Visualising Rough Time Intervals in a Two-Dimensional Space

Yi Qiang[1], Katrin Asmussen[1], Matthias Delafontaine[1], Birger Stichelbaut[2], Guy De Tré[3], Philippe De Maeyer[1], Nico Van de Weghe[1]

1 CartoGIS Cluster, Dept. of Geography, Ghent University, Belgium
2 Dept. of Archaeology and Ancient History of Europe, Ghent University, Belgium
3 Dept. of Telecommunications and Information Processing, Ghent University, Belgium

E-mail: {Yi.Qiang, Katrin.Asmussen, Matthias.Delafontaine, Birger.Stichelbaut, Guy.DeTre, Philippe.DeMaeyer, Nico.VandeWeghe}@UGent.be

Abstract *— A lot of disciplines (e.g. archaeology) have to process imprecise temporal information. There are different possibilities to handle this kind of information, amongst them e.g. fuzzy set theory and rough set theory. In this paper, due to its capability in the context of many data acquisition applications, the focus has been set on rough set theory. To illustrate temporal information, an interval is often visualised by means of a one-dimensional segment in a one-dimensional space. An alternative representation of time intervals is called the Triangular Model (TM) by which a time interval is represented by a point in a two-dimensional space. In this paper, rough set theory is applied into TM, which gets extended to the Rough Triangular Model (RTM). In RTM, Rough Time Intervals (RTI) and their mutual relations can be visualised diagrammatically, which offers opportunities to visualise and analyse imprecise temporal information. Aerial photos, taken during World War I, containing imprecise temporal information with archaeological background, are used to illustrate the potentials of the model in processing RTI.*

Keywords — Imprecise temporal information, rough set theory, temporal reasoning, temporal relation, time interval, triangular model.

1 Introduction

For time intervals, the most widely adopted representation is the linear model where time intervals are modelled as finite linear segments in a one-dimensional space. Much research has been carried out on representing and reasoning about time intervals, most of which is based on this linear concept and simple temporal relations [1] [3] [6] [7] [10]. Though Rough Time Intervals (RTI), i.e. intervals starting and/or ending at uncertain time stamp are also frequently used in many disciplines, there is still a shortage of methods and tools to visualise them. An alternative temporal model, the Triangular Model (TM), has been proposed in [11] [12] [2]. This model is based on the W-diagram introduced in [4] [5]. Up till now, TM has remained mainly a theoretical concept. However, it seems to offer a promising design for several applications.

A lot of disciplines (e.g. archaeology, geography, psychology, and philosophy) are faced by the problem of having imprecise temporal information. To handle this information it can be reverted to different approaches, for instance to fuzzy set theory [13]. This theory aims to formalise inherently fuzzy concepts by permitting the gradual assessment of the membership of an element in relation to a set. Another way to deal with imprecise temporal information can be found in rough set theory [8] [9]. This theory introduces a concept of lower and upper approximation and a boundary region, describing a set where elements can or can not be decisively classified into a set X. In contrast to fuzzy set theory, rough set theory is particularly useful in the context of many data acquisition applications. Therefore, in this paper the rough set theory has been chosen to be applied into the TM.

This paper extends TM to RTM in order to visualise and analyse RTIs. First, the basics of TM are introduced in section two. The remainder of this section describes RTIs and their visualisation in the RTM. In section three, we illustrate how to visualise rough time relations of RTIs by means of the RTM. This section is followed by a description of an application of the RTM to incomplete temporal data which is deriving from an archaeological background. Finally, conclusions are drawn and future work is pointed out.

2 The Triangular Model

2.1 Representing time intervals with TM

Time is usually conceptualised as a linear, one-dimensional time line (Fig. 1). In this classical concept, a temporal interval I is visualised by a segment that is bounded by a begin point I^- and end point I^+. In the linear model, the vertical dimension is only used to differentiate multiple overlapping intervals, if used at all. An interval without duration is visualised as a point (zero dimension). The basic concept of TM is the construction of two lines through the extremes of a linear time interval (Fig. 2). For each time interval I, two straight lines (L_1 and L_2) are constructed, with L_1 going through I^-; L_2 going through I^+, and where the angle $\alpha_1 = \alpha_2$. The intersection of L_1 and L_2 is called the interval point I. The position of I in the two-dimensional space completely determines both, the beginning and end point of the interval (Fig. 3).

ISBN: 978-989-95079-6-8

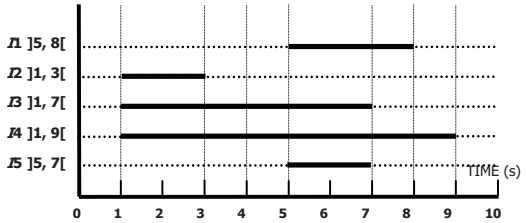


Figure 1: Linear representation of time intervals.

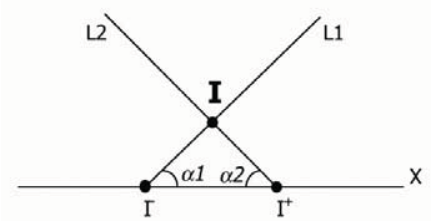


Figure 2: Construction of a simple interval point.

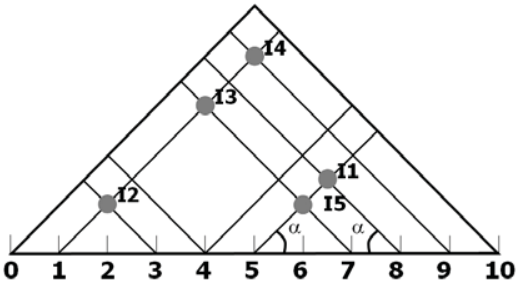


Figure 3: Time intervals in TM.

2.2 Representing RTIs with TM

It is often difficult for scientists to obtain precise temporal information about events or processes. Geological periods, for example, are always associated with uncertain starting and ending times. In history and archaeology as well, many events are lacking precise temporal demarcation. For instance, consider the impreciseness coming along with radiocarbon dating.

Scientifically underpinned treatments for imprecise information handling are already existing. For instance, fuzzy set theory [13] aims to formalize inherently fuzzy concepts by permitting the gradual assessment of the membership of an element in relation to a set. It extends conventional (crisp) set theory and handles the concept of partial truth, i.e. truth values between 0 (complete false) and 1 (complete true).

Another way to deal with imprecise information can be found in rough set theory [8] [9]. The main purpose of rough set theory is the induction of approximations of concepts, which are represented by the upper approximation $\overline{B}$ and the lower approximation $\underline{B}$. Within $\underline{B}$, elements can be decisively classified into a set X; outside $\overline{B}$, elements are not members of X. The difference between $\overline{B}$ and $\underline{B}$ forms a boundary region. If the boundary region is nonempty, a set is said to be rough; otherwise the set is crisp. In the boundary region, elements can not be decisively classified as members or not members of X.

Rough set theory is particularly useful in the context of temporal information, due to the nature of data acquisition in many scientific applications. Remote sensing, for instance, is a world-wide applied tool that relies on images taken at discrete time stamps. One can determine the state of a feature on these snapshots, whereas this state is uncertain in between two time stamps. However, we rely on the very natural assumption that this state does not change in between two snapshots which show similar states, i.e. the uncertain parts only remain in between two snapshots showing different states. Hence, we might consider a period of snapshots showing similar states (closed time interval) as a lower approximation for this state and its neighbouring periods of uncertainty (two open time intervals) as boundary region; both then cumulate into the state's upper approximation (open time interval) (Fig. 4). Thus, a feature's state can be considered as a rough set or more specifically a RTI. Next to remote sensing, this idea applies to numerous other fields, such as, soil core sampling, socio-economical census and surveys, and opinion polls.

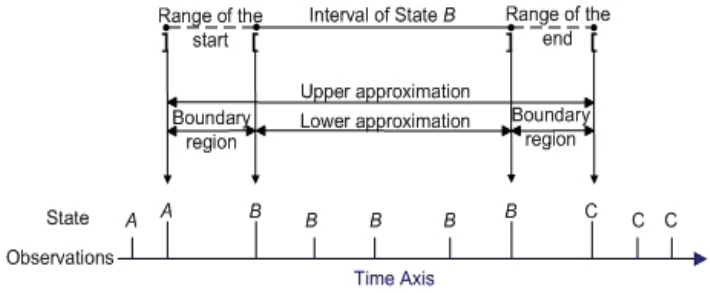


Figure 4: Rough set interval in remote sensing.

In the classical linear model, RTIs are represented as linear segments with an uncertain beginning range and an uncertain end range (Fig. 5). If there is a huge amount of rough intervals, it is quite difficult for humans to abstract information from this representation. Therefore the analyse capacities of the linear model are quite limited. For an advanced temporal analysis based on visualisation an alternative is needed.

In TM, a RTI is represented by a polygon. Four lines are constructed respectively from the earliest/latest beginning and the earliest/ latest end, forming a diamond (Fig. 6). This polygon indicates a zone within which the uncertain interval can be found. We call this representation the Rough Triangular Model (RTM).

In some cases, the beginning and end range have intersections. Take I_5 in Fig. 6 for example, we only know that it starts between 3 and 5, and ends between 3 and 5. In this case, the interval is represented as a triangle incumbent on the x-axis in RTM. In some other cases, intervals are partially rough: they have a precise beginning/end, but an imprecise end/beginning. These semi-rough intervals can be represented by lines (e.g. I_2 in Fig. 6). There are still other possibilities such as semi-open intervals, two-sided open intervals and totally indefinable intervals. In this paper, we are focussing on the

ISBN: 978-989-95079-6-8

first type of RTIs which are having a twofold, non-overlapping boundary region.

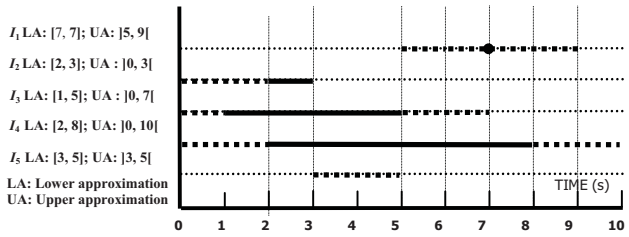


Figure 5: Linear representation of RTIs[1].

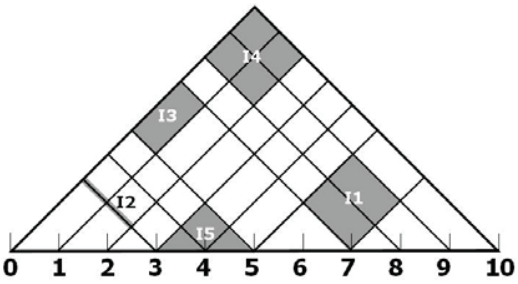


Figure 6: Visualising RTIs in RTM.

3 Rough temporal relations

When time intervals are rough, the relations between them also become rough. RTM provides the possibility of visualising and analysing temporal relations in a two-dimensional space

3.1 Representing fine temporal relations in TM

Based on the basic work of Allen [1], an interval I is represented as a pair (I^- ; I^+) with real numbers I^- and I^+, denoting respectively the beginning and end points of the interval. This means that Allen only deals with simple intervals having a specified duration. Let the beginning I^- and end point I^+ of two simple intervals have the following three possible relations: smaller than (<), equal (=) and larger than (>). Then, thirteen possible fine relationships between two intervals can be defined (see Tab. 1).

Table 1: Thirteen Allen relations [1].

Allen's temporal relations			
I_1 *equal* I_2	if $I_1^- = I_2^-$	^ $I_1^+ = I_2^+$	
I_1 *starts* I_2	if $I_1^- = I_2^-$	^ $I_1^+ < I_2^+$	
I_1 *started-by* I_2	if $I_1^- = I_2^-$	^ $I_2^+ < I_1^+$	
I_1 *finishes* I_2	if $I_1^+ = I_2^+$	^ $I_1^- > I_2^-$	
I_1 *finished-by* I_2	if $I_1^+ = I_2^+$	^ $I_2^- > I_1^-$	
I_1 *meets* I_2	if $I_1^+ = I_2^-$		
I_1 *met-by* I_2	if $I_2^+ = I_1^-$		
I_1 *overlaps* I_2	if $I_2^- > I_1^-$	^ $I_1^+ < I_2^+$	^ $I_1^+ > I_2^-$
I_1 *overlapped-by* I_2	if $I_1^- > I_2^-$	^ $I_1^- < I_2^+$	^ $I_2^+ < I_1^+$
I_1 *during* I_2	if $I_1^- > I_2^-$	^ $I_1^+ < I_2^+$	
I_1 *contains* I_2	if $I_2^- > I_1^-$	^ $I_2^+ < I_1^+$	
I_1 *before* I_2	if $I_1^+ < I_2^-$		
I_1 *after* I_2	if $I_2^+ < I_1^-$		

[1] In this paper, we use ISO standard notation to distinguish open intervals and close intervals. In this notation,]a, b[denotes open interval, [a, b] denotes close interval and]a, b] denotes left-open but right-close intervals. Since Allen's intervals are open, we use]a,b[here to denotes Allen's intervals.

