *16* are assumed to correspond to *new* (*unknown*) operation modes. This result reflects the real situation.

9 Conclusions

The *Onine* classification of operation modes in continuously working machines and systems proposed in this paper uses fuzzy inference for similarity analysis. It is also suitable for possible applications in other areas of interest, such as search engines, classification and similarity analysis of images etc.

Essential feature of the whole proposed computation scheme is its *incremental* ability, in a sense that the newly detected modes could be added as new members of the current Knowledge Base and further on used for online classification.

The main originality of the proposed classification scheme is that it is a *human-assisted* classification, in which human experience and preference is taken into account and included into the optimization criterion. This criterion is further on used for tuning the parameters of the membership functions and the singletons in the fuzzy inference procedure so that to achieve the best possible matching between the computer results and human preferences.

There are several directions to improve the current status of this research. One is to investigate other optimization criterions and to select good, effective optimization method. Another direction is to solve the problem of *ever-growing* Knowledge Base during long time operation, by appropriate *pruning* and *merging* the modes in KB. In such way the whole classification system would become more flexible, *evolving classification* system.

References

[1] J.C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*, Plenum Press, New York, 1981.

[2] Ch.M. Bishop, *Neural networks for pattern recognition.* Oxford University Press, Oxford, 2003.

[3] T. Martinetz, S. Berkovich and K. Schulten, Neural-Gas network for vector quantization and its application to time-series prediction, *IEEE Trans. Neural Networks*, vol. 4, No. 4, 558 – 569, 1993.

[4] Ya-Jun Zhang and Zhi-Quang Liu, Self-Splitting competitive learning: A new On-Line clustering paradigm, *IEEE Trans. on Neural Networks,* vol. 13, No. 2, 369 – 380, 2002.

[5] Xueju Fu, C. Ong, S. Keerthi, G. Huang and L. Goh, Extracting the knowledge embedded in support vector machines, Proc. of the *IEEE Int. Joint Conf. on Neural Networks*, IJCNN 2004, vol. 1, 291-296, 2004.

[6] N. Kwak and Ch-Ho Choi, Input feature selection for classification problems, *IEEE Trans. on Neural Networks,* vol. 13, No. 2, 369 – 380, 2002.

[7] G. Vachkov, Classification of images based on information compression and fuzzy rule based similarity analysis, CD-ROM Proc. *World Congress on Computational Intelligence* WCCI 2008, Hong Kong, June, 2326 – 2332, June, 2008.

ISBN: 978-989-95079-6-8

Introducing Relation Compactness for Generating a Flexible Size of Search Results in Fuzzy Queries

Xiaohui Tang Guoqing Chen Qiang Wei

School of Economics and Management, Tsinghua University
Beijing 100084, China
Email: weiq@sem.tsinghua.edu.cn

Abstract-*Generating relevant and manageable search results is of great interest in database queries and information retrieval nowadays. This paper proposes an approach to measuring relation compactness and to generating a compact set of query outcomes under a fuzzy relational data model. The approach has desirable properties that (1) the resultant compact set is unique and contains mutually distinct tuples at the specified degree, and (2) the resultant compact set is information-equivalent to the collection of original query outcomes. Moreover, it is deemed appealing that the approach provides a way to obtain query outcomes of different size in a flexible manner according to the specified degree of compactness or the preferred number of tuples by users.*

Keywords—Tuple closeness, Fuzzy databases and queries, Relation compactness, Information equivalence, Web search

1 Introduction

The commonly used database model nowadays is the relational database (*RDB*) model initiated by Codd [1] which is generally designed under the assumption that the data/information is precise and queries are crisp. However, decision makers often face a growing need for representing and processing imprecise and uncertain information. Particularly in a web environment, approximate match is often preferred in generating search outcomes when exact match is unable to produce satisfactory results from targeted databases.

Since 1980's, a number of fuzzy relational database (*FRDB*) models and fuzzy query techniques [2-6] have been proposed to deal with imprecision and uncertainty of fuzziness type by means of fuzzy logic [7]. For instance, with the data stored in classical *RDB*, Kacprzyk and Zadrozny [8] proposed a system called *FQUERY* that could execute fuzzy queries such as "find (all) records such that most of the (important) clauses are satisfied (to a degree in [0, 1])" for crisp data stored in the Microsoft Access DBMS. In [9], a method to transform query conditions described by fuzzy numbers into the classical *SQL* queries was presented. In [10], the authors extended the *SQL* queries into *SQLf* to allow the expression of imprecise queries. Furthermore, a method was developed to transform the *FRDB* into weighted *RDB* in that a nesting mechanism was introduced to support the expression of some query results such as projection-selection-join operations [11]. In [12-13], we extended basic algebraic operators to deal with fuzzy data modeling and queries in forms of database design as well as equivalent definitions and transformation rules of the algebraic operators.

Notably, when querying databases with data of high volume, the size of the outcomes can easily be very large, even massive in a web database search context. For instance, a fuzzy query such as "*Select* Customer Name *From R Where* Age is about 25 AND Location is near B" requires an approximate match between query conditions and values of corresponding attributes. Another example is a web search, say via Google. Just click and see how many pages of outcomes to appear when keying in *"Population of Beijing"* for an exact match and *Population of Beijing* (notably without quotation marks) for an approximate match. Efforts have been made to deal with queries and web search with imprecise information and approximate match measures, so as to enrich the representation semantics and strengthen the power of search engines [5, 14-16]. Generally speaking, approximate match provides flexible queries, which is considered meaningful and useful in many cases, whereas the size of its outcomes is usually larger than that of the exact match.

Hence, the size of query outcomes becomes an issue of concern. One notable approach was to provide the users with the top-*k* results according to a ranking based on a certain evaluation function *F* between query conditions and all data records [17-18]. While sometimes the top-*k* results are sufficient for fuzzy/soft queries with approximate match, it is desirable and meaningful if a query/search could also guarantee that the selected results have no information loss as far as all the query results are concerned. The focus of this paper is then to provide the users with a compact set of query outcomes, in that the compact set (a) is smaller than and information-equivalent to the set of all query outcomes, and (b) at the same time has its size flexibly specified by users upon their need and preference.

2 Fuzzy Relational Database Model

Fuzzy logic incepted by Zadeh [7] attempts at quantifying and reasoning with imprecision and uncertainty that is common in the real world. Possibility distributions provide a graded semantics to natural language statements such as "the customer is young", which are often used in our daily communications. For example, the linguistic term "Young" can be defined as

$$\pi_{Young}(age) = \begin{cases} 1 & 0 \le age \le 30 \\ (50 - age)/20 & 30 < age \le 50 \\ 0 & age > 50 \end{cases} \quad (1)$$

ISBN: 978-989-95079-6-8

Two specific fuzzy relations are of particular interest, namely, closeness relation and similarity relation. A closeness relation c is a fuzzy relation on $X\times X$ such that $\forall x, y\in X$, $c(x, x) = 1$ (reflexive) and $c(x, y) = c(y, x)$ (symmetric). A similarity relation s is a fuzzy relation on $X\times X$ such that $\forall x, y, z\in X$, $s(x, x) =1$ (reflexive), $s(x, y) = s(y, x)$ (symmetric), and $s(x, z) \geq \sup_{y\in X} \min(s(x, y), s(y, z))$ (sup-min transitive). Apparently, similarity relation is a special case of closeness relation.

In this paper, the extended-possibility-based model is considered to be the underlying fuzzy database model since it facilitates handling both imprecision in attribute values (e.g., Age is young) and fuzziness in domain elements (e.g., between classic and gold for the domain of (customer) Class) in terms of possibility distributions (including fuzzy sets and linguistic terms) and closeness relations, respectively [4, 19]. In light of data representation, this model is deemed to be a general setting compared with two known models, namely, the so-called possibility-based model [4] where attribute values can be possibility distributions, and the so-called similarity-based model [3] where domain elements can be associated by similarity relations (reflective, symmetric and sup-min transitive). Specifically, in the extended-possibility-based model, a relation R is a subset of $\Pi(D_1) \times \Pi(D_2) \times \ldots \times \Pi(D_g)$, where $\Pi(D_i) =\{\pi_{Ai} \mid \pi_{Ai}$ is a possibility distribution of attribute A_i on domain $D_i\}$, and a closeness relation c_i is associated with domain D_i ($1 \leq i \leq g$). In addition, an g-tuple t of R is of the form: $t(\pi_{A1}, \pi_{A2}, \ldots, \pi_{Ag})$. An example tuple recording a customer's Name, Age and Class can be (Tony, {0.7/28, 1.0/33, 0.8/36}, diamond). The closeness relation c_{Class} on domain D_{Class}={classic, silver, gold, diamond} can be pre-defined by the managers as shown in Table 1.

Table 1: A closeness relation c_{Class}

c_{Class}	classic	silver	gold	diamond
classic	1.0	0.5	0.0	0.0
silver		1.0	0.75	0.25
gold			1.0	0.75
diamond				1.0

Given two g-tuples $t_p(\pi_{A1}, \pi_{A2}, \ldots, \pi_{Ag})$, $t_q(\pi'_{A1}, \pi'_{A2}, \ldots, \pi'_{Ag})$, where π_{Ai} and π'_{Ai} are normalized ($1 \leq i \leq g$), an example measure for the closeness of two values is [22]:

$$E_c(\pi_{A_i}, \pi'_{A_i}) = \sup_{\substack{x,y\in D_i \\ c_i(x,y)\geq\alpha_i}} \min(\pi_{A_i}(x), \pi'_{A_i}(y)) \qquad (2)$$

Here c_i is a closeness relation on domain D_i, $\alpha_i \in [0, 1]$ is a threshold specified by experts or the database managers. Other measures can be found in [19-21]. Then, the closeness of two tuples t_p and t_q can be considered as:

$$F_c(t_p, t_q) = \min(E_c(\pi_{A_1}, \pi'_{A_1}), E_c(\pi_{A_2}, \pi'_{A_2}), \ldots, E_c(\pi_{A_g}, \pi'_{A_g})) \qquad (3)$$

3 Tuple Extraction

Regarding tuple extraction in fuzzy relational databases, there are two issues to consider. One is to evaluate tuple closeness, the other is to extract representative tuples.

3.1 The evaluation of tuple closeness

For the fuzzy database, given an attribute A with domain D, let $\pi_1, \pi_2, \ldots, \pi_n$ be n possibility distributions as values of A, then a closeness relation on $\Pi(D)\times\Pi(D)$ is a mapping from $\Pi(D)\times\Pi(D)$ to [0,1], which can be represented by an $n\times n$ fuzzy matrix M where $e_{ij} = E_c(\pi_i, \pi_j)$, $1\leq i, j \leq n$, as follows:

$$M = \begin{pmatrix} e_{11} & e_{12} & \ldots & e_{1n} \\ e_{21} & e_{22} & \ldots & e_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ e_{l1} & e_{l2} & \ldots & e_{nn} \end{pmatrix} \qquad (4)$$

Further, for $0 \leq \lambda \leq 1$, where λ is a threshold which is usually determined by users or domain experts, its λ-cut (M_λ) is a classical relation defined by:

$$(e_{ij})_\lambda = \begin{cases} 1 & e_{ij} \geq \lambda \\ 0 & e_{ij} < \lambda \end{cases} \qquad (5)$$

Moreover, it is known that if M is a closeness relation, then its transitive closure M^+ (i.e., M^+ is a series of max-min compositions onto M itself with $M^+ = M^p = M^{p+1}$, $p\geq 1$) is a similarity relation, and M^+ converges within n-1 compositions [23-25]:

$$M^+ = M^p, \ p \leq l - 1 \qquad (6)$$

Furthermore, the equivalence classes of $(M^+)_\lambda$ constitute a partition of the domain concerned [26].

Definition 1: Given a relation $R = \{t_1, t_2, \ldots, t_n\} \subseteq \Pi(D_1) \times \Pi(D_2) \times\ldots\times\Pi(D_g)$, where tuple t_i is $t_i(\pi_{i1}, \pi_{i2}, \ldots, \pi_{ig})$, and threshold $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_g)$ with equivalence classes of $(M^+)_{\lambda i}$, $i = 1, 2, \ldots, n$, then two tuples $t_i(\pi_{i1}, \pi_{i2}, \ldots, \pi_{ig})$ and $t_j(\pi_{j1}, \pi_{j2}, \ldots, \pi_{jg})$ are called $\boldsymbol{\lambda}$-close if for all $k \in \{1, 2, \ldots, g\}$, π_{ik} and π_{jk} are in the same equivalence class. □

Example 1: Given threshold $\boldsymbol{\lambda} = (0.7, 0.8)$ and a relation $S \subseteq \Pi(D_1) \times \Pi(D_2)$ as shown in Table 2. For the sake of simplicity, assume that a_1, a_2, a_3, and a_4 are mutually distinct; so are b_1, b_2, b_3 and b_4. Then the $\boldsymbol{\lambda}$-close tuples can be generated as follows.