Using TM [12], these relations can be visualised. Each relation thereby corresponds to a specific Fine Relation Zone (FRZ) within TM. Given a study period beginning at 0 and ending at 100, all examined intervals are located within the isosceles triangular of I]0, 100[. To obtain the best visualisation, the reference interval I_2]33,66[is chosen to be located in the centre of TM. As shown in Fig. 7b several intervals (I_{1a}, I_{1b}, I_{1c}, I_2) may exist *before* interval I_2]33,66[. All possible intervals for which $I_1^+ < I_2^-$ applies are generalised, with respect to interval I_2, into the FRZ *before,* displayed by the black triangle in Fig. 7c. Note that as Allen worked with open intervals, also the interval zone corresponding to I]0,33[is open. The right boundary of FRZ *before* represents all intervals for which applies $I_1^+ = I_2^-$. Therefore, the intervals have their end point at 33, resulting in the *meets* relationship.

Comparing the visualisations of TM with the linear model, both visualise the same fine interval relations, as displayed in Fig. 7a and Fig. 7c. The benefit of TM in visualising time relations gets spontaneously definite. The time intervals in Fig. 7c is faster to capture than the one in Fig. 7a. An overview of positions and names of the thirteen FRZs is given in Fig. 8.

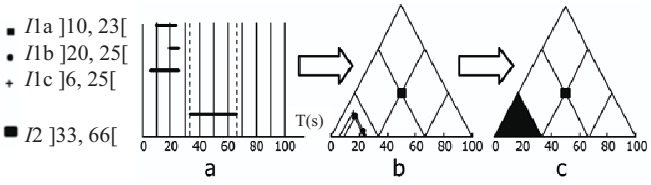


Figure 7: Visualisation of fine intervals by means of the linear model a) and TM b). FRZ *before* within TM c).

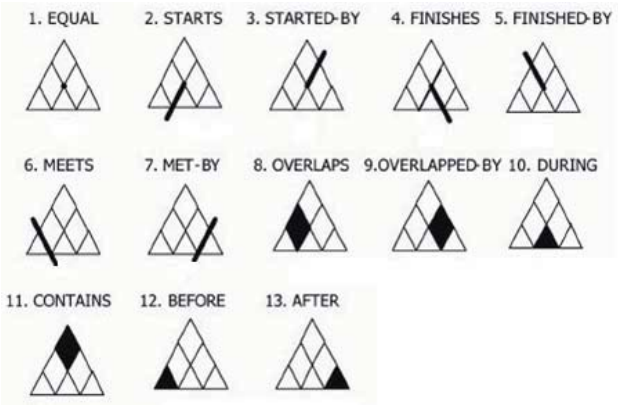


Figure 8: Thirteen FRZs in TM.

3.2 Relation zones in RTM

While TM represents fine intervals and fine interval relations, RTM represents rough intervals and its relation zones are Rough Relation Zones (RRZs).

To transform FRZs into RRZs, all point and line zones of TM (*equal, starts, started-by, finish, finished-by, meets,* and *met-*

ISBN: 978-989-95079-6-8

by) are expanded according to the duration of the boundary regions of the reference interval I_2. Hence, two polygon zones arise from the intersections of the expanded line zones *starts* and *meets* as well as from the intersections of the expanded line zones *finishes* and *met-by* (Fig. 9). Note that the absolute position of these zones may change according to the used data, while the relative position always remains.

Transforming the TM to RTM, fifteen relation zones are generated; nine of these are expanded zones (Fig. 9 and Tab. 2). The names of the expanded zones are based on the names of FRZs, but preceded by 'maybe'. Therefore these expanded relation zones are named *maybe* zones.

In TM, only one specific fine relation is possible within a FRZ (Tab. 2). Different from that, in RTM *maybe* zones represent zones where several fine relations are possible (Tab. 2). This is caused by the fact that a RRZ consists of parts of corresponding neighbouring FRZs in TM. The dashed lines in Fig. 9 represent the borders of original FRZs. In RTM, the positions of these dashed lines are uncertain and can be anywhere within the corresponding *maybe* zone. This is different from fuzzy set interval where the dashed lines have gradually-changing probability of appearance. Thus, we only use flat colour (white) to mark the interior of *maybe* zones. Note that each maybe zone contains the relations of its neighbouring zones in RTM (Fig. 9).

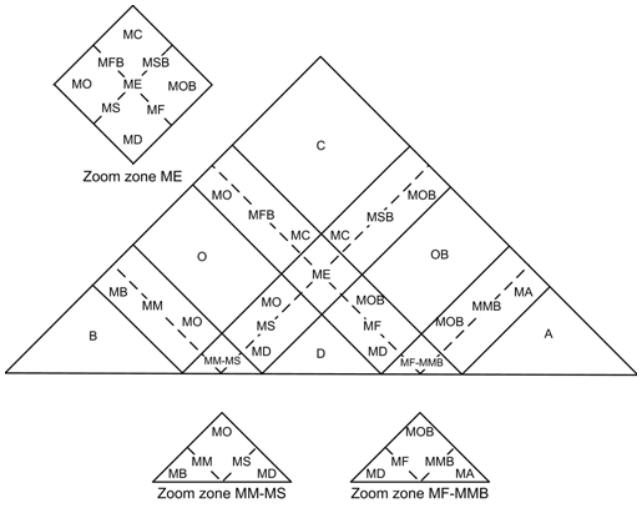


Figure 9: RRZs in RTM.

Table 2: Thirteen FRZs in TM and fifteen RRZs in RTM

FRZ		RRZ		
Name	**Abbr.**	**Name**	**possible relations**	**Abbr.**
equal	E	*maybe equal*	*contains, started-by, overlapped-by, finishes, during, starts, overlaps, finished-by, equal*	ME
starts	S	*maybe starts*	*starts, overlaps, during*	MS
started-by	SB	*maybe started-by*	*started-by, overlapped-by, contains*	MSB
finishes	F	*maybe finishes*	*finishes, during, overlapped-by*	MF
finished-by	FB	*maybe finished-by*	*overlapps, finished-by, contains*	MFB
meets	M	*maybe meets*	*meets, before, overlaps*	MM
met-by	MB	*maybe met-by*	*met-by, overlapped-by, after*	MMB
overlaps	O	*overlaps*	*overlaps*	O
overlapped-by	OB	*overlapped-by*	*overlapped-by*	OB
during	D	*during*	*during*	D
contains	C	*contains*	*contains*	C
before	B	*before*	*before*	B
after	A	*after*	*after*	A
/	/	*maybe meets or maybe starts*	*before, meet, overlaps, during, starts*	MM-MS
/	/	*maybe finishes or maybe met-by*	*during, finishes, overlapped-by, met-by, after*	MF-MMB

Unlike with FRZs, the border lines of RRZs do not make up separate relation zones but belong to one of their neighbouring RRZs. Due to their definition, the boundary regions in the beginning and the end of RTIs are open intervals (Fig. 4). The corresponding assignments are illustrated in Fig. 10, where the arrows are pointing into the zone to which the respective border line is assigned.

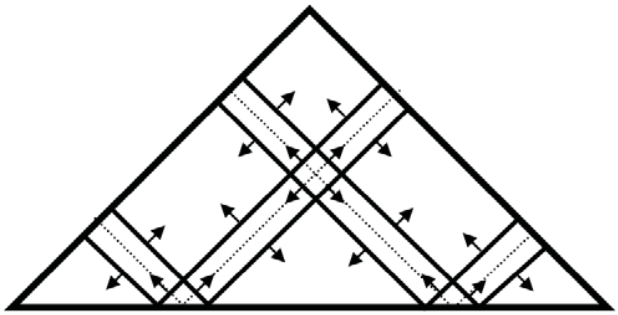


Figure 10: Assignment of the borders of RRZs.

3.3 Applying RTM

In order to evaluate whether RTM is helpful for the analysis of imprecise temporal information; we tested RTM with data deriving from an archaeological context. During World War I, aerial photos covering the Belgian-German front line in West-Flanders (Belgium) have been taken. From these aerial photos, we can observe whether a feature such as a fire trench, a gun position or a barrack, was not yet present, was present or was destroyed. For each feature, several photos with according date of acquisition exist. Hence, if we can find (1)

ISBN: 978-989-95079-6-8

the last photo where the feature is not yet found, (2) the first photo where the feature is found, (3) the last photo where the feature is found, and (4) the first photo where the feature is destroyed, we can derive a RTI representing the feature's uncertain lifetime (Fig. 4).

In RTM, fifteen RRZs can be created for one feature (Fig. 9). Lots of similar intervals will yield overlapped RRZs. If different grey shades are added to denote the number of overlaps in the concerning area, there will appear some patterns which reflect characteristics of the underlying data.

Fig. 11 and Fig. 12 visualise the overlaps of *maybe equal* zones of gun positions and breastworks respectively. More overlapped areas are marked with dark grey, and vice versa. Comparing these two visualisations, we can have a direct overview about how the two types of features are temporally distributed. As in the model most polygons of gun positions are distributed right of the breastworks zones, comparing Fig. 11 with Fig. 12, we may observe that gun positions exist relatively later than breastworks. It also can be observed that breastworks generally exist longer than gun positions because most of the breastworks zones are higher on the *y*-axis than the zones of the gun positions.

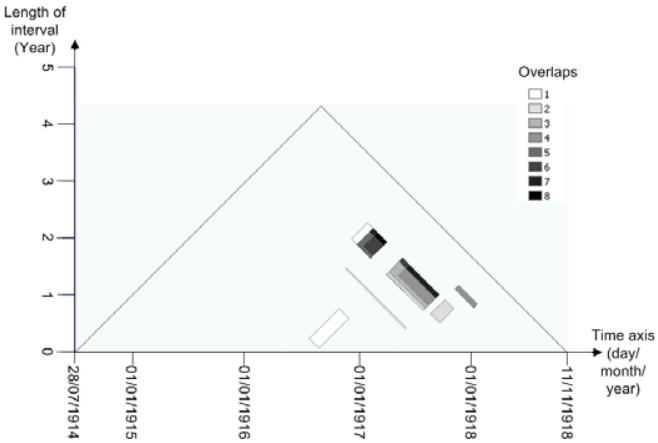


Figure 11: Overlaps of maybe equal zones of gun positions.

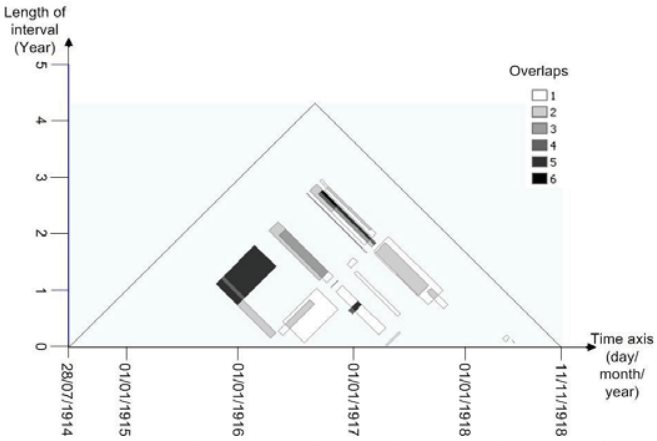


Figure 12: Overlaps of maybe equal zones of breastworks.

If we select *contain* interval zones of gun positions (Fig. 13), the display of the model can be divided into several zones according to the colour grade. If an interval point is in the dark zone, it contains intervals of most features. In other words, in the dark zone, all intervals contain most features lifetime intervals. In natural language, we could say: if an interval is in a dark zone, it witnesses the lifetime of most features. This approach can be applied to other RRZs and even to combinations of relation zones.

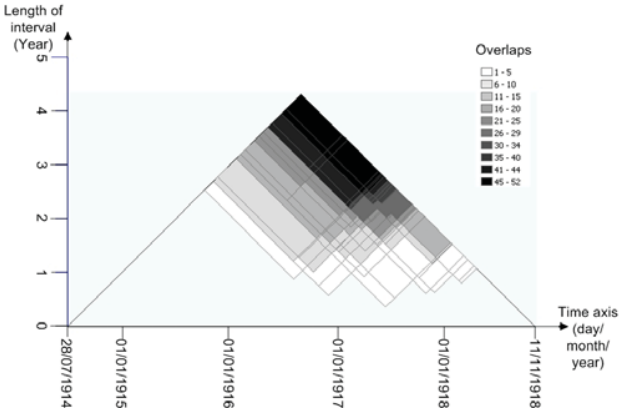


Figure 13: The overlaps of the *contain* zone of gun positions.

4 Conclusions and future work

Since a lot of disciplines (e.g. archaeology, geography, psychology, and philosophy) are faced by the problem of having imprecise temporal information we extended TM into RTM in order to visualise RTIs and temporal relations between them. We do not intend to create a new or extend an existing temporal calculus or temporal logic. Our approach takes special care for the intuitive and visualization aspects. In RTM, RTIs are represented by simple geometries (e.g. lines and polygons) in a two-dimensional space. Compared to the classical linear model, which has limited analytical capacities, RTM gives people a direct overview of the temporal distribution of RTIs. When handling a huge amount of rough intervals, RTM provides a compact visualisation pattern, which helps in further exploratory analysis. Furthermore, the temporal relations between RTIs can be visualised as zones, i.e. RRZs, in the two-dimensional space, which gives potentials in visual queries of rough temporal relations.

In this paper, we described the basics of RTM and RRZs. However, further research needs to be done.

First, as explained in 2.2; also other types of RTIs should be considered, e.g. partially rough intervals, one-sided open intervals and intervals with overlapped begin and end boundary regions. This would imply a wide modification of the RRZs. This major change in the division of the RRZ consequently has to be followed by an adaptation of visualisation and interpretation. Whether these visualisation and analyses are delivering feasible results has to be tested in further research.

ISBN: 978-989-95079-6-8

Second, the idea of conceptual neighbourhood has proved its importance in qualitative reasoning about time and space. When temporal relations become rough, there are more possibilities that one relation continuously changes to another relation without passing through other relations. Thus, the conceptual neighbourhood diagram of rough temporal relations will change accordingly. Also, the practical use of conceptual neighbourhood of temporal relations of rough intervals has to be studied in detail.

Third, so far we dedicated the borders of the RRZs of the RTM to one of the RRZs. But following the example of the TM, these lines could be relation zones by themselves. That would mean that additionally to the fifteen RRZ we would have nineteen line zones and probably six point zones. In total RTM would be build up out of forty zones. The additional line zones would include, for example, line zones of TM like *starts*, *started-by*, *finishes*, *finished-by, meets* and *met-by*. But also new line zones would be created like the line zones which are located between the zones *maybe starts or maybe meets* and *maybe meets*. Within this line zone a wide range of relations are possible. Whether this great number of RRZ provides useful visualisations and if the interpretation could be aided by a computer still needs to be determined. Also whether this interpretation would deliver helpful information still needs to be investigated.

Finally, this model can be more useful if it is implemented as an interactive tool. More flexible and interactive visualisation tools can help to better analyse complex temporal data.

Acknowledgment

The research work of Yi Qiang, Matthias Delafontaine and Birger Stichelbaut is funded by the Research Foundation–Flanders. The research work of Katrin Asmussen is co-funded by the Belgian Federal Science Policy.

References

[1] J.F. Allen. Maintaining Knowledge about Temporal Intervals. *Communications of the ACM,* 26:832-843, 1983.

[2] G. De Tré, et al. Towards a flexible Visualisation Tool for dealing with Temporal Data. *In proc. of "FQAS" LN AI*, 4027: 109-120, 2006.

[3] C. Freksa. Temporal Reasoning based on Semi-Intervals. *Artificial Intelligence*, 54:199-227, 1992.

[4] Z. Kulpa. Diagrammatic Representation for a Space of Intervals. *Machine Graphics & Vision* 6: 5-24, 1997.

[5] Z. Kulpa. Diagrammatic Representation of Interval Space in Proving Theorems about Interval Relations. *Reliable Computing* 3:209-217, 1997.

[6] V. Kumar, R. et al. Allen. Metadata Visualization for Digital Libraries: Interactive Timeline Editing and Review. *In proc. of "ACM Digital Libraries"*, Pittsburgh, USA, 1998.

[7] J.D. Mackinlay. The Perspective Wall: Detail and Context Smoothly Integrated. *In proc. "ACM CHI'91"*, New York, 1991.

[8] Z. Pawlak. Rough Sets. *Int. J. of Information and Computer Science*, 11:341-356, 1982.

[9] Z. Pawlak. Rough Sets: Theoretical Aspects of Reasoning About Data. *Kluwer Academic Publishers,* Dordrecht, The Netherlands, 1991.

[10] S. Schockaert, M.D. Cock, and E.E. Kerre. Fuzzifying Allen's Temporal Interval Relations. *IEEE Transactions on Fuzzy Systems*, 16: 517-533, 2008.

[11] N. Van de Weghe, et al. The Triangular Model as an Instrument for Visualising and Analysing Residuality. *J. of Archaeological Science* 34: 649-655, 2007.

[12] N. Van de Weghe. Development of a Conceptual Data Model for digital spatio temporal Geographical Information. *In proc. "ER", LNCS*, Tampere, Finland, 2002.

[13] L.A. Zadeh. Fuzzy Sets. *Information and Control* 8: 338-353, 1965.

ISBN: 978-989-95079-6-8

Repairing infeasibility in fuzzy goal programming

Mariano Jiménez[1], Amelia Bilbao[2], Mar Arenas[2], Mª Victoria Rodríguez-Uría[2]

1 Dpto. de Economía Aplicada I, Universidad del País Vasco,
Plaza Oñati 1. 20018 San Sebastián. Spain
2 Dpto. de Economía Cuantitativa, Universidad de Oviedo,
Avda. del Cristo s/n. 33006 Oviedo. Spain
E-mail: mariano.jimenez@ehu.es , ameliab@uniovi.es , mariamar@uniovi.es , vrodri@uniovi.es

Abstract*—The problem of solving a multi-objective programming problem, by assuming that the decision maker has fuzzy goals for each of the objectives, is addressed. In this case a fuzzy goal programming problem has to be solved. But when we have several objectives it can be a difficult task for the decision maker to determine coherent aspiration levels, so it is possible that the tolerance thresholds be inconsistent with each other, therefore obtaining an infeasible problem. We present a procedure in order to repair the infeasibility. We rebuild the unattainable fuzzy goals keeping their relaxation levels as small as possible while maintaining, at the same time, a good balance vis-à-vis the degree of achievement of the others goals.*

***Keywords*—**, Fuzzy goal programming, Goal programming, Infeasibility, Multi-objective programming.

1 Introduction

Let a multi-objective programming (MOP) problem with k objective functions $z_i(x) = c_i x$, $i = 1,...,k$ be

$$\text{(MOP)} \quad \text{Min} \quad z(x) = (z_1(x), z_2(x), ..., z_k(x))$$
$$\text{S. t.} \quad x \in X$$

Where $X = \{x \in \mathbb{R}^n \mid Ax \geq b, x \geq 0\}$, $c_i = (c_{i1}, ..., c_{in}) \in \mathbb{R}^n$, $i = 1,...,k$; $b = (b_1, ..., b_m) \in \mathbb{R}^m$ and A is a $\mathbb{R}^{m \times n}$ matrix.

For the sake of simplicity we suppose that all the objectives should be minimized. However, the procedure demonstrated is easily extendable were the case to include some maximizing objectives.

In (MOP) it is unlikely that all objectives will simultaneously achieve their optimal value. Therefore in practice the decision maker (DM) chooses a satisfying solution, according to the aspiration level fixed for each objective.

Assuming that the DM proposes imprecise aspiration levels such as, "$z_i(x)$ should be essentially less than or equal to some value g_i", model (MOP) can be written as

$$\text{(FGP)} \quad \text{Find } x$$
$$\text{Such that} \quad z_i(x) \lesssim g_i \quad i = 1,2,...,k$$
$$x \in \text{X}$$

Each expression $c_i x \lesssim g_i$ is represented by a fuzzy set called fuzzy goal, whose membership function, $\mu_i(z_i)$, $\mu_i : \mathbb{R} \to [0,1]$, provides the satisfaction degree λ_i to which the i^{th} fuzzy inequality is satisfied. In order to define the membership function $\mu_i(z_i)$ the DM has to provide the tolerance margins g_i+t_i that she/he is willing to accept. So $\mu_i(z_i)$ should be equal to 1 if $z_i \leq g_i$, strictly monotone decreasing from 1 to 0 over the interval $(g_i, g_i + t_i)$ and equal to 0 if $z_i \geq g_i + t_i$.

Models like this are named fuzzy goal programming (FGP) problems. (FGP) should be considered as an auxiliary model in order to solve (MOP).

As it has been widely seen, a FGP problem, using the fuzzy decision max-min of Bellman and Zadeh [1] and introducing the auxiliary variable λ, adopts the following formulation [2]

$$\text{(B-Z)} \quad \text{Max } \lambda$$
$$\text{S.t.} \quad 0 \leq \lambda \leq \mu_i(z_i(x)) \quad i = 1,\ldots,k$$
$$x \in X, \ \lambda \leq 1$$

But when there are several objective functions it can be a difficult task for the DM to determine coherent aspiration levels. Therefore it is possible that some tolerance thresholds will be inconsistent with each other or with the crisp constraints, some of them being unattainable (membership degree=0). In this case (B-Z) is infeasible. Chinneck [3] summarizes the state of the art in algorithms related to infeasibility in optimization. Most published research has been focused on the diagnosis of infeasibility. Little investigation has been made in infeasibility resolution and very limited literature can be found in infeasibility resolution using fuzzy approach. León and Liern [5] use a fuzzy sets approach to relax constraints to repair infeasibility in crisp mathematical programming problems. Gupta et al. [6] take a similar fuzzy approach to finding a best approximate solution to an infeasible mathematical programming problem. In this paper we propose a procedure in order to solve a FGP should the above described drawback occur. We demonstrate the

ISBN: 978-989-95079-6-8

operativeness of our approach by means of a numerical example.

2 Repairing infeasibility

Let us assume that (B-Z) is infeasible because the tolerance thresholds of some goals are unattainable. The first step (M.I) is to identify one minimum-cardinality subset of {1, 2,…, k}, that will be denoted *CS*, so that the goals {1, 2,…, k}−*CS* provide a feasible (B-Z) problem. This procedure is related to the minimum cardinality irreducible infeasible subsystem (IIS) set-covering problem (MIN IIS COVER) (see [3, 4]). Hence the Chinneck's algorithm [4] will be applied to the following sets of linear constraints

$$z_i(x) \le g_i + t_i \qquad i = 1,2,...,k$$
$$x \in \mathrm{X}$$

where only the k -first constraints (the goals) are elasticized. So we will work with the model

$$\text{(M.I)} \quad \text{Min} \sum_{i=1}^{k} p_i$$
$$\text{S.t.} \quad z_i - p_i \le g_i + t_i \quad i = 1,\ldots,k$$
$$x \in X$$

Next, in order to repair infeasibility, we have to shift the unattainable goals, i.e. the *CS*-goals, as low as possible without penalizing excessively the achievement degree of other goals. Taking this into account we consider the following two positions: 1) we allow the feasible goals to deteriorate to their corresponding tolerance threshold 2) we aim to attain a satisfactory solution for the feasible goals. Obviously the relaxation of the unattainable goals will be smaller in the first position than in the second. Therefore we assign a satisfaction degree equal to 1 to the relaxed values obtained in the first position and a satisfaction degree equal to 0 to that obtained in the second position, decreasing monotonously between these values. This way, we obtain the membership functions $\mu_i^*\left(z_i(x)\right)$ that rebuild the new fuzzy goals corresponding to the, previously, unattainable goals.