Table 2: A relation S with fuzziness

	A_1	A_2
t_1	$0.3/a_1, 0.7/a_2, 0.7/a_3, 0.7/a_4$	$0.53/b_1, 0.59/b_2, 0.62/b_3, 0.85/b_4$
t_2	$0.6/a_1, 0.4/a_2, 0.1/a_3, 0.8/a_4$	$0.92/b_1, 0.67/b_2, 0.09/b_3, 0.06/b_4$
t_3	$0.2/a_1, 0.6/a_2, 0.1/a_3, 1.0/a_4$	$0.45/b_1, 0.75/b_2, 0.06/b_3, 0.75/b_4$
t_4	$0.5/a_1, 0.1/a_2, 1.0/a_3, 0.9/a_4$	$0.61/b_1, 0.66/b_2, 0.35/b_3, 0.69/b_4$
t_5	$0.9/a_1, 0.6/a_2, 0.2/a_3, 0.2/a_4$	$0.56/b_1, 0.32/b_2, 0.98/b_3, 0.30/b_4$
t_6	$0.4/a_1, 0.2/a_2, 0.2/a_3, 0.4/a_4$	$0.81/b_1, 0.30/b_2, 0.93/b_3, 0.99/b_4$

First, for A_1 we can have:

$$M = \begin{pmatrix} 1 & 0.7 & 0.7 & 0.7 & 0.6 & 0.4 \\ 0.7 & 1 & 0.8 & 0.8 & 0.6 & 0.4 \\ 0.7 & 0.8 & 1 & 0.9 & 0.6 & 0.4 \\ 0.7 & 0.8 & 0.9 & 1 & 0.5 & 0.4 \\ 0.6 & 0.6 & 0.6 & 0.5 & 1 & 0.4 \\ 0.4 & 0.4 & 0.4 & 0.4 & 0.4 & 1 \end{pmatrix},$$

where each element on (i, j) in the matrix represents the closeness of t_i and t_j on A_1. Then, with $(M^+=M^3=M^2)_{0.7}$, three equivalence classes are obtained: $\{\pi_{11}, \pi_{21}, \pi_{31}, \pi_{41}\}$, $\{\pi_{51}\}$ and $\{\pi_{61}\}$. Consider attribute A_2, in the similar manner, there are three equivalence classes: $\{\pi_{12}, \pi_{22}, \pi_{52}, \pi_{62}\}$, $\{\pi_{32}\}$ and $\{\pi_{42}\}$.

ISBN: 978-989-95079-6-8

That means that values of t_1 and t_2 are in the same class for A_1 (i.e., π_{11}, π_{21}) as well as for A_2 (i.e., π_{12}, π_{22}), indicating that t_1 and t_2 are λ-close.

Finally, it is worthwhile to mention that, in case there are closeness relations on domains, the calculation of M can be readily made in consideration of c_i and α_i in E_c. For instance, suppose we have $c_1 = c_{\text{Class}}$ for A_1 with a_1 = classic, a_2 = silver, a_3 = gold and a_4 = diamond as shown in Table 1. Given α_1 = 0.7, since $c_1(a_2, a_3) = c_1(a_3, a_4) = 0.75$, we have $e'_{45} = 0.6 \neq e_{45} = 0.5$. In general, $M \subseteq M'$. □

3.2 Relation compactness

In addition to closeness of any two tuples, it is of interest to investigate the compactness of a given relation R composed of n tuples in the fuzzy relational database. Here the compactness of a relation refers to the degree of non-redundancy. For a given relation, the more redundant tuples it contains, the lower the degree of compactness. Straightforward ways of evaluating the compactness are to use the minimum closeness of any two tuples of R (*e.g.,* $1 - \min_{t,t' \in R} F_c(t, t')$) or the average closeness of any two tuples of R(*e.g.,* $1 - \sum_{t,t' \in R} F_c(t, t')/n(n-1)$) to reflect the compactness of R in whole. But it shall be noted that neither $1 - \min_{t,t' \in R} F_c(t, t')$ nor $1 - \sum_{t,t' \in R} Fc(t, t')/n(n-1)$ could describe the compactness of R in a satisfactory fashion. For example, suppose that all the n tuples in R are classical tuples for the sake of simplicity, i.e., for any $t, t' \in R$, $F_c(t, t') = 1$ or $F_c(t, t') = 0$. If there are $(n-1)$ tuples which are identical, only one tuple is not identical to them, then $1 - \min_{t,t' \in R} F_c(t, t') = 1$. But we know that there are a lot of tuples whose closeness degrees to each other are 1 in such a situation. On the other hand, if half of the tuples are identical to each other, and the other half of the tuples are also identical to each other (but different from the first half), then, $1 - \sum_{t, t' \in R} Fc(t, t')/n(n-1) = 2 \times (n/2) \times (n/2 - 1)/n(n-1) = 0.5 + 1/(2n - 2) > 0.5$. But in this situation, if n is large enough, e.g., $n = 100$, we will surely think that the compactness degree of R will be very close to 0, as 98% of the tuples can be deleted. Thus, a new measure, relation compactness, is proposed to evaluate a given relation R being compact. This measure not only works well in both of the above situations, but also possesses some good properties. Let us first consider it in classical relations, and then extend it into the relations in fuzzy databases.

In information theory [27], consider a single discrete information source, it may produce different kinds of symbol sets $A = \{a_1, a_2, \ldots, a_n\}$. For each possible symbol set A there will be a set of probabilities *pi* of producing the various possible symbols a_i ($\sum_{i=1}^{n} p_i = 1$), where these symbols are assumed successive and independent. Thus there is an entropy *Hi* for each a_i. The entropy of this given piece of information will be defined as the average of these H_i weighted in accordance with the probability of occurrence of the symbols in question,

$$H = H(p_1, p_2, \ldots, p_n) = -\sum_{i}^{n} p_i H_i = -\sum_{i}^{n} p_i \log p_i \qquad (7)$$

where the default log base is 2.

Similarly to [28], if there is a relation S with a set of classical tuples, which can be divided into m classes $C_1, C_2, \ldots, C_m$, the probability of a random tuple belonging to class C_i is n_i/n, where n_i is the number of tuples in C_i, and n is the number of tuples in S. Here, n_i/n is also called the probability of class C_i in S. The expected information for classifying this given relation S is:

$$H(n_1, n_2, \ldots, n_m) = -\sum_{i=1}^{m} p_i H_i = -\sum_{i=`}^{m} \frac{n_i}{n} \log \frac{n_i}{n} \qquad (8)$$

Definition 2: Let $R = \{t_1, t_2, \ldots, t_n\}$ be a classical relation with n tuples, R be divided into m classes $C_1, C_2, \ldots, C_m$ according to tuple identity (i.e., every tuple in C_i is identical to each other), and n_i be the number of tuples in C_i. Then the relation compactness of R is defined as

$$RC'(n_1, n_2, \ldots, n_m) = H(n_1, n_2, \ldots, n_m)/\log n$$
$$= -(\sum_{i=1}^{m} p_i H_i)/\log n = -(\sum_{i=1}^{m} \frac{n_i}{n} \log \frac{n_i}{n})/\log n \qquad (9)$$

For a relation R, its degree of compactness (RC') reflects the extent to which the tuples of R are not redundant, measured by the "amount of information" in R. In other words, the higher RC' is, the less redundant the tuples in R, meaning that the more information R contains.

Next, let us consider the fuzzy extension. Suppose there are n tuples with fuzziness involved in $R = \{t_1, t_2, \ldots, t_n\}$, where these tuples can be close to each other. To describe the degree at which a given relation with fuzziness is compact, the concept of relation compactness in the classical context could be extended to cope with the closeness of these tuples. A tuple t_k may not totally belong to class C_i, but belong to class C_i at a certain degree, e.g., $O_k^i \in [0, 1]$. Thus, we will use $\sum_{k=1}^{n_i} O_k^i$ instead of n_i in the extended relation compactness. Concretely, suppose we have n_i λ-close tuples in class $C_i = \{t_1, t_2, \ldots, t_{ni}\}$ of relation R. For each tuple t_k, $1 \le k \le n_i$, the degree of t_k belonging to C_i (or the closeness of t_k to C_i) can be defined as:

$$O_k^i = \sum_{j=1}^{n_i} F_c(t_k, t_j)/n_i \qquad (10)$$

Definition 3: Given a relation with fuzziness $R = \{t_1, t_2, \ldots, t_n\}$, where R can be divided into m classes $C_1, C_2, \ldots, C_m$ according to λ-close (i.e., every tuple in C_i is λ-close to each other), and n_i is the number of tuples in C_i. The relation compactness in R is,

$$RC(n'_1, n'_2, \ldots, n'_m) = -(\sum_{i=1}^{m} \frac{n'_i}{n} \log \frac{n'_i}{n})/\log n \qquad (11)$$

where $C_i = \{t_1, t_2, \ldots, t_{ni}\}$, $1 \le i \le m$, $1 < n_i \le n$, O_k^i is the degree of t_k belonging to C_i, and $n'_i = \sum_{k=1}^{n_i} O_k^i$ is the Σcount operation for the "effective number" of tuples in class C_i. □

As the classical relation is a special case of a relation with fuzziness, if all the tuples in $C_i = \{t_1, t_2, \ldots, t_{ni}\}$ are classical tuples, then they are equal to each other, i.e., $F_c(t_k, t_j) = 1$, $O_k^i = 1$, and $n'_i = n_i$ ($1 \le j, k \le n_i$). So $RC'(n_1, n_2, \ldots, n_m)$ is a special case of $RC(n'_1, n'_2, \ldots, n'_m)$. Note that one may also use some

ISBN: 978-989-95079-6-8

other measures (e.g., [19]) instead of O_k^i for the degree at which t_k belongs to C_i in (10). For *RC* in the fuzzy database context, the following proposition holds.

Proposition 1: Give a relation $R = \{t_1, t_2, ..., t_n\} \subseteq \Pi(D_1) \times \Pi(D_2) \times \ldots \times \Pi(D_g)$, where it is divided into m classes $C_1, C_2, ..., C_m$ according to $\boldsymbol{\lambda}$-close, $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, ..., \lambda_g\}$, and n_i is the number of tuples in C_i $(1 \le i \le m)$, then

(1) R is not $\boldsymbol{\lambda}$-close, i.e., $m = n$, $n'_1 = n'_2 = ... = n'_m = 1$, if and only if $RC(n'_1, n'_2, ..., n'_m) = 1$.

(2) Suppose $\min_{i=1}^{g} \lambda_i \ge e^{-1} \approx 0.37$, if R is totally $\boldsymbol{\lambda}$-close, i.e., $m = 1$, $n'_1 = n$, then $RC(n'_1)$ decreases when any two of the tuples's closeness $F_c(t_k, t_j)$ increases $(1 \le k, j \le n)$. Especially, if $\boldsymbol{\lambda} = \mathbf{1}$, then $RC(n'_1) = RC'(n_1) = 0$.

(3) $0 \le RC(n'_1, n'_2, ..., n'_m) \le 1$.

(4) Suppose $(n'_p + n'_q)/n \le e^{-1} \approx 0.37$, if n is fixed, $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, ..., \lambda_g\}$, $\lambda_i = \lambda$, $i = 1, 2, ..., g$, then *RC* decreases when any two of the original classes (e.g., C_p, C_q) are merged into one class C'_p according to $\boldsymbol{\lambda}$-close.

For different n and m, the value of *RC* is different. If $m = 2$, then $RC = -((n'_1/n)\log(n'_1/n) + ((n-n'_1)/n)\log((n-n'_1)/n))/\log n$, with different n, *RC* is shown in Figure 1. If $m = 3$, then $RC = -((n'_1/n)\log(n'_1/n) + (n'_2/n)\log(n'_2/n) + ((n-n'_1-n'_2)/n)\log((n-n'_1-n'_2)/n))/\log n$, with $n = 100$, *RC* is shown in Figure 2. Figures 1 and 2 reflect that RC has a characteristic of convexity.

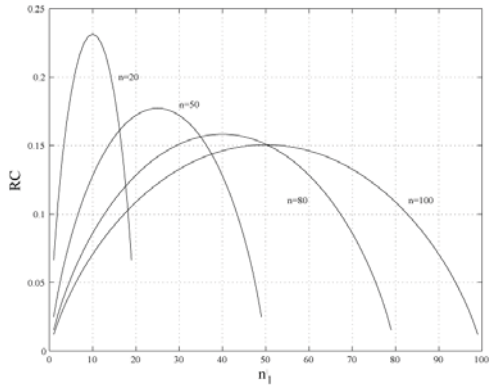


Figure 1: The value of *RC* ($m = 2$)

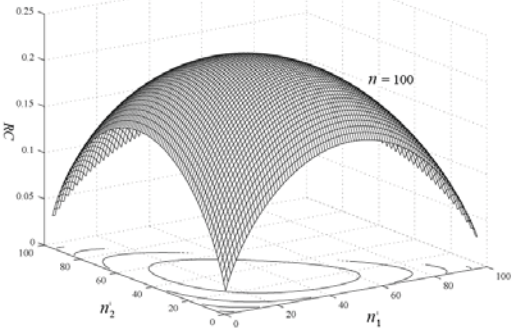


Figure 2: The value of *RC* ($m = 3$; $n = 100$)

3.3 Extraction of representative tuples

When it is known which tuples in a relation with fuzziness are $\boldsymbol{\lambda}$-close, then how to extract the representative tuples is the next problem of concern. It is considered ideal to have fewer tuples if they carry the same amount of original information. Usually, due to existence of tuple closeness, obtaining a smaller set of query outcomes becomes an effort of extracting representative tuples in light of "information equivalence".

Here we consider a center-based method for extracting the representative tuples. Conceptually, all the tuples in a same equivalence class are regarded to express approximately the same information and therefore one of them could be extracted to represent the class. Suppose we have n_i $\boldsymbol{\lambda}$-close tuples in class $C_i = \{t_1, t_2, ..., t_{ni}\}$ of a relation R. Then, we will keep the tuple whose closeness to C_i is the highest. This means, tuple t_p will be kept if $O_p^i = \max_{k=1}^{n_i} O_k^i$, $1 \le p \le n_i$.