Assuming, for the sake of simplicity, that the unattainable goals are the first *r*. Consider the aforementioned position 1, that is to say, we allow the feasible goals to deteriorate until their tolerance threshold and, being an inclusive condition, we look for the lowest relaxation levels for the unattainable goals by conventional goal programming approach [7]. Thus we propose to solve the following model

$$\text{(M.II)} \quad \text{Min} \sum_{i=1}^{r} w_i p_i$$
$$\text{S.t.} \quad z_i(x) - p_i \le g_i + t_i, \quad i = 1,...,r$$
$$z_i(x) \le g_i + t_i, \qquad i = r+1,...,k$$
$$x \in X, p_i \ge 0$$

where $\sum_{i=1}^{r} w_i = 1$ and $w_i \ge 0$ for all i=1,…,r. The variables p_i have the same significance as the positive deviation in goal programming (GP). (M.II) searches for a solution with the smallest weighted average relaxation levels in order to achieve feasibility, considering that the other goals could reach their corresponding tolerance threshold. Weight w_i represents the priority or importance of the relaxation levels of tolerance threshold g_i+t_i. It is well known that when $w_i>0$ for all i=1,…,r, any optimal solution of (M.II) is efficient [8], which give us a large set of possible goal shifts that provide a feasible solution.

Let x^{II} be an optimal solution of (M.II). In line with what we have said before, the values $z_i^{\mathrm{II}} = z_i\left(x^{\mathrm{II}}\right), i = 1,...,r$, have membership degree equal to 1 in the new (repaired) fuzzy goals that we are constructing.

Now let us consider the aforementioned position 2, that is to say we are looking for a good degree of satisfaction of the goals that are attainable in (M.I). In order to achieve this we solve the following max-min model in which only the feasible goals are included

$$\text{(M.III)} \quad \text{Max } \lambda$$
$$\text{S. t. } z_i(x) \ge z_i^{\mathrm{II}}, \ i = 1,...,r$$
$$0 \le \lambda \le \mu_i\left(z_i(x)\right) \quad i = r+1,\ldots,k$$
$$x \in X \text{ , } \lambda \le 1$$

Let x^{III} be an optimal solution of (M.III) and $z_i^{\mathrm{III}} = z_i\left(x^{\mathrm{III}}\right)$, i=r+1,…,k. We can consider the values $\max\left\{z_i^{\mathrm{III}}, g_i\right\}$, i=r+1,…,k as sufficiently satisfactory. In the process of searching for the lowest relaxation levels of the infeasible goals, maintaining the values of the feasible goals sufficiently satisfactory, we solve the following GP model

$$\text{(M.IV)} \quad \text{Min} \sum_{i=1}^{r} w_i p_i$$
$$\text{S.t.} \quad z_i(x) - p_i \le g_i + t_i, \quad i = 1,...,r$$
$$z_i(x) \ge z_i^{\mathrm{II}}, \qquad i = 1,...,r$$
$$z_i(x) \le \max\left\{z_i^{\mathrm{III}}, g_i\right\}, \qquad i = r+1,...,k$$
$$x \in X, p_i \ge 0$$

Let x^{IV} be an optimal solution of (M.IV). The values $z_i^{\mathrm{IV}} = z_i\left(x^{\mathrm{IV}}\right), i = 1,...r$, achieved by the unattainable goals, have membership degree equal to 0 in the new (repaired) fuzzy goals that we are constructing.

We write $z_i^{\mathrm{II}} \equiv g_i^*$ and $z_i^{\mathrm{IV}} \equiv g_i^* + t_i^*$.

This way we have obtained the new fuzzy aspiration levels $\mu_i^*\left(z_i(x)\right)$ which restore feasibility with the lowest relaxation levels (see figure 1).

Finally we solve the following max-min model, by replacing the unattainable aspiration levels substituted with the new aspiration levels, which restores feasibility.

$$\text{(B-Z*)} \quad \text{Max } \lambda$$
$$\text{S.t.} \quad \mu_i^*\left(z_i(x)\right) \ge \lambda \ge 0, i = 1,...,r$$
$$\mu_i\left(z_i(x)\right) \ge \lambda \ge 0, i = r+1,...,k$$
$$x \in X, \quad \lambda \le 1$$

ISBN: 978-989-95079-6-8

The optimal solutions of (B-Z*) can be considered as solutions of (B-Z) well balanced between attainable fuzzy goals an unattainable fuzzy goals. Moreover, if the solution of this problem is not fuzzy-efficient of (FGP) and/or not Pareto optimal of (MOP), the solution can be improved following a procedure showed in [9].

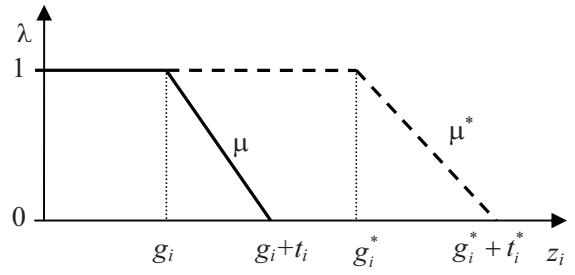


Figure1: Relaxation of an unattainable fuzzy aspiration level.

3 Example

To check the usefulness of our approach we consider the following numerical example.
Let a MOLP problem be

$$\begin{aligned}
&\text{Min } z_1=3x_1+3x_2+3x_3+2x_4+3x_5\\
&\text{Min } z_2=2x_1+x_2+2x_3+x_4+3x_5\\
&\text{Min } z_3=4x_1+4x_2+2x_3+3x_4+x_5\\
&\text{Max } z_4=x_1+6x_2+5x_3+7x_4+6x_5\\
&\text{Min } z_5= 5x_1+5x_2+4x_3+7x_4+5x_5\\
&\text{S.t. } 4x_1+2x_2+4x_3+2x_4+3x_5 \geq 18\\
&\quad x_1 \geq 1,\ 0 \leq x_3 \leq 3\\
&\quad x_2 \geq 0,\ x_4 \geq 0,\ x_5 \geq 0
\end{aligned}$$

Assume the DM proposes the following imprecise aspiration levels: "z_1 should be essentially less than or equal to 18", "z_2 should be essentially less than or equal to 6", "z_3 should be essentially less than or equal to 13", "z_4 should be essentially greater than or equal to 24" and "z_5 should be essentially less than or equal to 17", the respective tolerance threshold being 21, 8, 15, 21 and 20.

Step 1. We solve the (B-Z) problem

$$\begin{aligned}
&\text{Max } \lambda\\
&\text{S.t.}\\
&\lambda \leq \frac{1}{3}\left(21-\left(3x_1+3x_2+3x_3+2x_4+3x_5\right)\right)\\
&\lambda \leq \frac{1}{2}\left(8-\left(2x_1+x_2+2x_3+x_4+3x_5\right)\right)\\
&\lambda \leq \frac{1}{2}\left(15-\left(4x_1+4x_2+2x_3+3x_4+x_5\right)\right)\\
&\lambda \leq \frac{1}{3}\left(\left(x_1+6x_2+5x_3+7x_4+6x_5\right)-21\right)\\
&\lambda \leq \frac{1}{3}\left(20-\left(5x_1+5x_2+4x_3+7x_4+5x_5\right)\right)\\
&4x_1+2x_2+4x_3+2x_4+3x_5 \geq 18\\
&x_1 \geq 1,\ 0 \leq x_3 \leq 3\\
&x_2 \geq 0,\ x_4 \geq 0,\ x_5 \geq 0,\ 0 \leq \lambda \leq 1
\end{aligned}$$

This problem is infeasible because the tolerance thresholds of some goals are unattainable.

Step 2. We will identify one minimum-cardinality irreducible infeasible subsystem of goals by applying the Chinneck's algorithm to the following system of linear constraints where only the five first constraints (the goals) are elasticized:

$$\begin{aligned}
&3x_1+3x_2+3x_3+2x_4+3x_5 \leq 21\\
&2x_1+x_2+2x_3+x_4+3x_5 \leq 8\\
&4x_1+4x_2+2x_3+3x_4+x_5 \leq 15\\
&x_1+6x_2+5x_3+7x_4+6x_5 \geq 21\\
&5x_1+5x_2+4x_3+7x_4+5x_5 \leq 20\\
&4x_1+2x_2+4x_3+2x_4+3x_5 \geq 18\\
&x_1 \geq 1,\ 0 \leq x_3 \leq 3\\
&x_2 \geq 0,\ x_4 \geq 0,\ x_5 \geq 0
\end{aligned}$$

We obtain $CS=\{2, 5\}$.

Step 3. In order to repair infeasibility we propose to solve the following model (see M.II)

$$\begin{aligned}
&\text{Min } p_2 + p_5\\
&\text{S.t.}\\
&2x_1+x_2+2x_3+x_4+3x_5-p_2 \leq 8\\
&5x_1+5x_2+4x_3+7x_4+5x_5-p_5 \leq 20\\
&3x_1+3x_2+3x_3+2x_4+3x_5 \leq 21\\
&4x_1+4x_2+2x_3+3x_4+x_5 \leq 15\\
&x_1+6x_2+5x_3+7x_4+6x_5 \geq 21\\
&4x_1+2x_2+4x_3+2x_4+3x_5 \geq 18\\
&x_1 \geq 1,\ 0 \leq x_3 \leq 3\\
&x_2 \geq 0,\ x_4 \geq 0,\ x_5 \geq 0\\
&p_2 \geq 0,\ p_5 \geq 0
\end{aligned}$$

We get $z_2^{II}=9$ and $z_5^{II}=21.5$. These values have membership degree equal to 1 in the new (repaired) fuzzy goals that we are constructing.
In order to look for the balanced relaxation levels of the infeasible goal we need to find the values sufficiently satisfactory for the feasible goals.

Step 4. We solve the following model (M. III)

$$\begin{aligned}
&\text{Max } \lambda\\
&\text{S.t.}\\
&\lambda \leq \frac{1}{3}\left(21-\left(3x_1+3x_2+3x_3+2x_4+3x_5\right)\right)\\
&\lambda \leq \frac{1}{2}\left(15-\left(4x_1+4x_2+2x_3+3x_4+x_5\right)\right)\\
&\lambda \leq \frac{1}{3}\left(\left(x_1+6x_2+5x_3+7x_4+6x_5\right)-21\right)\\
&2x_1+x_2+2x_3+x_4+3x_5 \geq 9\\
&5x_1+5x_2+4x_3+7x_4+5x_5 \geq 21.5\\
&4x_1+2x_2+4x_3+2x_4+3x_5 \geq 18\\
&x_1 \geq 1,\ 0 \leq x_3 \leq 3\\
&x_2 \geq 0,\ x_4 \geq 0,\ x_5 \geq 0\\
&0 \leq \lambda \leq 1
\end{aligned}$$

As $z_i^{III}=15.48$, $z_3^{III}=11.32$, $z_4^{III}=26.52$ then in the process of searching for the lowest relaxation levels of the infeasible

ISBN: 978-989-95079-6-8

goals, maintaining the values of the feasible goals sufficiently satisfactory, we solve the following GP model:

Step 5. (see M.IV)

$$\begin{aligned}
&\text{Min } p_2 + p_5\\
&\text{S.t.}\\
&2x_1+x_2+2x_3+x_4+3x_5 - p_2 \le 8\\
&5x_1+5x_2+4x_3+7x_4+5x_5 - p_5 \le 20\\
&3x_1+3x_2+3x_3+2x_4+3x_5 \le 18\\
&2x_1+x_2+2x_3+x_4+3x_5 \ge 9\\
&4x_1+4x_2+2x_3+3x_4+x_5 \le 13\\
&x_1+6x_2+5x_3+7x_4+6x_5 \ge 24\\
&5x_1+5x_2+4x_3+7x_4+5x_5 \ge 21.5\\
&4x_1+2x_2+4x_3+2x_4+3x_5 \ge 18\\
&x_1 \ge 1,\ 0 \le x_3 \le 3\\
&x_2 \ge 0,\ x_4 \ge 0,\ x_5 \ge 0\\
&p_2 \ge 0,\ p_5 \ge 0
\end{aligned}$$

We get $z_2^{IV} = 9.5,\ z_5^{IV} = 24.96$. These values achieved by the unattainable goals, have membership degree equal to 0 in the new (repaired) fuzzy goals that we are constructing.

This way we have obtained the new fuzzy aspiration levels which restore the feasibility with the lowest relaxation levels: "z_2 should be essentially greater than 9" and "z_5 should be essentially greater than 21.5" with tolerance threshold equal to 9.5 and 25, respectively. Finally we solve the following (B-Z*) model

$$\begin{aligned}
&\text{Max } \lambda\\
&\text{s.t.}\\
&\lambda \le \frac{1}{3}\left(21-\left(3x_1+3x_2+3x_3+2x_4+3x_5\right)\right)\\
&\lambda \le \frac{1}{0.5}\left(9.5-\left(2x_1+x_2+2x_3+x_4+3x_5\right)\right)\\
&\lambda \le \frac{1}{2}\left(15-\left(4x_1+4x_2+2x_3+3x_4+x_5\right)\right)\\
&\lambda \le \frac{1}{3}\left(\left(x_1+6x_2+5x_3+7x_4+6x_5\right)-21\right)\\
&\lambda \le \frac{1}{3.5}\left(25-\left(5x_1+5x_2+4x_3+7x_4+5x_5\right)\right)\\
&4x_1+2x_2+4x_3+2x_4+3x_5 \ge 18\\
&x_1 \ge 1,\ 0 \le x_3 \le 3\\
&x_2 \ge 0,\ x_4 \ge 0,\ x_5 \ge 0\\
&0 \le \lambda \le 1
\end{aligned}$$

We get $x_1^* = 1,\ x_2^* = 0.79,\ x_3^* = 2.86,\ x_4^* = 0.25,\ x_5^* = 0.13$; $z_1\left(x^*\right) = 14.92,\ z_2\left(x^*\right) = 9.2,\ z_3\left(x^*\right) = 13.83,\ z_4\left(x^*\right) = 22.75$ and $z_5\left(x^*\right) = 22.95$. This can be considered as a solution of model (B-Z) well balanced between attainable fuzzy goals and unattainable fuzzy goals.

Acknowledgment

This work was supported by the *Ministerio Español de Educación y Ciencia* under grant no. MTM2007-67634. and by the *Universidad del País Vasco* under grant no. EHU08/28.

4 Conclusions

This paper describes a method to deal with infeasibility in fuzzy goal programming problems in case of the DM proposes unattainable tolerance thresholds for some goals. Considering the inherent flexibility of fuzzy approach we restore the feasibility by shifting the unattainable goals as low as possible but, at the same time, taking care not to penalize excessively the degree of achievement of the others goals. This way we rebuilt the unattainable goals obtaining a new feasible fuzzy goal programming problem whose solution is well balanced between attainable and unattainable fuzzy goals.

References

[1] R.E. Bellman and L.A. Zadeh, Decision making in a fuzzy environment, *Management Science,* 17: 141-164, 1970.

[2] H.J. Zimmermann, Fuzzy programming and linear programming with several objective functions, *Fuzzy Sets and Systems*, 1 (1): 45-55, 1978.

[3] J.W. Chinnek, *Feasibility and Infeasibility in Optimization. Algorithms and Computational Methods.* New York: Springer, 2008.

[4] J. W. Chinnek, Fast Heuristics for the Maximum Feasible subsystem Problem, *INFORMS Journal on Computing,* Vol. 13, Nº 3, Suumer: 210-223, 2001.

[5] T. León and V. Liern, A fuzzy method to repair infeasibility in linearly constrained problem, *Fuzzy Sets and Systems,* 122: 237-243, 2001.

[6] P. Gupta, M. Vlach and D. Bhatia, Fuzzy approximation to an infesible generalized linear complementary problem, *Fuzzy Sets and Systems,* 146: 221-233, 2004.

[7] J. Jang, Infeasibility resolution based on goal programming, Computers & Operations Research, 35:1483-1493, 2008.

[8] Y. Sagarawi, H. Nakayama, T. Tanino. Theory of Multiobjective Optimization. London: Academia Press, 1985.

[9] M. Jiménez and A. Bilbao, Pareto-optimal solutions in fuzzy multi-objective linear programming, *Fuzzy Set and Systems*, 2008, doi: 10.1016/j.fss.2008.12.005.

Sorted kernel matrices as cluster validity indexes

Francisco A. A. Queiroz[1] Antonio P. Braga[1] Witold Pedrycz[2]

1 Department of Electronics Engineering, Federal University of Minas Gerais
Belo Horizonte, Brazil
2 University of Alberta, Department of Electrical and Computing Engineering
Edmonton, Canada
Email: faaq@ufmg.br, apbraga@ufmg.br, pedrycz@ece.ualberta.ca

Abstract—*Two basic issues for data analysis and kernel-machines design are approached in this paper: determining the number of partitions of a clustering task and the parameters of kernels. A distance metric is presented to determine the similarity between kernels and FCM proximity matrices. It is shown that this measure is maximized, as a function of kernel and FCM parameters, when there is coherence with embedded structural information. We show that the alignment function can be maximized according FCM and kernel parameters. The results presented shed some light on the general problem of setting up the number of partitions in a clustering task and in the proper setting of kernel parameters according to structural information.*

Keywords— Affinity matrix, Clustering, Fuzzy C-Means (FCM), Kernel matrix, Reordering, Sorting.

1 Introduction

The importance of kernels has been much emphasized in the literature as the basic construct of learning machines like SVMs [6]. Kernel parameters play, therefore, a major role in achieving good performance of these classifiers. Setting up kernel parameters is usually accomplished by an exhaustive search carried out in the space of parameters. Quite often this search is completed without a full understanding of how kernels represent relations in the input space. The proper set-up of kernels is only expected to capture the input to feature space mapping that results on an acceptable overall machine performance, without any direct constraints imposed on the internal representation of relationships between patterns.

However, like Fuzzy C-Means (FCM) [7] proximity matrices (FPM) [8], kernels may incorporate important structural information about the data. In fact, data clustering can be accomplished by using information already contained in kernels, by finding the proper order of columns and rows of the kernel matrix [2, 3, 4, 5]. A similar affinity matrix [1] representation can also be obtained for FPMs, which may suggest that kernels and FPMs embody similar information about the data if their parameters are set accordingly. This observation also suggests that supervised and unsupervised learning may exhibit a closer relationship by the exchange of information between kernels and FPMs.

The problem of approximating kernels and FPMs relies, therefore, on setting their parameters so that they yield matrices that are close to each other or, in other words, that represent the same structural information. The parameters to be set are basically the width *r* of the Gaussian functions, for RBF kernels, and the number of partitions c, the basic parameter of FPMs. In this paper, we investigate how these two parameters are related to the alignment between RBF kernels and FPMs.

Experimental results on synthetic data yielded maximum alignment on values of *c* that coincide with the number of clusters of the data generator functions. This suggests that, for a given kernel matrix, the proper number of clusters can be induced by maximizing the similarity of FPM and the kernel matrix. In addition, it is also shown that, for the same problems, SVMs [6] can also be designed by setting the kernel parameter to the corresponding value of *r* that maximizes the similarity.

The paper is organized as follows. We start with a general overview of the main concepts including kernels and FPM. The concept of affinity matrix representation of kernels and FPMs is then described. Another important link between these two representations of data is presented in section 5 where kernels are shown to incorporate the same structural information the FPMs. The metric for similarity measure is then presented, which is followed by the description of the optimization algorithm, the main results, discussions and conclusions.

2 Mercer Kernels

The main characteristic of (Mercer) kernels [9] in the supervised learning context is that they allow implicit mapping of input data into feature space without actually computing the mapping itself. Instead, the mapping is accomplished by computing the internal product located in the input space. The kernel matrix $\mathbf{K}=[k(x_i,x_j]$ is regarded as a Mercer kernel if it is symmetric and positive semi-definite. In general, the kernel matrix can be considered as a matrix capturing similarity between all pairs of points of a data set. Kernels can be implemented by polynomials, linear and sigmoidal functions as well as Gaussian functions. A Gaussian, or RBF, kernel is described as follows

$$k(x_i, x_j) = e^{\frac{-(x_i - x_j)^2}{2r^2}} \quad (1)$$

where *r* is the radius of the Gaussian function.

ISBN: 978-989-95079-6-8

As an example, let us consider Fig. 1, where data were sampled from two Gaussian distributions. The corresponding RBF kernel where *r=0.25* is shown in Fig. 2. As patterns are not ordered according to the generator functions, no structural information can be directly visualized in the kernel matrix. In order to observe pattern relation properties directly from the kernel matrix, it is necessary to order the patterns accordingly, as it will be described in the next sections.

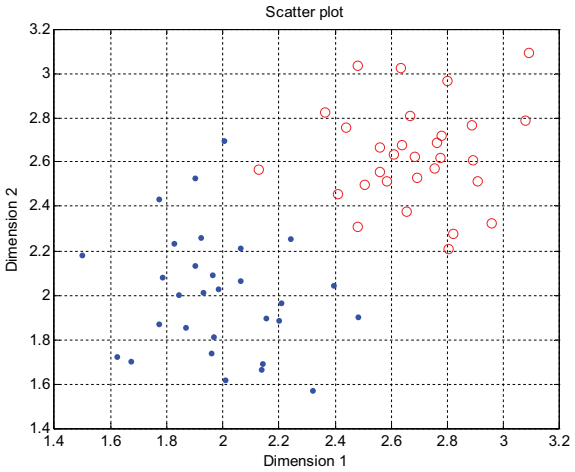


Figure 1: Data sample from two Gaussian distributions.

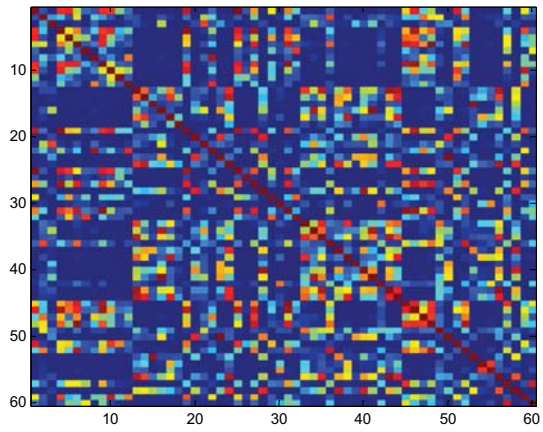


Figure 2: Unordered RBF kernel matrix (r=0.25) for the data of Fig. 1.

3 FCM Proximity Matrix

Fuzzy clustering realized by the well-known FCM algorithm [7] is formulated as a constrained optimization problem, where the partition matrix $\mathbf{U}=[u_{ik}]$ satisfies the constraint $\sum_{i=1}^{c} u_{ik} = 1$, while the objective function to be minimized reads as follows

$$J = \sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^{m} \| \mathbf{x}_k - \mathbf{v}_i \|^2 \qquad (2)$$

where c is the number of partitions (clusters) in the data set, N is the data set size, m is the fuzziness coefficient (fuzziness factor) and $\mathbf{v}_i$ is the prototype of cluster i.

The obtained partition matrix $\mathbf{U}=[u_{ij}]$ can be used to generate the *NxN* proximity matrix $\mathbf{P}=[p_{kl}]$, according to the known relationship available in the literature [8] that is presented in (3). Likewise kernels, the proximity matrix **P** embodies relationships between patterns according to the clusters represented in the partition matrix **U**.

$$p_{kl} = \sum_{i=1}^{c} \min(u_{ik}, u_{il}) \qquad (3)$$

4 Affinity matrices

Given a data set $\Gamma_u = \{x_k\}_{k=1}^{N}$, where N is the number of patterns, the elements s_{ij} of the Affinity Matrix $\mathbf{S}=[s_{ij}]$ contain a measurement or estimation of the affinity of the pair of patterns $(\mathbf{x}_i, \mathbf{x}_j)$, where affinity is defined as a likeness based on relationship or causal connection [1]. For reflexive affinities, **S** is symmetrical, what implies on $s_{ij}=s_{ji}$. In general, the Affinity Matrix can be represented as a block diagonal symmetrical matrix as the one of (4).

$$S = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1k} \\ S_{21} & S_{22} & \cdots & S_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ S_{k1} & S_{k2} & \cdots & S_{kk} \end{bmatrix} \qquad (4)$$

where $\boldsymbol{S}_{ij}=\boldsymbol{S}_{ji}$ are sub-matrices of **S** and k is the total number of sub-groups of **S**.