Theorem 1: Suppose there are h $\boldsymbol{\lambda}$-close tuples in class $C = \{t_1, t_2, ..., t_h\}$ of a relation R, $C_i = C - \{t_i\}$, $1 \le i \le h$, and the relation compactness of C and C_i are *RC* and RC_i respectively. Then, tuple t_p $(1 \le p \le h)$ will be retained if and only if $RC_p = \max_{i=1}^{h} RC_i$.

Theorem 1 indicates that we can extract the representative tuples in light of relation compactness. It will keep the tuple that, without this tuple, the relation compactness of the other tuples will be the highest.

Example 2: Suppose we have had the $\boldsymbol{\lambda}$-close tuples of a fuzzy query about customers' age and salary as shown in Table 3.

Table 3: Customer's age and salary (in part)

	Age	Salary($)
t_1	0.6/25, 0.9/35, 0.6/40	0.8/4000, 0.9/5000, 0.8/6000
t_2	0.7/28, 0.8/35, 0.7/42	0.7/4500, 0.8/5000, 0.7/6000
t_3	0.6/25, 0.9/35, 0.8/42	0.9/4000, 0.7/5000, 0.6/6000

According to (2): $e_{12}(\pi_{11}, \pi_{21}) = 0.8$, $e_{13}(\pi_{11}, \pi_{31}) = 0.9$, $e_{23}(\pi_{21}, \pi_{31}) = 0.8$; $e_{12}(\pi_{12}, \pi_{22}) = 0.8$, $e_{13}(\pi_{12}, \pi_{32}) = 0.8$, $e_{23}(\pi_{22}, \pi_{32}) = 0.7$. Then, according to (3) and (10): $F_c(t_1, t_2) = 0.8$, $F_c(t_1, t_3) = 0.8$, $F_c(t_2, t_3) = 0.7$, $F_c(t_1, t_1) = F_c(t_2, t_2) = F_c(t_3, t_3) = 1$. $O_1 = 0.87$, $O_2 = 0.83$, $O_3 = 0.83$. Thus, t_2 and t_3 will be eliminated, t_1 will be kept.

Also, from the viewpoint of relation compactness, Table 3 can be considered as a class C, so, $C_1 = \{t_2, t_3\}$, $C_2 = \{t_1, t_3\}$, $C_3 = \{t_1, t_2\}$. For RC_1, $n'_1 = (F_c(t_2, t_2) + F_c(t_2, t_3))=2 + (F_c(t_3, t_3) + F_c(t_2, t_3))/2 = 1.7$. $RC_1 = -(1.7/2)\log(1.7/2)/\log 2 = 0.20$. In the same way, $RC_2 = RC_3 = 0.14$, thus, t_1 will be kept.

As far as the complexity of the tuple extraction is concerned, while introducing possibility distributions enriches the semantics that the data model represents, approximate match may lead to an increase in computational complexity. Further analysis shows that the complexity is polynomial and generally manageable. Suppose there are n tuples in relation R, the matrix M has $n \times n$ entries. For every entry, suppose there are, on average, p values in the two imprecise attribute values to compare, the cost is $O(p^2)$. The total cost to get M is then $O(p^2n^2)$. A reasonable estimate of p could be within 10. Therefore, the total cost to get M could be simplified to be $O(ln^2)$ where $l < 100$. For the transitive closure of M, the total cost is at $O(n^2)$ level for large n [24-25].

Furthermore, in consideration of generating matrix M, though detailed treatments go beyond the scope of this paper, there

ISBN: 978-989-95079-6-8

have been effective approaches in database queries, informational retrieval and web search, using a variety of measures, in respective fields, for keyword similarity, text match, semantic proximity, etc., in order to compare pair-wise closeness of data/information of interest (e.g., [30-34]).

4 Further Discussions

First, whether the extraction result is unique and representative tuples are not redundant is an interesting and important issue. Importantly, it can be proven that the result of the tuple extraction treatment in Section 3 is unique, and the resulting new relation is not $\boldsymbol{\lambda}$-close.

Second, we can also have the number of tuples in the outcome to be controlled by the setting of $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_g)$. When λ_i increases, there will be more classes, thus, after the extraction, more tuples will be kept. Especially, with $\lambda_i = 1$, the treatment turns into the classical situation, in that no tuples will be eliminated.

Third, two representations are information equivalent if the transformation from one to the other entails no loss of information, i.e., if each can be constructed from the other [35]. In other words, the information in one is also inferable from the other, and vice versa [36-37]. Here, "information equivalence" in our discussions could be described as follows. Suppose there is an original query result R_1. The classified relation is R_2 (i.e., the $\boldsymbol{\lambda}$-close tuples are in the same classes), and the resultant relation (compact set) containing representative tuples is R_3, which is not $\boldsymbol{\lambda}$-close.

Compared with original relation R_1 of the query result, R_2 and R_3 contain the classification information. We can say that, in terms of $\boldsymbol{\lambda}$-close, R_2 and R_3 are information equivalent. If we have had R_2, then R_3 could be derived (i.e., via extracting) with respect to tuple closeness/relation compactness. If we have had R_3, suppose $R_3 = \{t_u, t_w, t_x, ...\}$, then every tuple of R_3 is not $\boldsymbol{\lambda}$-close, and can be considered as a class labeled as $C_u, C_w, C_x, ...$. For every tuple t_i of R, if t_i satisfies the fuzzy query conditions, then it could be compared with the tuples in the labeled classes $C_u, C_w, C_x, ...$, if there is a tuple t in one labeled class, e.g., C_u such that t_i and t are $\boldsymbol{\lambda}$-close, then add t_i to C_u. Consequently, we will get relation R_2, as the tuples in R_3 and the equivalence classes in R_2 are bijective.

Finally, let us discuss the approach from a more applied perspective. While searching via a search engine (e.g., Google or the like) or querying available databases for some information, it would be desirable if the database/web search management systems are capable of generating outcomes that are representative and of a manageable size. Even better is to control the query/search outcomes with different compactness degrees or with different sizes. A way to view different results is to specify the degree of relation compactness. If R_C is higher, there will be fewer outcomes, otherwise, more outcomes.

Example 3: Someone uses a search engine to browse for the literature about "fuzzy queries". To help discuss the procedure and treatment details of how the search results can be shortened or extended, we hereby only chose 10 resultant items as R_1, merely for the sake of illustrative simplicity:

a_1: A Fuzzy Database Model for Supporting a Concept-Based Query ...
a_2: Fuzzy Database Query Languages and Their Relational Completeness ...
a_3: FSQL (Fuzzy SQL), a Fuzzy Query Language
a_4: Fuzzy Database Modeling with XML
a_5: Fuzzy database systems - Fuzzy Systems, 1995. International Joint ...
a_6: Amazon.com: Fuzzy Database Modeling of Imprecise and Uncertain ...
a_7: Amazon.com: Fuzzy Database Modeling with XML
a_8: Database Schema with Fuzzy Classification and Classification Query ...
a_9: Type-2 Fuzzy Logic - Publications Database
a_{10}: Fuzzy Clustering for Content-based Indexing in Multimedia Database

Suppose the threshold set by the user is $\lambda = 0.7$ and the closeness of these articles generated with a keywords match technique (e.g., [30-34]) is shown in Table 4.

Table 4: The closeness degrees among $a_1, a_2, \ldots, a_{10}$

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
a_1	1	0.75	0.9	0.6	0.5	0.5	0.5	0.6	0.6	0.6
a_2		1	0.82	0.6	0.6	0.6	0.6	0.5	0.6	0.6
a_3			1	0.6	0.5	0.5	0.6	0.5	0.6	0.6
a_4				1	0.85	0.9	0.9	0.6	0.6	0.6
a_5					1	0.7	0.75	0.6	0.5	0.5
a_6						1	0.78	0.6	0.6	0.6
a_7							1	0.6	0.5	0.6
a_8								1	0.9	0.95
a_9									1	0.7
a_{10}										1

Now, let us consider how to obtain the outcomes with $RC|_{0.7}$ being 1 and 0.6 respectively. In doing so, the database management systems or search processing functions could execute the extracting process with $\lambda = 0.7$, $\lambda_1 = 0.8$, $\lambda_2 = 0.85$, $\lambda 3 = 0.9$ and $\lambda 4 = 1$ at the back-end as follows.

For $\lambda = 0.7$, R_1 can be classified as $\{a_1, a_2, a_3\}$, $\{a_4, a_5, a_6, a_7\}$ and $\{a_8, a_9, a_{10}\}$, extracting the representative tuples as described in Section 3.3, the final outcome is $S = \{a_3, a_4, a_8\}$, $RC(S)|_{0.7} = 1$.

For $\lambda_1 = 0.8$, R_1 can be classified as $\{a_1, a_2, a_3\}$, $\{a_4, a_5, a_6, a_7\}$ and $\{a_8, a_9, a_{10}\}$, $S_1 = \{a_3, a_4, a_8\}$, $RC(S_1)|_{0.8} = RC(S_1)|_{0.7} = 1$.

For $\lambda_2 = 0.85$, R_1 can be classified as $\{a_1\}$, $\{a_2\}$, $\{a_3\}$, $\{a_4, a_5, a_6, a_7\}$ and $\{a_8, a_9, a_{10}\}$, $S_2 = \{a_1, a_2, a_3, a_4, a_8\}$, $RC(S_2)|_{0.85} = 1$, $RC(S_2)|_{0.7} = 0.6$.

In the same manner, for $\lambda_3 = 0.9$, $S_3 = \{a_1, a_2, a_3, a_4, a_5, a_8\}$, $RC(S_3)|_{0.9} = 1$, $RC(S_3)|_{0.7} = 0.57$.

For $\lambda_4 = 1$, $S_4 = R_1 = \{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}\}$, $RC(S_4)|_1 = 1$, $RC(R_1)|_{0.7} = RC(S_4)|_{0.7} = 0.47$.

Thus, if the threshold $\lambda = 0.7$, the original result of the query is compact at a degree of $RC(R_1)|_{0.7} = 0.47$. The back-end can give the user a totally compact outcome $S = \{a_3, a_4, a_8\}$ (with $RC(S)|_{0.7} = 1$) by default, which is composed of the three items:

"FSQL (Fuzzy SQL), a Fuzzy Query Language"
"Fuzzy Database Modeling with XML"

ISBN: 978-989-95079-6-8

"Database Schema with Fuzzy Classification and Classification Query ..."

If user wants to review more outcomes, e.g., $RC|_{0.7} = 0.6$, then the DBMS can generate $S_2 = \{a_1, a_2, a_3, a_4, a_8\}$, with $RC(S_2)|_{0.7} = 0.6$. Two more specific "fuzzy query" related items are listed: "A Fuzzy Database Model for Supporting a Concept-Based Query ..." and "Fuzzy Database Query Languages and Their Relational Completeness ...". In this way, the list of the literature could be shortened or extended depending upon the need and preference of the user. □

5 Conclusion

This paper has introduced a measure, namely relation compactness, to express the degree of a given relation being compact. Then an approach to evaluating and extracting representative tuples has been proposed. It has been proven that the resultant set of query outcomes is unique, smaller in size, and information equivalent to the original result. Moreover, the approach enables the query/search users to obtain various sizes of results upon their need and preference in light of compactness.

Ongoing research is centering on an application with real data on tuple extraction in both database query and web search contexts.

Acknowledgements

The work was partly supported by the National Natural Science Foundation of China (70890083/70621061) and Tsinghua University's Research Center for Contemporary Management.

References

[1] E.F. Codd, "A relational model for large shared data banks", *Communications of the ACM*, vol. 13, no. 6, pp. 377-387, 1970.

[2] J.F. Baldwin, S.Q. Zhou, "A fuzzy relational inference language", *Fuzzy Sets and Systems*, vol. 14, pp. 155-174, 1984.

[3] B.P. Buckles, F.E. Petry, "A fuzzy representation of data for relational databases", *Fuzzy Sets and Systems*, vol. 7, pp. 213-226, 1982.

[4] H. Prade, C. Testemale, "Generalizing database relational algebra for the treatment of incomplete or uncertain information and vague queries", *Information Sciences*, vol. 34, pp. 115-143, 1984.

[5] V. Owei, "An Intelligent Approach To Handling Imperfect Information In Concept-Based Natural Language Queries", *ACM Transactions on Information Systems*, vol. 20, no. 3, pp. 291-328, 2002.

[6] Z.M. Ma, W.J. Zhang, W.Y. Ma, "Extending object-oriented databases for fuzzy information modeling", *Information Systems*, vol. 29, no. 5, pp. 421-435, 2004.

[7] L.A. Zadeh, "Fuzzy sets as a basis for a theory of possibility", *Fuzzy Sets and Systems*, vol. 1, no. 1, pp. 3-28, 1978.

[8] J. Kacprzyk, S. Zadrozny, "Computing with words in intelligent database querying: standalone and internet-based applications", *Information Sciences*, vol. 134, pp. 71-109, 2001.

[9] S.M. Chen, W.T. Jong, "Fuzzy Query Translation For Relational Database Systems", *IEEE Transactions on System, Man, and Cybernetics-Part B*, vol. 27, no. 4, pp. 714-721, 1997.

[10] P. Bosc, O. Pivert, "SQLf: a relational database language for fuzzy querying", *IEEE Transactions on Fuzzy Systems*, vol. 3, no. 1, pp.1-17, 1995.