For ordered data sets, kernels and proximity matrices can be represented in the form of (4), where sub-matrix $\mathbf{S}_{ii}$ represents within-cluster affinities and $\mathbf{S}_{ij}$ represents inter-cluster affinities, for $i \neq j$. Consider, for instance, the data set of Fig. 1 and the corresponding kernel presented in Fig. 2, which is now ordered according to the generator distributions and presented in Fig. 3. The block-diagonal affinity matrix form of the kernel presented in Fig. 3 shows clearly the structure of the data set. The block-diagonal matrices $\mathbf{S}_{11}$ and $\mathbf{S}_{22}$ represent the two clusters of Fig. 1 with 30 elements each, whereas the matrices $\mathbf{S}_{21}$ and $\mathbf{S}_{12}$ represent the relationships between data of matrices $\mathbf{S}_{11}$ and $\mathbf{S}_{22}$. Matrices $\mathbf{S}_{21}$ and $\mathbf{S}_{12}$ have in fact very small values, indicating that the two clusters are not quite related to each other, as can be confirmed in Fig. 1. A similar result to the one presented in Fig. 3 would have been obtained if the proximity matrix **P** were obtained and ordered according to the partition matrix **U** for c=2.

ISBN: 978-989-95079-6-8

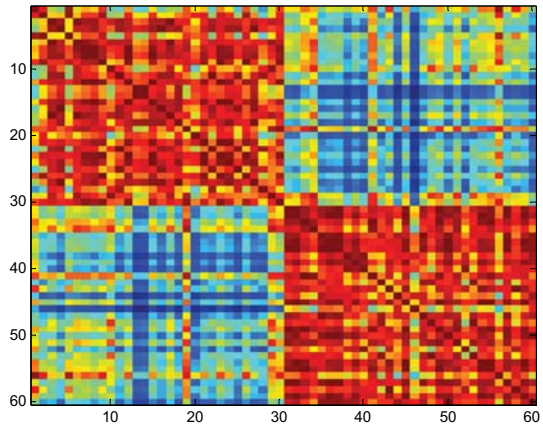


Figure 3: Ordered RBF kernel matrix ($r=0.25$) for the data set of Fig. 1.

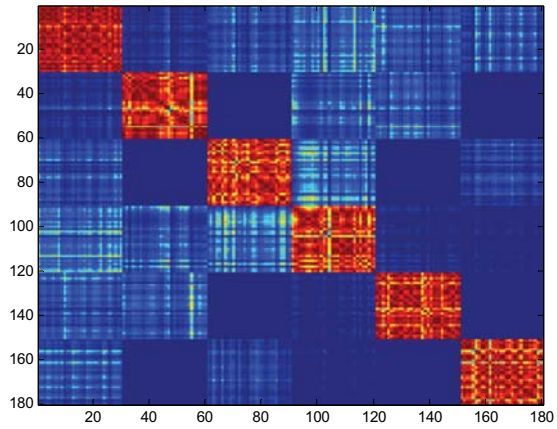


Figure 5: Affinity matrix representation of the ordered RBF kernel matrix of the data set of Fig. 4.

5 Clustering by Sorting Kernels

In the previous section we have pointed out to the close relationship between kernels and proximity matrices as represented by affinities of the data. Clustering algorithms can be described with the objective of sorting kernel matrices. There are in fact clustering algorithms in the literature that work by sorting rows and columns of the kernel matrix [2, 3, 4, 5] in order to obtain affinity matrices like the one presented in Fig. 3. Once the affinity matrix has been formed, embedded cluster information can be directly extracted from the kernels.

The eigenvector ordering method [1, 2] was applied to the data set presented in Fig. 4, which was generated from 6 Gaussian generator functions. The corresponding ordered RBF kernel is presented in Fig. 5. As it can be observed, the ordered kernel shows clearly the presence of 6 groups of data in the input space. The affinity matrix representation of kernels simply reveal the information that is already present in the kernel, since only permutation operations were accomplished in the rows and columns of the original matrix. This indicates that (properly set) kernels already contain information about data structure that is equivalent to that obtained by clustering methods.

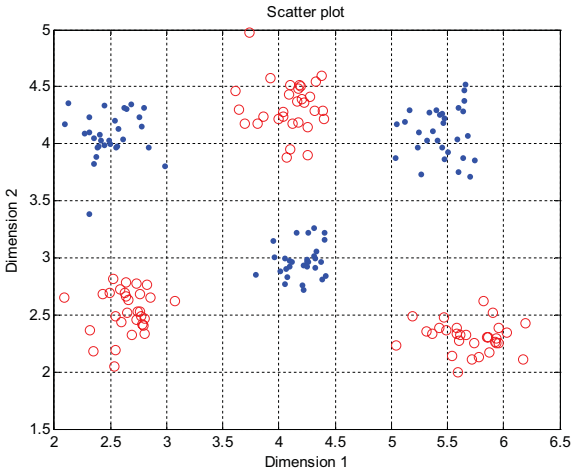


Figure 4: Data set obtained from 6 Gaussian functions.

6 Alignment of Kernels and FPMs

In order to quantify the similarity between kernels and proximity matrices, the Empirical Alignment described in [2] was adopted. The alignment quantity A is described in the form

$$A(K,P) = \frac{\langle K,P\rangle_F}{\sqrt{\langle K,K\rangle_F \langle P,P\rangle_F}} \tag{3}$$

where K and P, respectively, are the kernel and proximity matrices and $\langle .,. \rangle_F$ is the Frobenius inner product [2],

$$\langle K,P\rangle_F = \sum_{i=1}^{N}\sum_{j=1}^{N} K(i,j)P(i,j) \tag{4}$$

7 Alignment as an Optimization Problem

As discussed in the previous sections, depending on their parameters, kernels and FPM may incorporate analogous information about patterns and cluster relationships which hold in the input space. For the RBF kernels, structural information may be revealed by the proper setting of the width r, whereas FPM ability to describe pattern relations depends on the previously set number of clusters c. Nevertheless, these parameters are usually blindly configured in advance by the user. In the SVM design, the value of r can be fine-tuned according to the overall learning machine performance, while further cluster analysis can also give a hint for fine tuning of the value of c. When the setting of (c, r) results on affinity matrices that are coherent with the actual data structure, the corresponding kernel and FPM will be aligned to each other, according to the metric presented in (3). These arguments suggest that A(K,P) has a maximum on (c^*, r^*), where c^* and r^* are, respectively, the number of partitions and RBF width that best describe the original data structure. This characterizes the optimization problem presented in (5). The objective is therefore to obtain the values of the parameters c and r that maximize the alignment, which are expected to be the ones that describe better the data set for both kernel and proximity matrix.

$$\arg \max_{(c,r)} A(K,P) \tag{5}$$

ISBN: 978-989-95079-6-8

In order to demonstrate the arguments above, the values of A(K,P) were computed as a function of *c* and *r* for the data sets coming from Fig. 1 and 4, respectively. The obtained results are shown in Fig. 6 and 7. The same calculations were also completed for another data set with 4 original Gaussian generator functions. The obtained graph is presented in Fig. 8. As can be observed in Fig. 6, 7 and 8, the maximum occurs for *c=2*, *c=6* and *c=4*, respectively, which correspond to the original number of generator functions used in each data set. These results confirm the principle that the alignment function (3) has a maximum corresponding to the number of generator functions and that (5) could be optimized in order to obtain the values of *c* and *r* that result on the best representation of the data set. In the next section, the optimization of (5) will be implemented with a simple Genetic Algorithm (GA) [10] approach.

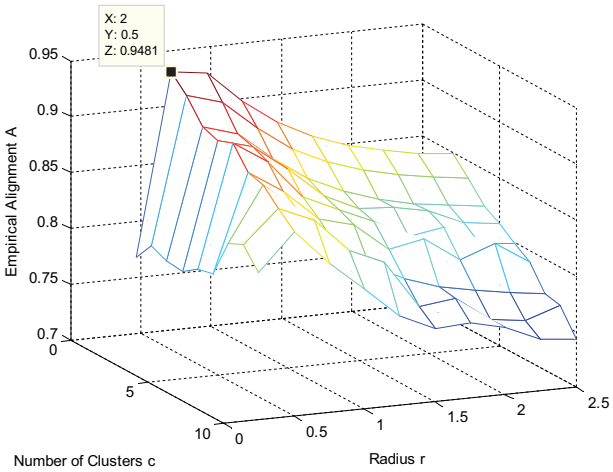


Figure 6: Alignment A(K,P) as a function of c and r for the data set of Fig. 1.

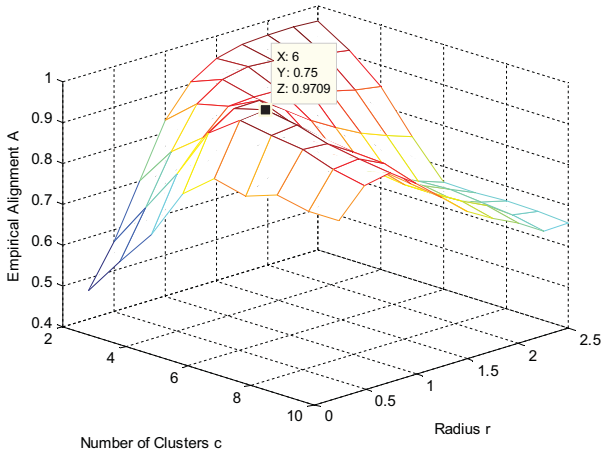


Figure 7: Alignment A(K,P) as a function of c and r for the data set of Fig. 4.

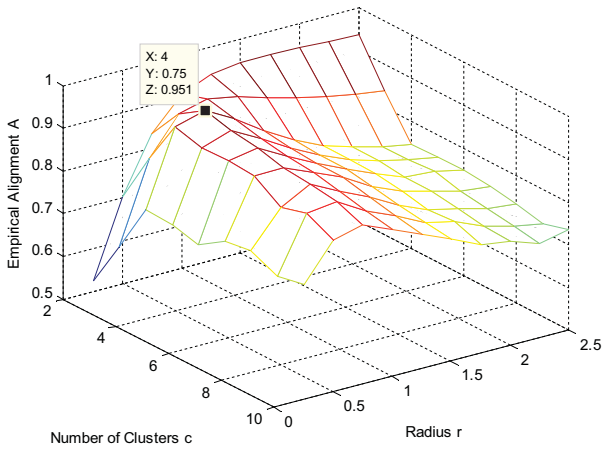


Figure 8: Alignment A(K,P) as a function of c and r for a data set with 4 Gaussian generator functions.

8 Evolutionary Optimization

A simple Genetic Algorithm (SGA) [10] was implemented in order to optimize (5) for the three problems represented by the objective functions of Fig. 6, 7, and 8. GA parameters are presented in Table 1 and the obtained results in Table 2. As can be observed, the maximum obtained for the three objective functions are quite close to the ones observed in the figures. As for the number of clusters, the optimization yielded exactly the same number of generator functions.

Table 1: Parameters of the SGA for optimization of the objective functions of Fig. 6, 7 and 8.

Number of variables in search space	2
Population size	50
Maximum number of generations	65
Crossover rate	0.6
Mutation rate	0.03
Biased linear cross-over coefficient	0.9
Linear cross-over coefficient	0.5

Table 2: SGA solutions for the objective functions of Fig. 6, 7 and 8.

Number of generator functions	***Number of iterations***	***r***	***c***	**A**
Two	52	0.5894	2	0.9540
Four	19	0.6808	4	0.9537
Six	63	0.6551	6	0.9758

The SGA was also applied to the data set of Fig. 9, represented as a binary classification problem. In order to distinguish between structural information and data set labels for classification problems, one of the classes was sampled from two Gaussian distributions, as can be observed in Fig. 9. In such a situation, there are three generator functions and two classes. This can be visualized by the ordered kernel of Fig. 10, which suggests the existence of the three clusters, in spite of the number of labels. The SGA optimization resulted on *r=0.66959*, *c=3* and *A=0.92225*,

ISBN: 978-989-95079-6-8

what is consistent with the number of clusters of the original distribution.

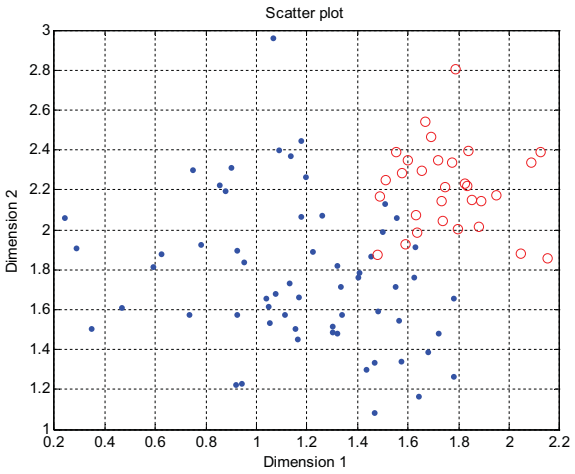


Figure 9: Binary classification problem sampled from three generator functions.