[11] P. Bosc, O. Pivert, "About project-selection-join queries addressed to possibilistic relational databases", *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 1, pp. 124-139, 2005.

[12] G.Q. Chen, E.E. Kerre, J. Vandenbulcke, "Normalization based on fuzzy functional dependency in a fuzzy relational data model", *Information Systems*, vol. 21, no. 3, pp. 299-310, 1996.

[13] X.H. Tang, G.Q. Chen, "Equivalence and Transformation of Extended Algebraic Operators in Fuzzy Relational Databases", *Fuzzy Sets and Systems*, vol. 157, no. 12, pp. 1581-1596, 2006.

[14] P. Buche, C. Dervin, O. Haemmerle, R. Thomopoulos, "Fuzzy querying of incomplete, imprecise, and heterogeneously structured data in the relational model using ontologies and rules", *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 3, pp. 373-383, 2005.

[15] D.Y. Choi, "Enhancing the power of Web search engines by means of fuzzy query", *Decision Support Systems*, vol. 35, pp. 31-44, 2003.

[16] S. Fox, K. Karnawat, M. Mydland, S. Dumais, T. White, "Evaluating Implicit Measures to Improve Web Search", *ACM Transactions on Information Systems*, vol. 23, no. 2, pp. 147-168, 2005.

[17] C. Li, K. C. Chang, I. F. Ilyas, S. Song, "RankSQL: Query Algebra and Optimization for Relational Top-k Queries", in *Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data*, 2005, pp. 131-142.

[18] H. Yu, S. Hwang, K. C. Chang, "Enabling soft queries for data retrieval", *Information Systems*, vol. 32, pp. 560-574, 2007.

[19] G.Q. Chen, J. Vandenbulcke, E.E. Kerre, "A general treatment of data redundancy in a fuzzy relational data model", *Journal of The American Society for Information Science*, vol. 43, no. 4, pp. 304-311, 1992.

[20] J.C. Cubero, M.A. Vila, "A new definition of fuzzy functional dependency in fuzzy relational databases", *Intelligent Systems*, vol. 9, no. 5, pp. 441-448, 1994.

[21] K.V.S.V.N. Raju, A.K. Majumdar, "Fuzzy functional dependencies and lossless join decomposition of fuzzy relational database systems", *ACM Transactions on Database Systems*, vol. 13, no. 2, pp. 129-166, 1988.

[22] G.Q. Chen, *Fuzzy logic in data modeling: semantics, constraints, and database design*, Kluwer Academic Publishers, Boston, 1998.

[23] S. Tamura, S. Higuchi, K. Tanaka, "Pattern classification based on fuzzy relations", *IEEE Transactions on Systems, Man and Cybernetics*, vol. 1, no. 1, pp. 61-66, 1971.

[24] H.S. Lee, "An optimal algorithm for computing the max-min transitive closure of a fuzzy similarity matrix", *Fuzzy Sets and Systems*, vol. 123, pp. 129-136, 2001.

[25] H.L. Larsen, R.R. Yager, "Efficient computation of transitive closures", *Fuzzy Sets and Systems*, vol. 38, no. 1, pp. 81-90, 1990.

[26] S. Shenoi, A. Melton, Fan, L.T. "An equivalence classes model of fuzzy relational databases", *Fuzzy Sets and Systems*, vol. 38, pp. 153-170, 1990.

[27] C.E. Shannon, "A Mathematical Theory of Communication", *The Bell System Technical Journal*, vol. 27, pp. 379-423, pp. 623-656, 1948.

[28] J. Han, M. Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann publishers, 2001.

[29] T.M. Cover, J.A. Thomas, *Elements of Information Theory*, John Wiley, 1991.

[30] Y.Z. Cao, M.S. Ying, G.Q. Chen, "Retraction and Generalized Extension of Computing With Words", *IEEE Transactions on Fuzzy Systems*, vol. 15, no. 6, pp. 1238-1250, 2007.

[31] S. Medasani, R. Krishnapuram, and Y. Choi, "Graph matching by relaxation of fuzzy assignments", *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 1, pp. 173-182, 2001.

[32] A. Koeller, V. Keelara, "Approximate matching of textual domain attributes for information source integration", in *Proc. of the 2nd international workshop on Information quality in information systems*, 2005, pp. 77-86.

[33] M. Mitra, B.B. Chaudhuri, "Information Retrieval from Documents: A Survey", *Information Retrieval*, vol. 2, pp. 141-163, 2000.

[34] E. Rahm, P.A. Bernstein, "A survey of approaches to automatic schema matching", *The International Journal on Very Large Data Bases*, vol. 10, no. 4, pp. 334-350, 2001.

[35] H.A. Simon, "On the forms of mental representation" in *Minnesota Studies in the Philosophy of Science: Perception and cognition: issues in the foundations of psychology*, C. W. Savage Eds. University of Minnesota Press, 1978.

[36] J.H. Larkin, H.A. Simon, "Why a diagram is (sometimes) worth ten thousand words", *Cognitive Science*, vol. 11, pp. 65-99, 1987.

[37] K. Siau, "Informational and Computational Equivalence in Comparing Information Modeling Methods", *Journal of Database Management*, vol. 15, no. 1, pp. 73-86, 2004.

Robot Cooperation without Explicit Communication by Fuzzy Signatures and Decision Trees

Áron Ballagi[1] László T. Kóczy[2,3] Tamás D. Gedeon[4]

1 Department of Automation,
Széchenyi István University,
H-9026, Győr, Egyetem tér 1., Hungary

2 Inst. of Mechanical, Electrical Engineering and Information Technology,
Széchenyi István University,
H-9026, Győr, Egyetem tér 1., Hungary

3 Department of Telecommunications and Media Informatics,
Budapest University of Technology and Economics,
H-1117, Budapest, Magyar tudósok krt. 2., Hungary

4Department of Computer Science,
Australian National University,
Acton, ACT (Canberra), Australia

Email: ballagi@sze.hu, koczy@tmit.bme.hu, tom.gedeon@anu.edu.au

Abstract—*This paper presents a novel action selection method for multi robot task sharing problem. Two autonomous mobile robots try to cooperate for push a box to a goal position. Both robots equipped with object and goal sensing, but do not have explicit communication ability. We explore the use of fuzzy signatures and decision making system to intention guessing and efficient action selection. Virtual reality simulation is used to build and test our proposed algorithm.*

Keywords— Mobile robots, cooperation, fuzzy signatures, action selection, intention guessing.

1 Introduction

In this paper we focus on manipulation tasks with sufficiently challenging dynamics to require the careful cooperation of two or more robots. Sensing and actuation is noisy and uncertain in mobile robot domains, resulting in partial knowledge about the world. We explore the use of fuzzy signatures to efficient action selection and intention guessing in this environment. The intention guessing is the base of a meta-communication method between the robots. In this setup is not any explicit communication line.

We chose box-pushing as the problem domain because it has both theoretical interest and practical applications as it is an instance in large class of practical object manipulation tasks that appear to require tight feedback and control of real-world physics and dynamics [1-9]. From a theoretical standpoint, box-pushing is a variant on canonical object manipulation problem that draws on issue in fine motion as well as high level planning and control. Box-pushing is related to the well-known "piano-movers problem" [10], in that it requires the achievement of top level goal of delivering the box to a particular location, as well as the maintenance of low-level requirements including obstacle avoidance, maintaining contact with the box, and maintaining forward motion. From a practical point of view, box-pushing is a prototypical problem for studying various tasks requiring cooperation of number of smaller robots moving larger objects [6].

We use simulation in our experiments where two robots push a box to a goal position. Each robot has the own behaviour based control system thus they are fully autonomous. An action selection mechanism works in our behaviour based control in which the decision about selection is done by fuzzy signature based state describing algorithm.

2 Experimental task and environment

The actual stage of our research we use simulation of our real differential driven autonomous micro-robots (Fig. 1). The physical simulation is exact model of our robots in the case of scale, weight, mechanical systems and sensors.

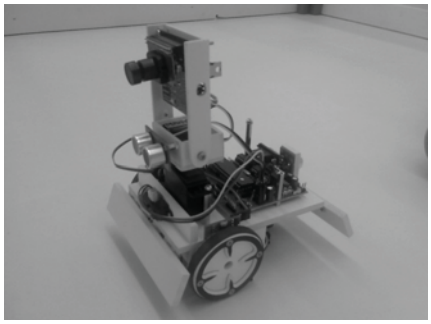


Figure 1. The real box-pushing robot

The two robots operate in a 2 x 2 m square arena. The box to be pushed is a 20 cm high and 40 cm wide and long. The

ISBN: 978-989-95079-6-8

goal region, located in one corner of the arena (Fig. 2), indicated by light sources detectable by the robot's sensor. In the experiments described here the goal location is fixed but it can be moved before and during experiments.

Each robot pushes the box with two "whiskers" which have a pair of force sensors. The left and right whiskers each provides an analog force signal which is combined to give information the relative position to the box and via the control loop keep it contact on both sides and thus perpendicular to the box.

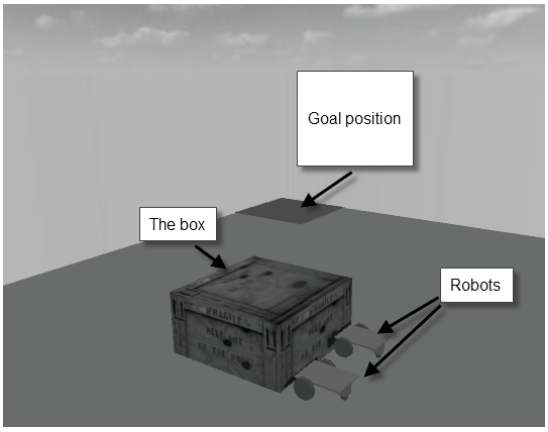


Figure 2. The simulation environment

In the description two direction sign systems are used, the absolute direction with letters N, E, S, W as in the usual sense for North, East, South and West. The second direction sign system is a box relative system where the sides of the box are N_B, E_B, S_B and W_B respectively (Fig. 3). The position of the objects (boxes and robots in this case) always can be described by the absolute course, latitude and longitude of the object. One object relative position to a box is described by the box relative system, i.e. which side of the box is touched by that object. For simplicity we assume that the sides of the box are always parallel with the N-S and E-W axes, so there is not necessary any rotation.

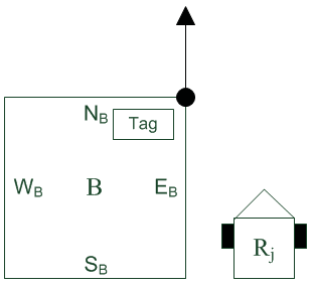


Figure 3. Symbols of boxes and robots

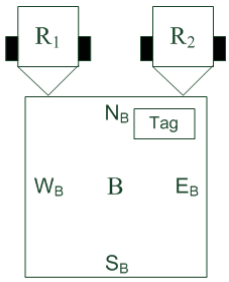
There are just a few essentially different robot positions allowed. Because two robots are needed for pushing the box, at each side of the boxes, two spaces are available for the robots manipulating them: the "counterclockwise position" and the "clockwise position" (see Fig. 4). The position is described by $P_r = [S,T]$ where r is the number of the robot, S is the side of the box where the robot touch it (N_B, E_B, S_B and W_B respectively) and the T is the turning position that means "counterclockwise position" or "clockwise position" (CC or CW).

The cooperating combination of robots is denoted by $C_{i,j}$ where i,j is the number of the robots. $C_{1,2} = P$ is the "pushing or shifting combination", when two robots (R_1 and R_2) are side by side at the same side of the box as Fig. 5 shows. In this case R_1 and R_2 are in the relative North (N_B) position. Of course, all the other three directions are similarly allowed. Any other combination of two robots is illegal, except see the next paragraph ("stopping combination").



$P_j = [N_b, CC]$ $\quad$ $P_j = [N_b, CW]$

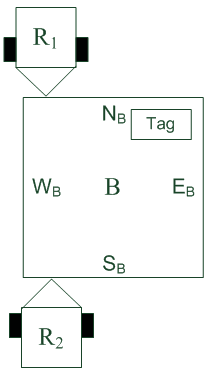
Figure 4. Robot positions at the N_B side of the box



$$\left.\begin{aligned} P_1 &= [N_B, CC] \\ P_2 &= [N_B, CW] \end{aligned}\right\} C_{1,2} = P$$

Figure 5. Allowed combinations of two robots for moving the box

Eventually, in Fig. 6, the combinations are shown where one robot intends to do a move operation, and another robot that has recognized the goal box configuration positions itself to prevent a certain move. This is an exception where a two robot combination other than the ones listed in Fig. 5 is legal as a temporary combination, clearly signalizing "stop this attempt as it is in contrary to the goal ".



$$\left.\begin{aligned} P_1 &= [N_B, CC] \\ P_2 &= [S_B, CW] \end{aligned}\right\} C_{1,2} = ST$$

Figure 6. The stopping combination

ISBN: 978-989-95079-6-8

3 Build up the codebook

After having overviewed the possible positions and combinations, let us build the part of the codebook that enables robots to recognize a situation and take action accordingly. In initial position both robots can see the goal light and know the actual position of the box. There are four good position and two pushing combination that the robots can take up.

How can take the starting position? The possible initial actions are:

1. The both robots move to reach the nearest possible good position.
2. Which reach the position first wins the temporal leader role if the other knows or guesses that this position is well for forming a useful combination. In this case the second robot turns and moves to the free position for this combination as the Fig. 7 shows. If the second robot rejects this combination then they try to take up a new form.
3. After the forming this starting combination the next is the push or shift task. We assume there are only two shifting axis, the N-S and E-S. The robots push the box a given force.

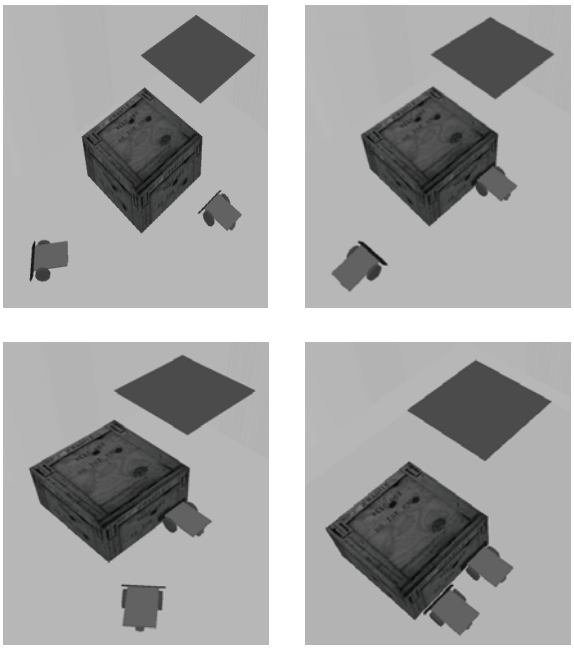


Figure 7. The robots take the starting combination

From this point a lot of scenarios are possible. Let us see some examples.

The R2 robot loses the goal sign, so it does not know the direction exactly. The R2 will slow down. The R1 robot detects the rotation of the box via its force sensors and guesses the R2 is slacken. The R1 has to make a decision, in other words it selects an action or reaction. There are two ways:

1. The R1 knows the right direction and forces to shift toward. It speeds up slightly, the R2 senses this and has to decide its reaction. The R2 makes a second self-examination and if it has not any other problem but the lost of goal sign then it switches to the *blind push* action or behavior. It means R2 pushes the box lean on R1 as follower. If R2 finds any other reason of slacken, e.g.: external obstacle or any internal error (e.g.: low battery), it keeps its own speed or decreases it. In this case the R1 gives a new reaction and so on, as long as they reach a right deal or give up the task.
2. The R1 does not sure the right direction, slow down to the speed of R2 and both robots search the goal light. If one of they find the goal takes the leader role and forces the other to push the box. If not, after a given time they stop pushing the box and move to take a new combination. After the some new unsuccessful attempt they give up.

Based on the above example and considerations it is possible to build up some elements of the action selection algorithm as a codebook. It will take the form of a decision tree, where the inputs are the direct observation, the first level outputs are intention guesses and the second level outputs the concrete actions of the corresponding robot.

Now let us see a relevant part of the codebook contains a decision tree with fuzzy elements. It is a part of the above presented example.

In R1 control system:
- Does R2 {slow down}?
  - If no then *No Action*
  - else
  - Do I know the {goal position}?
    - If no then *Goal Searching Action*
    - else
    - *Force Move Action*
  - Does R2 {accelerate}?
    - If no then *Slow Down Action*
    - else
    - *No Action*

In R2 control system:
- Can I see the {goal light}?
  - If no then *Slow Down Action*
  - .
  - .
  - .
- Does R1 {force the move}?
  - If no then *No Action* (or *Goal Searching Action*)
  - else
  - Do I have any internal or external {obstacle}?
- If no then *Blind Push Action*
- else
- *Slow Down Action* (or *Keep Move Action*)

Note that in the condition parts there are fuzzy notions which are between curly brackets. It is usually hard to judge that the R2 slows down really or the box hits an obstacle as the information come from some simple sensors that might provide only approximate results.

This simple example illustrates clearly that the meta-communication among intelligent robots by intention guessing and fuzzy evaluation of the situation might lead to effective cooperation and the achievement of task that cannot be done without collaboration and communication.

ISBN: 978-989-95079-6-8

4 Fuzzy signatures and decisions

4.1 Fuzzy signatures

In 1999 Vámos, *et al.* introduced the concept of Fuzzy Signatures [11]. Some further advanced versions of the concept and it possible use for describing complex data were later proposed in [12,13,14].

The original definition of fuzzy sets was $A : X \to [0,1]$, and was soon extended to *L-fuzzy sets* by Goguen [15]

$$A_s : X \to [a_i]_{i=1}^{k}, a_i = \begin{cases} [0,1] \\ [a_{ij}]_{j=1}^{k_i} \end{cases}, a_{ij} = \begin{cases} [0,1] \\ [a_{ijl}]_{l=1}^{k_{ij}} \end{cases} \tag{1}$$

$A_L : X \to L$, L being an arbitrary algebraic lattice. A practical special case, *Vector Valued Fuzzy Sets* was introduced by Kóczy [16], where $A_{V\,k} : X \to [0,1]^k$, and the range of membership values was the lattice of k-dimensional vectors with components in the unit interval. A further generalization of this concept is the introduction of fuzzy signature and signature sets, where each vector component is possibly another nested vector (right).

Fuzzy signature can be considered as special multidimensional fuzzy data. Some of the dimensions are interrelated in the sense that they form sub-groups of variables, which jointly determine some feature on higher level. Let us consider an example. Fig. 8 shows a fuzzy signature structure.

The fuzzy signature structure shown in Fig. 8 can be represented in vector form as follow:

$$x = \begin{bmatrix} \begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix} \\ \begin{bmatrix} x_{21} \\ \begin{bmatrix} x_{221} \\ x_{222} \\ x_{223} \end{bmatrix} \\ x_{23} \end{bmatrix} \\ \begin{bmatrix} x_{31} \\ x_{32} \end{bmatrix} \end{bmatrix}^T \tag{2}$$

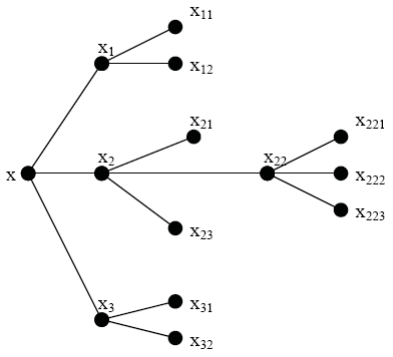


Figure 8. A Fuzzy Signature Structure

Here $[x_{11}\ x_{12}]$ from a sub-group that corresponds to a higher level compound variable of x_1. $[x_{221}\ x_{222}\ x_{223}]$ will then combine together to form x_{22} and $[x_{21}\ [x_{221}\ x_{222}\ x_{223}]\ x_{23}]$ is equivalent on higher level with $[x_{21}\ x_{22}\ x_{23}] = x_2$. Finally, the fuzzy signature structure will become $x = [x_{221}\ x_{222}\ x_{223}]$ in the example.

The relationship between higher and lower level is govern by the set of fuzzy aggregations. The results of the parent signature at each level are computed from their branches with appropriate aggregation of their child signature. Let a_1 be the aggregating associating x_{11} and x_{12} used to derive x_1, thus $x_1 = x_{11}a_1x_{12}$. By referring to Fig. 8, the aggregations for the whole signature structure would be a_1, a_2, a_{22} and a_3. The aggregations a_1, a_2, a_{22} and a_3 are not necessarily identical or different. The simplest case for a_{22} might be the min operation, the most well known t-norm. Let all aggregation be min except a_{22} be the averaging aggregation. We will show the operation based on the following fuzzy signature values for the structure in the example.

Each of these signatures contains information relevant to the particular data point x_0; by going higher in the signature structure, less information will be kept. In some operations it is necessary to reduce and aggregate information obtained from another source (some detail variables missing or simply being locally omitted). Such is when interpolation within a fuzzy signature rule base is done, where the fuzzy signature flanking an observation are not exactly of the same structure. In this case the maximal common sub-tree must be determined and all signatures must be reduced to that level in order to be able to interpolate between the corresponding branches or roots in some cases [17].

$$x = \begin{bmatrix} \begin{bmatrix} 0.3 \\ 0.4 \end{bmatrix} \\ \begin{bmatrix} 0.2 \\ \begin{bmatrix} 0.6 \\ 0.8 \\ 0.1 \end{bmatrix} \\ 0.9 \end{bmatrix} \\ \begin{bmatrix} 0.1 \\ 0.7 \end{bmatrix} \end{bmatrix}^T \tag{3}$$

After the aggregation operation is perform to the lowest branch of the structure, it will be described on higher level as:

$$x = \begin{bmatrix} 0.3 \\ \begin{bmatrix} 0.2 \\ 0.5 \\ 0.9 \end{bmatrix} \\ 0.1 \end{bmatrix}^T \tag{4}$$

Finally, the fuzzy signature structure will be:

ISBN: 978-989-95079-6-8

$$x = \begin{bmatrix} 0.3 \\ 0.2 \\ 0.1 \end{bmatrix}^T \quad (5)$$

4.2 Fuzzy signatures for box-pushing robot cooperation

Now, let us construct the fuzzy signatures for robot cooperation. The fuzzy signatures presented here are implemented in R1 robot action selection system as a part of the meta-communication codebook. Of course here is described only a slice of the whole signatures.

Fig. 9 presents the signatures which describe R2 behaviour. The information communicated to R1 is partly the observation about the last move of R2 and partly the evaluation of the situation with the box, according to those membership degrees will be attached to each leave of the actual signature.

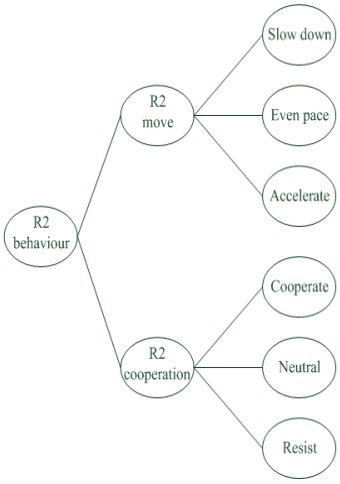


Figure 9. R2 robot's behavior fuzzy signatures

Figure 11 presents a portion of decision tree applied by R1 when it observes some action or reaction of R2 described by above mentioned behavior fuzzy signatures.

After all action selection the R1 control system reevaluates the values of the membership degrees of observed R2 behavior, environmental properties and the own state as the lines partially show in Fig. 11. After the evaluation the robot makes a decision which action is taken.

Parallel, in the R2 control system runs a similar task which reacts to the R1 decision and action. Thus, the two robots produce a circle of action and reaction (Fig. 10), where an initial action triggers the circulation of reactions and composes a meta-communication between the robots.

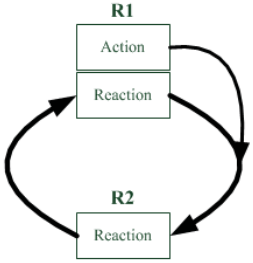
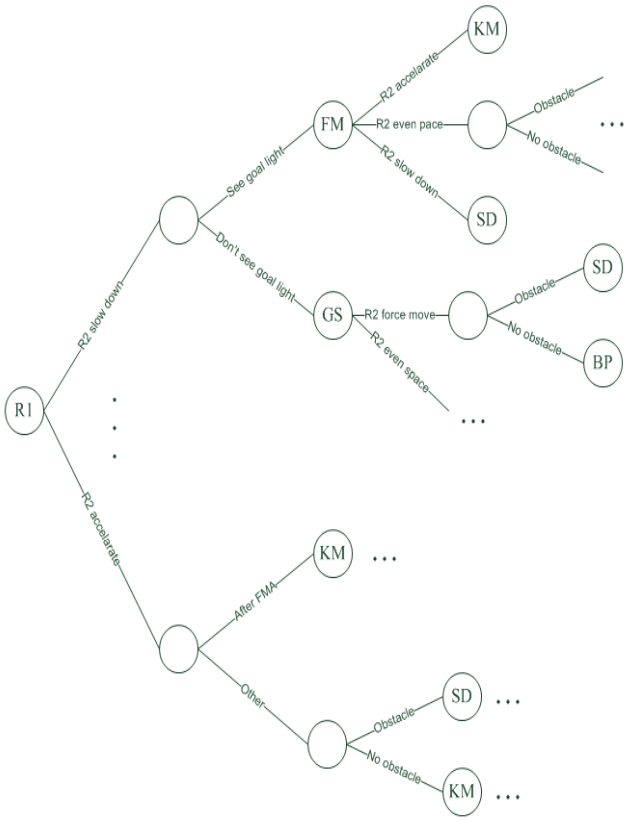


Figure 10. Action – reaction circle

This is a type of context dependent or fuzzy communication [18, 19], which means every robot has the own codebook and communicate in a vague, compressed or quasi channel. Then the robots build up and interpret this imprecise information with their codebooks.



Where the actions are

- BP – Blind push
- FM – Force move
- GS – Goal Searching
- KM – Keep move
- SD – Slow down

Figure 11. The decision tree of R1 robot

5 Conclusions

Fuzzy communication contains vague or imprecise components and it might lack abundant information. If two robots are communicating by a fuzzy channel, it is necessary that both ends possess an identical part within the codebook. The codebook might partly consist of common knowledge but it usually requires a context dependent part that is learned by communicating. Possibly it is continuously adapting to the input information. If such a codebook is not available or it contains too imprecise information, the information to be transmitted might be too much distorted and might lead to misunderstanding, misinterpretation and serious damage. If however the quality of the available codebook is satisfactory, the communication will be efficient i.e. the original contents of the message can be reconstructed. At the same time it is cost effective, as fuzzy communication is compressed as compared to traditional communication. This advantage can be deployed in many areas of engineering, especially where the use of the communication channels is expensive in some sense, or where there is no proper communication channel available at all.