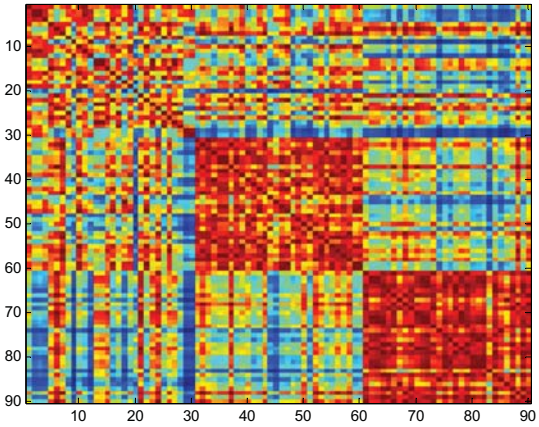


Figure 10: Affinity matrix representation of the ordered RBF kernel matrix of the data set of Fig. 9.

In order to show the consistency of the RBF width resulted from SGA optimization for classification purposes, a SVM was designed with the obtained value of $r=0.66959$. The corresponding separating surface is presented in Fig. 11 and, as can be observed, the resulting surface is coherent with the classification problem (the margin parameter was set to $C=2$). This suggests that the maximum obtained from (5) also points out to a proper tuning of kernel parameters for SVM design. Similar procedures were applied to the Iris data set [12], which resulted on $c=2$ in the maximum alignment. This result is consistent with the literature and with statistical analysis of the data set [13].

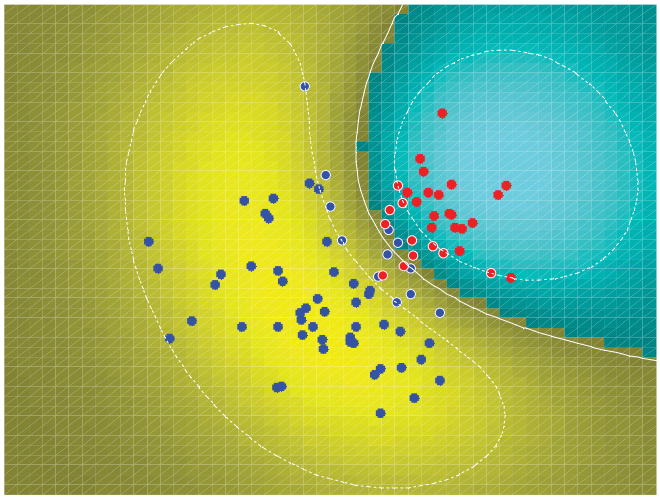


Figure 11: SVM separation surface for Fig. 9. RBF kernel was selected with and $r=0.66959$ obtained from SGA optimization.

9 Discussions and conclusions

Kernel and clustering design was presented in this paper in the perspective of function optimization. Our main argument to support this principle is that when both kernel and FPM are coherent with the embedded structural information contained in the data set, they should be aligned to each other. Alignment is measured in this paper according to the Frobenius inner product [2] between kernel and FPM, computed as a function of the parameters c and r. Therefore, we argue that maximum alignment should result on a proper setting of the parameters c and r. This approach differs from other kernel FCM algorithms [11], since the original objective functions are used.

This principle sheds some light on the discussion of determining the number of partitions of a clustering task. It has been shown throughout the paper that coherent partitions can be inferred according to the maximum alignment principle. Likewise, proper kernel design can also be obtained from the alignment. This suggests that kernel can be designed by exploring the structural data properties instead of performing an exhaustive blind search in the space of parameters with the aim of solely accomplishing the input-output mapping. According to these arguments, kernels should be more representative of the underlying data instead of serving uniquely as a mapping engine. This also suggests that kernels and FCM clustering should be co-designed and not described independently. According to the same principle, supervised and unsupervised learning may exhibit a closer relationship by collaborative interaction between kernels and FPMs.

References

[1] Y. Weiss, Segmentation using eigenvectors: a unifying view, *Proceedings of the International Conference on Computer Vision*, Volume 2, p. 975, May 20-25, 1999.

[2] N. Cristianini, J. Kandola, A. Elisseeff and J. Shaw-Taylor, On kernel target alignment, *Journal of Machine Learning Research*, 1, 2002.

[3] L. Võhandu, R. Kuusik, A. Torim, E. Aab, G. Lind, Some algorithms for data table (re) ordering using Monotone Systems, *Proceedings of the 5th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases: 5th WSEAS International*

ISBN: 978-989-95079-6-8

Conference on Artificial Intelligence, Knowledge Engineering and Data Bases (AIKED'06) , 417-422, Madrid, Spain, Feb 15-17, 2006.

[4] D. Tsafrir, I. Tsafrir, L. Ein-Dor, O. Zuk, D. A. Notterman and E. Domany, Sorting points into neighborhoods (SPIN): data analysis and visualization by ordering distance matrices. *Bioinformatics*, 21:2301-8, 2005.

[5] P. Arabie, L. J. Hubert, The bond energy algorithm revisited, *IEEE Transactions on Systems, Man and Cybernetics*, 20 (1): 268-274, Jan-Feb 1990.

[6] C. Cortes, V. Vapnik, Support Vector Network, *Machine Learning*, vol.20, pp.273-297.

[7] J. C. Dunn, A Fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, *Journal of Cybernetics* 3: 32-57. 1973.

[8] V. Loia, W. Pedrycz and S. Senatore, Proximity fuzzy clustering for web context analysis, *Proceedings of EUSFLAT 2003*, pp. 59-62.

[9] N. Cristianini and J. Shawe-Taylor, Support vector machines and other kernel-based learning methods. Cambridge University Press, 2000.

[10] D. E. Goldberg, Genetic algorithms in search, optimization and machine learning. Kluwer Academic Publishers, Boston, MA, 1989.

[11] W. Cai, S. Chen and D. Zhang, (2007). Robust fuzzy relational classifier incorporating the soft class labels**,** *Pattern Recognition Letters*, vol. 28(16), pp 2250-2263.

[12] A. Asuncion, 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.

[13] S. R. Gunn, Support vector machines for classification and regression (Technical Report). Image Speech and Intelligent Systems Research Group, University of Southampton, 1997.

ISBN: 978-989-95079-6-8

Complementary source information cooperation within a decision system for crop monitoring

Mahmoud El Hajj[1] Serge Guillaume[2] Agnès Bégué[3]

1 NOVELTIS, Ramonville Saint-Agne, F-31520 France
2 Cemagref, UMR ITAP, Montpellier, F-34196 France
3 CIRAD, UMR TETIS, Montpellier, F-34093 France
Email: mahmoud.elhajj@noveltis.fr, serge.guillaume@montpellier.cemagref.fr, agnes.begue@teledetection.fr

Abstract— *Multi-source information fusion usually deals with redundant sources; the goal is to aggregate decisions, or evidence which supports one or another class. Many frameworks are then available: OWA, fuzzy integrals, etc. In this paper we are interested in managing complementary information sources. The integration is done using a fuzzy rule base. The rules are generated using a fuzzy decision tree algorithm. The application consists in the design of a crop harvest monitoring system. The three information sources are a time series of satellite images, a crop growth model and expert knowledge. The results show that the system is ready to use in an operational way.*

Keywords— Crop growth modeling, Decision, Expert knowledge, Fuzzy rules, Learning, Remote sensing.

1 Introduction

Over the past decade, time series of satellite images acquired at high spatial resolution have proven to be an important source of information for different agricultural applications. Several authors have recognized the benefits of this kind of data for monitoring agricultural lands [1], classifying land cover [2-4], mapping seasonal patterns and crop rotations [5, 6], and for many other uses (see papers collected in [7, 8]).

Nevertheless, the quantity of information extracted from time series of optical images is often restricted by several factors: acquisition gaps, atmospheric conditions, imperfect radiometric normalization, radiometric confusion, etc. Therefore, to make credible decisions, this information needs to be supplemented with data from other sources.

Let us focus on the sugarcane industry. One of its principal needs is to have continuous information about the harvest progress throughout the season. Such information helps increasing the effectiveness in fields and factory. The main method reported in the literature, to monitor the sugarcane harvest is based on a multispectral classification [9-11]. The major limitations of this method are the subjectivity and the considerable time to invest in the photo-interpretation phase and the confusion between the various states of a sugarcane field when the gap between two released images is important (more than 2 months). Since the harvest campaign extends over several months, it is difficult in some cases to distinguish between a standing crop and the regrowth in a field harvested at the beginning of the season.

In order to improve and automate the monitoring of sugarcane harvest using time series of high spatial resolution satellite images, it is necessary to develop a new method that process the time series in co-operation with other sources. This co-operation allows compensating for the possible lack of data in the time series of images by the integration of temporal and environmental constraints.

As the sources are complementary and of different nature, usual frameworks of fusion information cannot be used. In this paper we propose to make the source integration within a fuzzy inference system. Next section introduces the study site and the data used. The characteristics of the three information sources are presented in section 3. Section 4 describes the fuzzy inference system and section 5 goes into detail of the experiments. The last section deals with concluding remarks.

2 Study site and data sets

The study site consists of two sugarcane farms located in the north-east part of Reunion Island, which is a small territory of ~2,500 km² in the Indian Ocean (21°7' to 19°40' S, 55°13' to 61°13' E), where sugarcane is the main crop. The first farm is at an average altitude of 70 m and includes 33 fields with an average size of 5.4 ha. The second farm is at an altitude ranging from 400 to 700 m, and has 46 fields with an average size of 3.5 ha. As the study area is located in a tropical zone, the year is divided into two seasons: a hot rainy season from November to April, and a cool dry season from May to October.

The satellite data set used in this study consists of 18 SPOT5 images acquired over Reunion Island between January 10th, 2003 and December 7th, 2004. Both SPOT5 instruments (HRG1 and HRG2) acquire radiation in four spectral bands[1] with high spatial resolution: 10 m for the Green, Red, and Near Infra-Red (NIR) bands, and 20 m for the Short Wave Infra-Red (SWIR) band. The images belong to the Kalideos-ISLE REUNION database

[1] Green (0.50 - 0.59 µm), Red (0.61 - 0.68 µm), Near Infra-Red (0.78 - 0.89 µm), Short Wave Infra-Red (1.58 - 1.75 µm).

ISBN: 978-989-95079-6-8

set up by the CNES[2] [12, 13]. All images were ortho-rectified and co-registered to the UTM coordinate system (zone 40 South) with a root mean square error of less than 0.5 pixel per image.

The radiometry of the images was corrected so that pixel values represent the top of canopy reflectances in the four spectral bands [14, 15]. Cloud mask was available for each image.

Block field boundaries for all Reunion Island were provided by the DDAF[3] and were refined by the CIRAD[4] to define the boundaries of each field in the study site.

Daily climatic data recorded at La Mare meteorological station near the two sugarcane farms were collected for the period covered by the satellite time series. These data are daily estimations of rainfall (mm), potential evapo-transpiration (mm), global radiation (J/m^2), and minimum, maximum and mean temperature values (°C). Climatic data were required to run the crop growth model.

A ground truth database was built by using harvest dates reported by farmers for each field during the 2003 and 2004 harvest campaigns. This database indicates the status of each field (whether it was harvested or not) between each pair of consecutive satellite acquisition dates in the time series.

3 Three information sources

3.1 SPOT5 time series

The time series of SPOT5 images are the principal source of information. By using field boundaries, the temporal profiles of reflectances in the four spectral bands, Green, Red, NIR and SWIR, were extracted from the images for each field in the study site. This extraction was carried out by calculating the average value of field pixels in each spectral band. The calculation was done after discarding cloud pixels using the cloud masks. Temporal profiles of NDVI (Normalized Difference Vegetation Index) where then calculated for each field using reflectance values in the Red, ρ_{Red}, and NIR, ρ_{NIR}, bands:

$$NDVI = \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}} \quad (1)$$

The temporal profile of NDVI provides useful information about the actual field status, and about its different historical stages. In general, this profile can be divided into two periods: a period in which NDVI values increase, corresponding to the vegetative development of the field crop, and another period with steady or decreasing values, corresponding to the maturation phase. Fig.1 shows an example of the temporal profile of a sugarcane field NDVI, extracted from a time series of SPOT5 images acquired with high temporal repeatability.