ISBN: 978-989-95079-6-8

Here we illustrate clearly that the communication among intelligent robots by intention guessing and fuzzy evaluation of the situation might lead to effective cooperation and the achievement of tasks that cannot be done without collaboration and communication.
We simulated many scenarios and almost got acceptable results, but sometimes the robots made deadlock combination and gave up the work. In future we want to work out new algorithms to solve these situations.

Acknowledgment

The research was supported by HUNOROB project (HU0045), a grant from Iceland, Liechtenstein and Norway through the EEA Financial Mechanism and the Hungarian National Development Agency, the Széchenyi University Main Research Direction Grant 2009, and a National Scientific Research Fund Grant OTKA T048832 and K75711.

References

[1] P. Caloud, W. Choi, J. C. Latombe, C. L. Pape, and M. Yim, Indoor automation with many mobile robots, *in IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 67-72, 1990.

[2] B. Donald, J. Jennings, and D. Rus, Analyzing teams of cooperating mobile robots, *in IEEE International Conference on Robotics and Automation,* pp. 1896-1903, 1994.

[3] C. R. Kube and H. Zhang, Collective robotic intelligence, *in Second International Confernece on Simulationof Adaptive Behavior: From Animals to Animats*, pp. 460-468, MIT Press, (Cambridge, MA), 1992.

[4] C. R. Kube and H. Zhang, Task modelling in collective robotics, *Autonomous Robots* 4(1), pp. 53-72, 1997.

[5] C. R. Kube and E. Bonabeau, Cooperative transport by ants and robots, *Tech. Rep.* 99{01-008, Santa Fe Institute, Santa Fe, NM, 1999.

[6] M. J. Matari_c, M. Nilsson, and K. T. Simsarian, Cooperative multi-robot box-pushing, *in IEEE International Conference on Intelligent Robots and Systems*, vol. 3, pp. 556-561, 1995.

[7] F. R. Noreils, An architecture for cooperative and autonomous mobile robots, *in IEEE International Conference on Robotics and Automation*, pp. 2703-2710, 1992.

[8] L. E. Parker, ALLIANCE: An architecture for fault tolerant, cooperative control of heterogeneous mobile robots, *in IEEE/RSJ/GI International Conference on Intelligent Robots and Systems*, pp. 776-783, 1994.

[9] D. J. Stilwell and J. S. Bay, Toward the development of a material transport system using swarms of ant-like robots, *in IEEE International Conference on Robotics and Automation*, pp. 766-771, 1993.

[10] J. Schwartz and M. Sharir, On the Piano Mover's problem. II. General techniques for computing topological properties of real algebraic manifolds', *Advances in Applied Math*, 1983.

[11] T. Vámos, Gy. Bíró and L. T. Kóczy, Fuzzy signatures, *in SIC-EUROFUSE'99,* Budapest, pp. 210-217, 1999.

[12] B. S. U. Mendis, T. D. Gedeon and L. T. Kóczy, Investigation of Aggregation in Fuzzy Signatures, *in 3rd International Conference on Computational Intelligence, Robotics and Autonomous Systems*, (CD proc), Singapore, 2005.

[13] K. W. Wong, T. D. Gedeon and L. T. Kóczy, Construction of fuzzy signature from data, *in FUZZ-IEEE 2004*, Budapest, pp. 1649-1654, 2004.

[14] K. W. Wong, A Chong, T. D. Gedeon, L. T. Kóczy, and T. Vámos, Hierarchical fuzzy signature structure for complex structured data, *in International Symposium on Computational Intelligence and Intelligent Informatics (ISCIII),* Nabeul, pp. 105-109, 2003.

[15] J. A. Goguen, L-fuzzy Sets, *J. Math. Anal. Appl.* 18 pp. 145-174, 1967.

[16] L. T. Kóczy, Vectorial I-fuzzy Sets, *in M. M. Gupta & E. Sanchez (eds.): Approximate Reasoning in Decision Analysis*, North Holland, Amsterdam, pp. 151-156, 1982.

[17] T. Vámos, L.T. Kóczy and Gy. Biró, Fuzzy Signatures in Data Mining, *in IFSA World Congress and 20th NAFIPS International Conference*, Vancouver, pp. 2842-2846, 2001.

[18] L.T. Kóczy and T.D. Gedeon, Context Dependent Reconstructive Communication, *3rd International Symposium on Computational Intelligence and Intelligent Informatics*, ISCIII 2007, Agadir, Morocco, March 28-30, pp. 13-19, 2007.

[19] A. Ballagi, L.T. Kóczy and T.D. Gedeon, Local Codebook Construction for Fuzzy Communication in Cooperation of Mobile Robots, *in First Győr Symposium on Computational Intelligence*, Győr, 2008.

ISBN: 978-989-95079-6-8

A Fuzzy Filter for the Removal of Gaussian Noise in Colour Image Sequences

Tom Mélange, Mike Nachtegael, Etienne E. Kerre

Ghent University, Dept. of Applied Math. and Computer Science
Fuzziness and Uncertainty Modelling Research Group
Krijgslaan 281-S9, 9000 Ghent, Belgium
Email: {tom.melange,mike.nachtegael,etienne.kerre}@ugent.be

Abstract— *In this paper a new fuzzy filter for the removal of additive Gaussian noise in colour image sequences is presented. The proposed method, which is a colour extension of our previous work on greyscale images, consists of two subfilters. The first subfilter averages the noise in each of the colour bands separately. In this averaging, the weights of the neighbouring pixels are determined by fuzzy rules. Even though the filtering is performed in each colour band separately, information from the other colour bands is also used in the fuzzy rules. In this way, we can be more confident whether a neighbouring pixel is similar to the pixel being filtered. In order to further improve the results, a second subfilter is applied too. This colour restorating filter is based on the simplified assumption that for similar pixels the pixel value difference should be approximately the same in all colour bands. A pixel is estimated from its neighbours by estimating the differences in each colour band equal to the average over all colour bands. Experimental results show that, in terms of average PSNR and average NCD, the proposed colour extension of our previous work outperforms the usual colour extension of a greyscale filter, in which the Y component in the YUV transform is filtered with the original greyscale method and the chrominance bands U and V are averaged.*

Keywords— Video, noise filter, colour, Gaussian noise, fuzzy rule.

1 Introduction

Images sequences are among the most important information carriers in today's world. They are used in numerous applications such as broadcasting, video-phone, traffic observations, surveillance systems and autonomous navigation to name a few. The used sequences are however often affected by noise due to bad acquisition, transmission or recording. In this paper we will concentrate on image sequences corrupted with additive white Gaussian noise of zero mean and variance σ^2:

$$I_{n,i} = I_{o,i} + \epsilon_i, \qquad i = 1, \ldots, P \tag{1}$$

where $I_{n,i}$ and $I_{o,i}$ denote the i-th pixel from the noisy and the original frame respectively, $\epsilon_i \sim N(0, \sigma^2)$ and P is the number of pixels per frame.

The goal of noise filtering is not only visual improvement but also an improvement of the further analysis or coding of the sequences. During this filtering process a compromise needs to found between the removal of noise and the preservation of fine image details.

Most video filters that exist in literature are designed for greyscale sequences. Some examples are e.g. [2, 3, 4, 5]. These greyscale methods can nevertheless be extended to the RGB colour space [1] in a straightforward way by filtering each of the colour bands R, G and B separately. This might however result in the introduction of colour artefacts since the correlation between the different colour bands is neglected. Therefore, the commonly used alternative is to filter only the luminance component Y of the YUV-transform with the given greyscale method, possibly with an additional averaging of the chrominance components U and V.

The proposed filtering framework can be seen as a colour extension of our previous work presented in [4], which was inspired by the multiple class averaging filter from [5]. The presented filter consists of two separate subfilters. In the first subfilter, we add colour information to the fuzzy logic framework from our work in [4]. In this first fuzzy subfilter, each of the colour bands is denoised separately by averaging the noise. The weights assigned to the pixels considered in the averaging are determined by fuzzy rules. Even though the filtering is performed in each colour band separately, the fuzzy rules also require information from the other colour bands. Due to this increase in information, we can expect a more reliable estimation of the degree to which a neighbouring pixel is similar to the pixel that is filtered. However, especially around edges in the image, some colour artefacts might appear because sometimes not enough similar neighbours can be found to completely average the noise and it might also happen that a neighbouring pixel is wrongly considered as belonging to the same object (similar). To cope with this problem, the first subfilter is combined with an additional second subfilter, which is an extension of the second subfilter in [6]. This subfilter is based on the simplified assumption that for similar pixels the pixel value differences in the three different colour components should all three be approximately the same. The pixel being filtered is estimated from a neighbour by estimating the differences in each band equal to the average over the different colour bands.

The experimental results show that the proposed colour video denoising framework performs better than the YUV-colour extension in terms of average PSNR and NCD.

The paper is structured as follows: Section 2 gives some preliminaries about fuzzy set theory. Next, the proposed colour video denoising algorithm is explained in Section 3. Finally, experimental results are presented in Section 4.

2 Preliminaries

In this section, we will introduce some basic notions concerning fuzzy set theory and fuzzy if-then rules.

2.1 Fuzzy Sets

In classical set theory, an element x in a universe X belongs or does not belong to a certain set C, defined over the uni-

ISBN: 978-989-95079-6-8

verse X. This can be modelled by membership degrees belonging to $\{0,1\}$. In fuzzy set theory [7], there is a more gradual transition between 'belonging to' and 'not belonging to'. A fuzzy set F in the universe X is characterized by a $X \rightarrow [0,1]$ mapping μ_F, which assigns a degree of membership $\mu_F(x) \in [0,1]$ of x in the fuzzy set F, with every element x in X. This membership degree may also lie between 0 and 1, which makes fuzzy sets very useful and more natural than crisp sets to work with complex systems and human knowledge, where linguistic variables are used. In this paper, we will use a linguistic variable "large" for several parameters that will be introduced further. Using fuzzy set theory, a parameter is not necessarily large or not large, but can also be large to some degree.

2.2 Fuzzy Rules and Fuzzy Logical Operators

The general form of a fuzzy rule is "IF A THEN B", where the premise A and the consequent B are (collections of) propositions containing linguistic variables. These propositions can contain AND, OR and NOT operators, which correspond to respectively the intersection, the union and the complement of fuzzy sets.

The membership degree of an element x in the intersection (respectively union) of two fuzzy sets A and B in X is obtained through a mapping T, (respectively S) that maps the membership degrees of an element in the fuzzy sets A and B onto its membership degree in the fuzzy set $A \cap B$ (respectively $A \cup B$): $\mu_{(A \cap B)}(x) = T(\mu_A(x), \mu_B(x))$, $\forall x \in X$ (respectively $\mu_{(A \cup B)}(x) = S(\mu_A(x), \mu_B(x))$,$\forall x \in X$). In fuzzy logic for the mappings T and S, respectively a triangular norm [8] and a triangular conorm [8] are used. For the results in this paper, we have used the algebraic product and the probabilistic sum as triangular norm and conorm respectively. This norm and conorm led to the best results for our filter and are the most simple from a computational point of view.

To specify the complement of a fuzzy set A in X, we use a mapping N to derive the membership degree of an element x in the complement of the fuzzy set A from its membership degree in the fuzzy set A: $\mu_{coA}(x) = N(\mu_A(x))$, $\forall x \in X$. In fuzzy logic for the mapping N an involutive negator [8] is used. In this paper we have chosen for the well-known standard negator $N_s(x) = 1 - x, \forall x \in [0,1]$.

Let's take for example the following fuzzy rule:

Fuzzy Rule 1. `IF` (u *is* `U` `AND` v *is* V) `OR` w *is* `NOT` `W` `THEN` z *is* `Z`.

The membership degree $\mu_Z(z)$ of z in Z, which corresponds to the activation degree of the rule, i.e. the degree of thruthfulness, is then calculated as:

$$\mu_Z(z) = (\mu_U(u) \cdot \mu_V(v)) + (1 - \mu_W(w)) - (\mu_U(u) \cdot \mu_V(v)) \cdot (1 - \mu_W(w)). \quad (2)$$

3 The proposed algorithm

In this section, we will outline the proposed filtering framework. The method is a superposition of two subfilters, presented in respectively Subsection 3.1 and Subsection 3.2.

In the first subfilter a $3 \times 3 \times 2$ sliding window is used, which is moved through the frame from top left to bottom right, each time filtering the central position in the window. This window consists of 3×3 pixels in the current frame and 3×3 pixels in the previous frame as shown in Fig. 1. The central position in the window is denoted by $(\mathbf{r}, t)$, where $\mathbf{r} = (x, y)$ and t respectively stand for the spatial and temporal position in the image sequence. An arbitrary pixel position in the sliding window (which may also be the central position) is denoted by $(\mathbf{r'}, t')$, with $\mathbf{r'} = (x + k, y + l)$, $(-1 \leq k, l \leq 1)$ and $t' = t$ or $t' = t - 1$. Further, the second subfilter uses a 3×3 window in the current frame for which similar notations will be used as for the $3 \times 3 \times 2$ window.

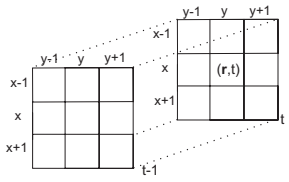


Figure 1: The $3 \times 3 \times 2$ filtering window consisting of 3×3 pixels in the current frame and 3×3 pixels in the previous frame.

3.1 First Subfilter

The subfilter explained in this subsection, can be seen as a colour extension of the filtering framework of our previous work, presented in [4]. This method was inspired by the multiple class averaging filter [5], from which we adopted the following ideas: (i) temporal blur should be avoided by taking into account only pixels from the current frame when motion is detected; (ii) image details should be preserved by filtering less when large spatial activity (details) is detected. More noise will be left, but the eye is less sensitive for the high spatial frequencies to which large spatial activity corresponds [9]. In homogeneous areas on the other hand, as much noise as possible should be removed by strong filtering.

The filtering is based on averaging the noise using the pixel values in the neighbourhood that are similar to the given pixel value and probably belong to the same object. Each colour band is filtered separately, but in the filtering of each colour band, the information from the other colour bands is used to confirm that a neighbouring pixel does indeed belong to the same object.

In the following the output of the first fuzzy subfilter is denoted by I_f, while the noisy input sequence is denoted by I_n. The output of the first subfilter for the central pixel in the window is determined as a weighted mean of the pixel values in the $3 \times 3 \times 2$ window ($i = 1, 2, 3$):

$$I_f(\mathbf{r}, t, i) = \frac{\sum_{\mathbf{r'}} \sum_{t'=t-1}^{t} W(\mathbf{r'}, t', \mathbf{r}, t, i) I_n(\mathbf{r'}, t', i)}{\sum_{\mathbf{r'}} \sum_{t'=t-1}^{t} W(\mathbf{r'}, t', \mathbf{r}, t, i)}, \quad (3)$$

The weights $W(\mathbf{r'}, t', \mathbf{r}, t, i)$ in the above weighted means are determined using fuzzy logic. The weight of a pixel is chosen equal to its membership degree in the fuzzy set "large weight". This membership degree is calculated as the activation degree of a fuzzy rule that corresponds to the ideas given at the beginning of this subsection. If large detail is detected, i.e., if a calculated detail value is large, then we should filter less by averaging only over pixels that are quite similar, i.e.,

ISBN: 978-989-95079-6-8

for which there is no large difference in the considered colour component and also in at least one of the other components. On the other hand, if there is not much detail detected, i.e., in the case that the calculated detail value is not large, strong filtering should be performed, i.e., we don't put a condition on the difference between the considered and the filtered pixel in the considered colour band (the check in the other colour bands remains for the case that the calculated detail value was not completely reliable). Further, if the pixel for which the weight is calculated belongs to the previous frame, we only want to give it a large weight if there is no motion detected in the filtering window, i.e., if a calculated motion value is not large. In the filtering of the red colour band, this results in the following two fuzzy rules, depending on whether the pixel lies in the current or the previous frame:

Fuzzy Rule 2. *Assigning the membership degree in the fuzzy set "large weight" of the weight $W(\mathbf{r'}, t', \mathbf{r}, t, 1)$ for the red value at position $\mathbf{r'}$ in the current frame ($t' = t$) of the window with central pixel position $(\mathbf{r}, t)$:*

IF $\Big($*(the detail value $d(\mathbf{r},t)$ is* LARGE AND $\Delta_1(\mathbf{r'}, t', \mathbf{r}, t)$ *is* NOT LARGE AND $(\Delta_2(\mathbf{r'}, t', \mathbf{r}, t)$ *is* NOT LARGE OR $\Delta_3(\mathbf{r'}, t', \mathbf{r}, t)$ *is* NOT LARGE$)\Big)$

OR $\Big($*(the detail value $d(\mathbf{r},t)$ is* NOT LARGE$)$ AND $(\Delta_2(\mathbf{r'}, t', \mathbf{r}, t)$ *is* NOT LARGE OR $\Delta_3(\mathbf{r'}, t', \mathbf{r}, t)$ *is* NOT LARGE$)\Big)\Big)$

THEN *the red value at position $\mathbf{r'}$ has a* LARGE *weight $W(\mathbf{r'}, t', \mathbf{r}, t, 1)$ in (3).*

Fuzzy Rule 3. *Assigning the membership degree in the fuzzy set "large weight" of the weight $W(\mathbf{r'}, t', \mathbf{r}, t, 1)$ for the red value at position $\mathbf{r'}$ in the previous frame ($t' = t - 1$) of the window with central pixel position $(\mathbf{r}, t)$:*

IF $\Big($*(the detail value $d(\mathbf{r},t)$ is* LARGE AND $\Delta_1(\mathbf{r'}, t', \mathbf{r}, t)$ *is* NOT LARGE AND $(\Delta_2(\mathbf{r'}, t', \mathbf{r}, t)$ *is* NOT LARGE OR $\Delta_2(\mathbf{r'}, t', \mathbf{r}, t)$ *is* NOT LARGE$)\Big)$

OR $\Big($*(the detail value $d(\mathbf{r},t)$ is* NOT LARGE$)$ AND $(\Delta_3(\mathbf{r'}, t', \mathbf{r}, t)$ *is* NOT LARGE OR $\Delta_2(\mathbf{r'}, t', \mathbf{r}, t)$ *is* NOT LARGE$)\Big)\Big)$

AND *the motion value $m(\mathbf{r},t)$ is* NOT LARGE

THEN *the red value at position $\mathbf{r'}$ has a* LARGE *weight $W(\mathbf{r'}, t', \mathbf{r}, t, 1)$ in (3).*

Similar fuzzy rules, switching the role of the red colour band and the colour band that needs to be filtered, are used to determine the weights $W(\mathbf{r'}, t', \mathbf{r}, t, 2)$ and $W(\mathbf{r'}, t', \mathbf{r}, t, 3)$ in expression (3) to filter the green and blue colour band respectively. In these fuzzy rules, a detail value $d(\mathbf{r}, t)$, three difference values $\Delta_1(\mathbf{r'}, t', \mathbf{r}, t)$, $\Delta_2(\mathbf{r'}, t', \mathbf{r}, t)$ and $\Delta_3(\mathbf{r'}, t', \mathbf{r}, t)$ (one for each colour band) and a motion value $m(\mathbf{r}, t)$ are used, which we will now discuss. In our proposed method, only one detail value $d(\mathbf{r}, t)$ is used for all three colour bands. This detail value depends however on three detail values computed on each colour band separately. These three detail values are equal to the standard deviation calculated in the respective colour bands on the 3×3 pixels of the filtering window belonging to the current frame ($i = 1, 2, 3$):

$$I_{av}(\mathbf{r}, t, i) = \frac{1}{9} \sum_{\mathbf{r'}} I_n(\mathbf{r'}, t, i),$$

$$d_i(\mathbf{r}, t) = \Big(\frac{1}{9} \sum_{\mathbf{r'}} \big(I_n(\mathbf{r'}, t, i) - I_{av}(\mathbf{r}, t, i) \big)^2 \Big)^{\frac{1}{2}}.$$

We don't need to know the exact value of $d(\mathbf{r}, t)$. Only the membership degree $\mu_d\big(d(\mathbf{r}, t)\big)$ of $d(\mathbf{r}, t)$ in the fuzzy set "large detail value" is needed to calculate the activation degree of Fuzzy Rules 2 and 3 that determine the weights in the expression (3). This membership degree corresponds to the activation degree of the following fuzzy rule:

Fuzzy Rule 4. *Assigning the membership degree in the fuzzy set "large detail value" of the detail value $d(\mathbf{r}, t)$ for the pixel at the central position $(\mathbf{r}, t)$ in the filtering window of the current step:*

IF $d_1(\mathbf{r}, t)$ *is* LARGE AND $d_2(\mathbf{r}, t)$ *is* LARGE AND $d_3(\mathbf{r}, t)$ *is* LARGE

THEN $d(\mathbf{r}, t)$ *is* LARGE.

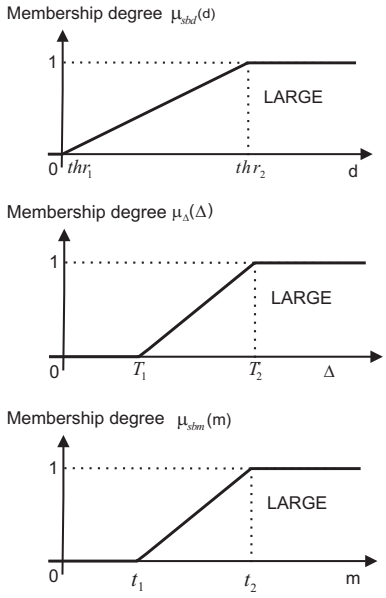


Figure 2: The respective membership functions μ_{sbd}, μ_Δ and μ_{sbm} of the respective fuzzy sets "large single band detail value", "large difference" and "large single band motion value".

The membership function μ_{sbd} of the fuzzy set "large single band detail value" is given in Fig. 2. The membership degree $\mu_d\big(d(\mathbf{r}, t)\big)$ is thus given by:

$$\mu_d\big(d(\mathbf{r}, t)\big) = \mu_{sbd}\big(d_1(\mathbf{r}, t)\big) \cdot \mu_{sbd}\big(d_2(\mathbf{r}, t)\big) \cdot \mu_{sbd}\big(d_3(\mathbf{r}, t)\big). \quad (4)$$

The three difference values $\Delta_1(\mathbf{r'}, t', \mathbf{r}, t)$, $\Delta_2(\mathbf{r'}, t', \mathbf{r}, t)$ and $\Delta_3(\mathbf{r'}, t', \mathbf{r}, t)$ that are used to determine the weights in (3) are given by ($i = 1, 2, 3$):

$$\Delta_i(\mathbf{r'}, t', \mathbf{r}, t) = |I_n(\mathbf{r'}, t', i) - I_n(\mathbf{r}, t, i)|. \quad (5)$$

The membership function μ_Δ of the fuzzy set 'large difference' is given in Fig.2.

ISBN: 978-989-95079-6-8

Analogously to the detail value $d(\mathbf{r},t)$ also only one motion value $m(\mathbf{r},t)$ is used for the filtering of all three colour bands. This value depends however again on three values computed for each of the colour bands separately. These three single band motion values are ($i=1,2,3$):

$$\begin{aligned} m_i(\mathbf{r},t) &= \Big|I_{av}(\mathbf{r},t,i) - I_{av}(\mathbf{r},t-1,i)\Big| \\ &= \Big|\frac{1}{9}\sum_{\mathbf{r'}} I_n(\mathbf{r'},t,i) - \frac{1}{9}\sum_{\mathbf{r'}} I_v(\mathbf{r'},t-1,i)\Big|. \end{aligned}$$

Just as it was the case for the detail value $d(\mathbf{r},t)$, we don't need to know the exact value of $m(\mathbf{r},t)$ either. Only its membership degree $\mu_m\big(m(\mathbf{r},t)\big)$ in the fuzzy set "large motion value" is needed for the Fuzzy Rules 2 and 3 (or thus the calculation of the weights in (3)). This membership degree is obtained from the following fuzzy rule:

Fuzzy Rule 5. *Assigning the membership degree in the fuzzy set "large motion value" of the motion value $m(\boldsymbol{r},t)$ for the pixel at the central position $(\boldsymbol{r},t)$ in the filtering window of the current step:*

IF ($m_1(\boldsymbol{r},t)$ *is* LARGE AND $m_2(\boldsymbol{r},t)$ *is* LARGE) OR
($m_1(\boldsymbol{r},t)$ *is* LARGE AND $m_3(\boldsymbol{r},t)$ *is* LARGE) OR
($m_2(\boldsymbol{r},t)$ *is* LARGE AND $m_3(\boldsymbol{r},t)$ *is* LARGE)
THEN $m(\boldsymbol{r},t)$ *is* LARGE.

The membership function μ_{sbm} of the fuzzy set "large single band motion value" is given in Fig 2. The membership degree $\mu_m\big(m(\mathbf{r},t)\big)$ is thus given by:

$$\mu_m\big(m(\mathbf{r},t)\big) = \alpha + (\beta+\gamma-\beta\cdot\gamma) - \alpha\cdot(\beta+\gamma-\beta\cdot\gamma), \quad (6)$$

with

$$\begin{aligned} \alpha &= \Big(\mu_{sbm}\big(m_1(\mathbf{r},t)\big)\cdot\mu_{sbm}\big(m_2(\mathbf{r},t)\big)\Big), \\ \beta &= \Big(\mu_{sbm}\big(m_1(\mathbf{r},t)\big)\cdot\mu_{sbm}\big(m_3(\mathbf{r},t)\big)\Big), \\ \gamma &= \Big(\mu_{sbm}\big(m_2(\mathbf{r},t)\big)\cdot\mu_{sbm}\big(m_3(\mathbf{r},t)\big)\Big). \end{aligned} \quad (7)$$

Summarized, the membership degree of the pixel at position $(\mathbf{r'},t')$ in the fuzzy set "large weight", which corresponds to the weight $W(\mathbf{r'},t',\mathbf{r},t,1)$ in (3) is thus given by

$$W(\mathbf{r'},t',\mathbf{r},t,1) = \omega\cdot\theta\cdot\phi + (1-\omega)\cdot\phi - (\omega\cdot\theta\cdot\phi)\cdot((1-\omega)\cdot\phi), \quad (8)$$

where

$$\begin{aligned} &\omega = \mu_d(d(\mathbf{r},t)) \\ &\theta = (1-\mu_\Delta(\Delta_1(\mathbf{r'},t',\mathbf{r},t))) \\ &\phi = (1-\mu_\Delta(\Delta_2(\mathbf{r'},t',\mathbf{r},t))) + (1-\mu_\Delta(\Delta_3(\mathbf{r'},t',\mathbf{r},t))) \\ &\quad - (1-\mu_\Delta(\Delta_2(\mathbf{r'},t',\mathbf{r},t)))\cdot(1-\mu_\Delta(\Delta_3(\mathbf{r'},t',\mathbf{r},t))) \end{aligned}$$

For pixel positions in the window belonging to the current frame. For pixel positions in the window belonging to the previous frame, an extra factor $1-\mu_m(m(\mathbf{r},t))$ is needed.

3.2 Second Subfilter

Because sometimes not enough similar neighbours can be found to completely average the noise in the first subfilter and because some pixels might have been wrongly considered similar in the first subfilter, some colour artefacts might still be present after applying the first subfilter. To further improve the result, the first subfilter is combined with an additional second subfilter, which is an extension of the second subfilter in [6]. Based on the simplified assumption that the difference between similar pixels is approximately the same in all three colour bands, a pixel is estimated from a neighbour by estimating a difference in a given colour component equal to the average over all three colour bands. So a difference that is larger than the average is made smaller and vice versa. The final output is a weighted average over the estimations obtained from the different neighbours, where the weight is the degree to which we believe that the neighbour belongs to the same object. The weights are introduced because for neighbours not belonging to the same object, the simplified assumption does not hold.

3.2.1 Local Differences and Correction Terms

As mentioned before, for this second subfilter, a 3×3 sliding window is used. In each step the central pixel in this window, at position $(\mathbf{r},t)$ in the image sequence, is filtered. For each pixel in the sliding window, local differences (gradients) in the three colour bands (each separately) are calculated. The differences in the red, green and blue neighbourhoods are respectively denoted by LD_1, LD_2 and LD_3 and they are calculated based on the output of the first subfilter ($i=1,2,3$):

$$LD_i(\mathbf{r'},t,\mathbf{r},t) = I_f(\mathbf{r'},t,i) \;-\; I_f(\mathbf{r},t,i). \quad (9)$$

Next, for each position in the window one correction term is determined using the calculated local differences. This correction term is defined as the average of the local difference in the red, green and blue component at the given position:

$$\epsilon(\mathbf{r'},t,\mathbf{r},t) = \frac{1}{3}\sum_{i=1}^{3} LD_i(\mathbf{r'},t,\mathbf{r},t). \quad (10)$$

3.2.2 Output of the second subfilter

In [6], the output for each component of the central pixel is an average of the corresponding components of the neighbourhood pixels, corrected with the corresponding correction term ($i=1,2,3$):

$$Out(\mathbf{r},t,i) = \frac{\sum_{\mathbf{r'}}\Big(I_f(\mathbf{r'},t,i) - \epsilon(\mathbf{r'},t,\mathbf{r},t)\Big)}{9}. \quad (11)$$

However, pixels that belong to another object and that have another colour, have a negative influence on the output. In homogeneous areas, neighbouring pixels are expected to be almost the same, and the local differences to be almost 0. So the method further averages the remaining differences caused by the noise. For a pixel belonging to another object however, the assumption that the local differences are expected to be equal in all components does not always hold. Therefore we assign weights $WT(\mathbf{r'},t,\mathbf{r},t)$ to the neighbouring pixels, based on whether they are expected to belong to the same object or not. To make this decision, we use the Euclidian distance between the central pixel and the considered neighbourhood pixel, given by

$$\delta(\mathbf{r'},t,\mathbf{r},t) = \big(\sum_{i=1}^{3} LD_i(\mathbf{r'},t,\mathbf{r},t)^2\big)^{\frac{1}{2}} \quad (12)$$

ISBN: 978-989-95079-6-8

The weights themselves are then calculated using the following fuzzy rule that expresses that the value $\delta(\mathbf{r'}, t, \mathbf{r}, t)$ should not be large. Otherwise, the considered pixel is expected to belong to another object.

Fuzzy Rule 6. *Assigning the weight in the second subfilter for the pixel at position* $(\boldsymbol{r'}, t)$ *in the filtering window:*

IF $\delta(\boldsymbol{r'}, t, \boldsymbol{r}, t)$ *is* NOT LARGE

THEN *the pixel at position* $(\boldsymbol{r'}, t)$ *has a* LARGE WEIGHT $WT(\boldsymbol{r'}, t, \boldsymbol{r}, t)$ *in the second subfilter.*

The membership function μ_δ that determines the fuzzy set "large Euclidian distance" is depicted in Fig 3. The weights in

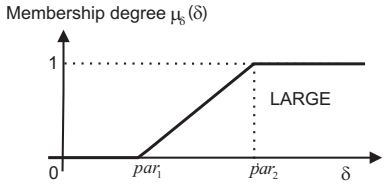


Figure 3: The membership function μ_δ of the fuzzy set "large Euclidian distance".

the filtering are again chosen equal to their membership degree in the fuzzy set "large weight", i.e., $WT(\mathbf{r'}, t, \mathbf{r}, t) = 1 - \mu_\delta(\delta(\mathbf{r'}, t, \mathbf{r}, t))$.

Finally, if not $WT(\mathbf{r'}, t', \mathbf{r}, t) = 0$ for all neighbouring pixels in the 3×3 window, the output of the second subfilter for the central pixel in the window is determined as follows ($i = 1, 2, 3$):

$$Out(\mathbf{r}, t, i) = \frac{\sum_{\mathbf{r'}} WT(\mathbf{r'}, t, \mathbf{r}, t)\Big(I_f(\mathbf{r'}, t, i) - \epsilon(\mathbf{r'}, t, \mathbf{r}, t)\Big)}{\sum_{\mathbf{r'}} WT(\mathbf{r'}, t', \mathbf{r}, t)} \tag{13}$$

where $\epsilon(\mathbf{r'}, t, \mathbf{r}, t)$ is the correction term for the components of the neighbouring pixel at position $(\mathbf{r'}, t')$. If the central pixel is so corrupt that all neighbouring pixels get a weight equal to zero, the output is calculated by giving all neigbhouring pixels in the window a weight equal to 1 and the corrupt central pixel the weight 0 ($i = 1, 2, 3$):

$$Out(\mathbf{r}, t, i) = \frac{\sum_{\mathbf{r'} \neq \mathbf{r}} \Big(I_f(\mathbf{r'}, t, i) - \epsilon(\mathbf{r'}, t, \mathbf{r}, t)\Big)}{8} \tag{14}$$

4 Experimental Results

In this section we present some experimental results obtained from the test sequences "Salesman", "Tennis", "Chair" and "Flowers", corrupted with additive Gaussian noise of zero mean and standard deviation $\sigma_n = 5, 10, 15, 20, 25$.

To draw conlusions about the proposed filtering framework, we have used the peak signal to noise ratio (PSNR) and the normalized colour difference (NCD) as measures of objective similarity and dissimilarity between the original and the filtered frames [6]. The higher the PSNR and the lower the NCD, the more similar the original and the filtered frame.

4.1 Parameter Selection

The parameters, that determine the membership functions in the above described filtering framework, have been set as follows. For the respective noise levels $\sigma_n = 5, 10, 15, 20$, the optimal parameters in terms of the mean PSNR values averaged over the sequences "Salesman", "Tennis", "Flowers" and "Chair" have been determined by letting them vary over a range of possible values. As illustrated in Fig. 4 for the parameter thr_2, this led to a linear relationship between these optimal values and the noise level. Hence, the parameters are set as the best fitting line through the observations, as shown in Fig. 4. The equations of those straight lines are given in Table 1, where σ_n stands for the standard deviation of the Gaussian noise. If this standard deviation is not known, it can be estimated by e.g. the method from [10].

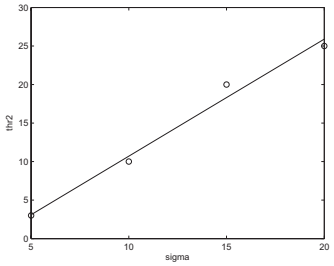


Figure 4: Optimal parameter values in terms of the PSNR.

Table 1: The used parameter values.

Parameter	Value	Parameter	Value
thr_1	0	t_1	$0.72\sigma_n - 4.0$
thr_2	$1.52\sigma_n - 4.5$	t_2	$2.22\sigma_n - 4.5$
T_1	0	par_1	$1.1\sigma_n - 7.5$
T_2	$3.14\sigma_n - 1.0$	par_2	$6\sigma_n + 35$

4.2 Experiments

In the experiments, we have compared the proposed filter (denoted by FMDAF-CR) to the filtering scheme in which a wavelet extension of the original pixel domain greyscale method [4], which outperforms both the pixel domain filter and other state-of-the-art greyscale methods of a similar complexity (as shown in [4]), is applied on the Y component of the YUV transform, followed by an averaging (3×3 window) of the chrominance components U and V (denoted by FMDAF-YUV). The average PSNR and NCD values found for the test sequences corrupted with different noise levels and processed by these two above approaches are given in Tables 2 and 3 respectively. From these tables, we can conclude that the proposed colour extension performs better in terms of average PSNR and average NCD than the commonly used YUV-filtering. For a visual comparison, we have made the original and noisy ($\sigma_n = 15$) "Salesman" sequence and the results after applying respectively the FMDAF-CR and the FMDAF-YUV filter available on `http://users.ugent.be/~tmelange/colourrule`. If

ISBN: 978-989-95079-6-8

Table 2: Average PSNR value for the processed sequences.

Sequence	σ_n	FMDAF-CR	WRFMDAF-YUV
"Salesman"	5	38.34	37.37
	10	34.36	33.64
	15	32.07	31.20
	20	30.45	29.35
	25	29.18	27.88
"Tennis"	5	36.51	33.40
	10	32.75	29.83
	15	30.35	27.92
	20	28.68	26.68
	25	27.39	25.59
"Chair"	5	39.27	39.92
	10	36.17	35.67
	15	33.94	32.99
	20	32.17	30.89
	25	30.67	29.16
"Flower garden"	5	33.32	29.03
	10	30.07	26.39
	15	27.65	24.98
	20	25.87	23.67
	25	24.47	22.44

Table 3: Average NCD value for the processed sequences.

Sequence	σ_n	FMDAF-CR	WRFMDAF-YUV
"Salesman"	5	0.0446	0.0531
	10	0.0676	0.0829
	15	0.0861	0.1121
	20	0.1027	0.1397
	25	0.1183	0.1652
"Tennis"	5	0.0269	0.0339
	10	0.0376	0.0475
	15	0.0470	0.0601
	20	0.0555	0.0724
	25	0.0638	0.0848
"Chair"	5	0.0132	0.0122
	10	0.0182	0.0215
	15	0.0238	0.0309
	20	0.0296	0.0404
	25	0.0353	0.0498
"Flower garden"	5	0.0543	0.0792
	10	0.0743	0.0945
	15	0.0907	0.1064
	20	0.1041	0.1190
	25	0.1157	0.1319

we concentrate on the side of the phone, we see that for the FMDAF-YUV method more colour artefacts (green and red spots) are visible than for the FMDAF-CR filter, which might be an explanation for the better PSNR and NCD values. Further, we also see that the wavelet domain method FMDAF-YUV has removed more noise and produces a smoother result. This smoother result can however be attributed to the use of a wavelet domain filter. If the original pixel domain greyscale method would have been used, we would also have had a little more noise remaining. Remark also that the smoother result also has as a result that the details have been smoothed a little more. This trade-off between noise removal and detail preservation is one of the main challenges in the development of a noise filter.

Acknowledgment

This work was financially supported by the GOA-project B/04138/01 IV 1 of Ghent University.

References

[1] S.J. Sangwine and R.E.N. Horne. *The colour image processing handbook*, Chapman&Hall, London, 1998.

[2] E.J. Balster, Y.F. Zheng and R.L. Ewing. Combined spatial and temporal domain wavelet shrinkage algorithm for video denoising. *IEEE Transactions on Circuits and Systems for Video Technology*, 16(2):220–230, Feb. 2006.

[3] S.M. Mahbubur Rahman, M. Omair Ahmad and M.N.S. Swamy. Video denosing based on inter-frame statistical modeling of wavelet coefficients. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(2):187–198, Feb. 2007.

[4] T. Mélange et al. A new fuzzy motion and detail adaptive video filter. *Proc. of ACIVS 2007, LNCS*, 4678:640–651, 2007.

[5] V. Zlokolica, A. Pizurica and W. Philips. Video denoising using multiple class averaging with multiresolution. *International Workshop VLBV03, LNCS,* 2849:172–179, 2003.

[6] S. Schulte et al. A New Fuzzy Additive Noise Reduction Method. *Proc. of ICIAR 2007, LNCS* 4633:12–23, 2007.

[7] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353,1965.

[8] S. Weber. A general concept of fuzzy connectives, negations and implications based on t-norms and t-conorms. *Fuzzy Sets and Systems*, 11(2):115–134, 1983.

[9] E.B. Bellers and G. De Haan. *De-interlacing: A Key Technology for Scan Rate Conversion*, Elsevier Science B.V., Sara Burgerhartstraat, Amsterdam, 2000.

[10] D. Donoho and I. Johnstone. Ideal spatial adaptation by wavelet shrinkage, *Biometrika*, 8:425–455, 1994.

ISBN: 978-989-95079-6-8