3.2 Sugarcane growth model

To cope with missing data or cloudy images, the sugarcane growth model MOSICAS [16] was used. It provides information on the field harvest possibility that is independent of the satellite images and based essentially on climatic data. This information is particularly interesting when the gap between two cloud-free satellite images is more than two months.

MOSICAS is a dynamic model that estimates the growth of the sugarcane on a daily time scale by using climatic and biophysical data of the environment of the sugarcane field. It allows the simulation of LAI (Leaf Area Index) profiles (Fig.1) that can be converted into NDVI profiles using a linear model [17]. An illustration of such a simulated profile is plotted in Fig.1. From the simulated NDVI temporal profiles, we build a helpful indicator for harvest detection: this indicator represents the nominal time required (in days) to reach a given threshold of NDVI starting with a given harvest date.

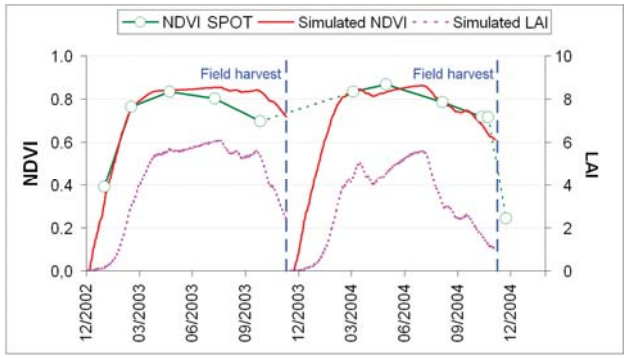


Figure 1: Example of temporal profiles of a sugarcane field NDVI extracted from SPOT5 time series, as well as NDVI and LAI temporal profiles simulated using the sugarcane growth model MOSICAS.

3.3 Expert Knowledge

The third source of information is the knowledge about phenological stages of the sugarcane, as well as its cropping system. This knowledge allows a better understanding of the relationship between the dynamics of the signals extracted from the satellite image time series and the sugarcane field status; it offers also very useful temporal constraints that particularly help in making decisions when there is a lack of satellite data. Examples of parameters defining these temporal constraints are the nominal length of the sugarcane cycle and the dates of opening and closing of the sugar factories which determine the harvest campaigns, as illustrated in Fig.2.

Various variables are built from the three information sources and used as inputs of the decision system. The inputs are divided into three main groups:

- The first group (4 inputs) is based on the NDVI and SWIR values calculated at field scale using each satellite image.
- The second group (5 inputs) relates to the dynamics in the temporal profiles of NDVI and SWIR.
- The third group (5 inputs) represents the temporal constraints resulting from the sugarcane growth model and the expert knowledge.

[2] *Centre National d'Etudes Spatiales* ; French Spatial Agency.

[3] *Direction Départementale de l'Agriculture et de la Forêt* ; Departmental Directorate of Agriculture and Forestry.

[4] *Centre de coopération Internationale en Recherche Agronomique pour le Développement* ; French Agricultural Research Centre for International Development.

ISBN: 978-989-95079-6-8

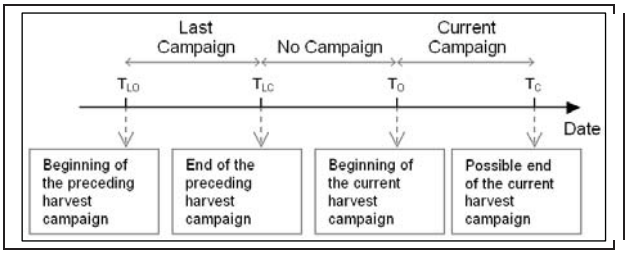


Figure 2: Temporal intervals used for the classification of image acquisition dates. T_{LO}, T_{LC}, T_O and T_C are opening and closure dates of the preceding and the current campaign respectively.

A thorough description of the inputs can be found in [18].

4 Decision system

The information coming from the three sources (SPOT5 time series, sugarcane growth model and expert knowledge) is heterogeneous and of different nature. The information extracted from the SPOT5 images is based on reflectance "measurements". The sugarcane model provides "estimates" using parameters that are not directly connected to the studied phenomenon (*i.e.* the field harvest). Expert knowledge contribution is neither "measurements" nor "estimates". Another important characteristic to highlight is that the sources are not redundant but complementary. That means that none of them is able to make a decision on its own; the cooperation is needed within the decision process. This discards usual aggregation techniques such as votes, OWA or fuzzy integrals. The general framework used to combine the different and complementary information sources is a fuzzy inference system.

The rules are of the following form: "*if X Then Y*", e.g.

$$\begin{aligned}
&\text{Rule}_1\text{: if } x_1 \text{ is } A_1^1 \text{ and } x_2 \text{ is } A_2^1 \ldots \text{ and } x_n \text{ is } A_n^1 \text{ then } y \text{ is } B^1\\
&\text{Rule}_2\text{: if } x_1 \text{ is } A_1^2 \text{ and } x_2 \text{ is } A_2^2 \ldots \text{ and } x_n \text{ is } A_n^2 \text{ then } y \text{ is } B^2\\
&\quad\vdots\\
&\text{Rule}_r\text{: if } x_1 \text{ is } A_1^r \text{ and } x_2 \text{ is } A_2^r \ldots \text{ and } x_n \text{ is } A_n^r \text{ then } y \text{ is } B^r
\end{aligned} \quad (2)$$

where A_k^r is the fuzzy set of the k^{th} input variable involved in the rule r, and B^r is the conclusion of the r^{th} rule.

The rule conclusion is a symbolic label, related to the field status. Two are possible: H (harvested) or NH (not harvested).

For each input universe a partition is defined. Some of the input partitions are crisps (see Fig.2), the other are fuzzy. In this case, we built standardized partitions according to expert knowledge. Fig.3 illustrates the fuzzy sets of NDVI and an example of NDVI profiles plotted according to thermal time for several sugarcane fields. NDVI fuzzy sets were designed according to expert knowledge about the phenology and field status of sugarcane as well as about its NDVI temporal profiles.
Linguistic labels are related to the NDVI:

- "Low" NDVI values (< 0.30) generally correspond to residues and bare soil after field harvesting.
- "Medium" NDVI values (between 0.30 and 0.75) are observed in the growth and senescence phases. The growth phase is characterized by an accumulation of biomass and an increasing length of the stems. The senescence is caused by the sugarcane ageing, the reduction in the temperature and the lack of water.
- "High" NDVI values (>0.75) are obtained at the end of the growth stage and before senescence. In this phase sugarcane fields are well developed and green covered.

The inference technique used in our system is based on Mamdani's method [19]. The weight attributed to the conclusion of an activated rule is calculated by combining the membership degrees of rule premises in a conjunctive way using the *min* operator. The aggregation of the distinct conclusions of the activated rules is done in a disjunctive way using the *max* operator.
Once system inputs and output as well as inference parameters are defined, the last step is reasoning rule description. In a previous version [18], the rules were defined according to expert knowledge. 116 rules were needed to cover all possible situations. The system proved to be accurate. A comparison between fuzzy and crisp partitions showed that the former ones increase a 15% the overall accuracy. The rules can also be generated by an automatic procedure. A fuzzy decision tree induction algorithm, which is a fuzzy extension of the well known ID3 algorithm [20], is used. The main advantage of the decision trees is to generate incomplete rules, only defined by a subset of the available input variables. The generated rules are informative for experts to the condition that the partitioning is carefully defined. The Fispro [21] implementation is used with the partitions proposed by the experts, those which proved efficient in the expert fuzzy system.

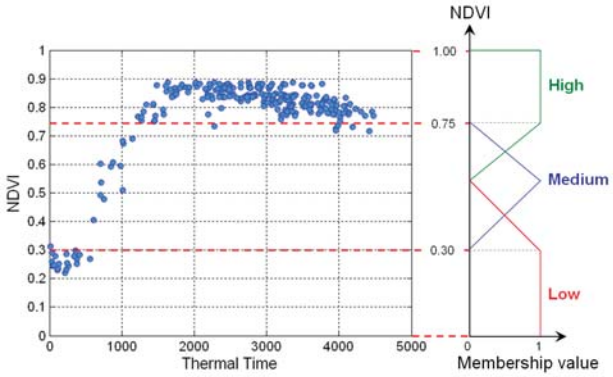


Figure 3: NDVI profiles plotted according to thermal time for several sugarcane fields. On the right are the fuzzy sets of NDVI-defined inputs.

5 Results and discussion

The learning process includes a cross validation. It has been carried out with various configurations to assess the system sensitivity to important parameters: the size of the training set and the number of satellite images per year. Finally, the contribution of each of the information sources is evaluated.
The data set has been randomly split to build training and test sets. For each proportion, the given result is the

ISBN: 978-989-95079-6-8

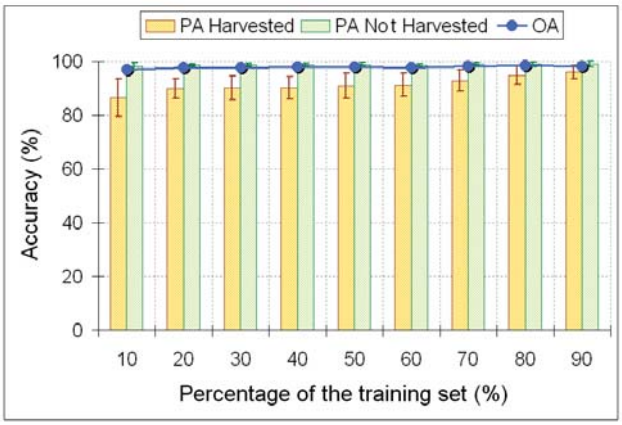


Figure 4: System performances using different percentages of training set (OA: overall accuracy; PA Harvested and PA Not Harvested: producer's accuracy for the two classes "Harvested" and "Not Harvested" respectively). These performances are obtained from confusion matrices.

average of ten runs. Nine percentages have been tested for the training set: 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80% and 90%. The performance is measured over the test set.

Fig.4 shows the average values and the standard deviations (for ten iterations) of the overall accuracy (OA) of the system and the producer's accuracy for the classes "Harvested" and "Not Harvested" (PA Harvested and PA Not Harvested, respectively). Let's notice that the performance for the Not Harvested decision is quite independent from the size of the training set, while the Harvested one increases with the sample size. This is due to the respective proportions of both cases in a given image: the not harvested fields are much more numerous than the harvested ones. So, only a small number of images is needed to learn the not harvested case.

The other key parameter related to system robustness, is the number of available images. It is likely to vary from a year to another because of the lack of acquisition or cloudy weather which makes unusable the satellite images.

Fig.5 shows the selected images for each of the studied years. The selection takes into account the campaign organisation. For instance, in the case of 3 images per year, one is chosen before, the other between and the last one after the harvest campaign.

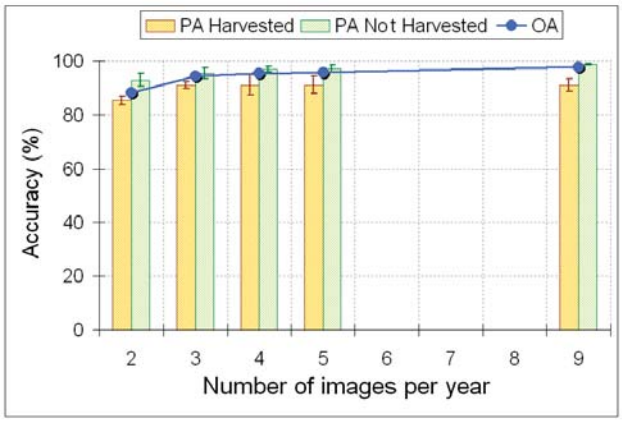


Figure 6: System performances according to the number of satellite images per year. (OA: overall accuracy; PA Harvested and PA Not Harvested: producer's accuracy for the two classes "Harvested" and "Not Harvested" respectively).

Fig.6 shows that the system performance is still acceptable even when the number of images per year drops to 3.

The number of generated rules also evolves with the number of images per year. It reaches an average of 38 for the 5 image configuration as shown in Fig.7.

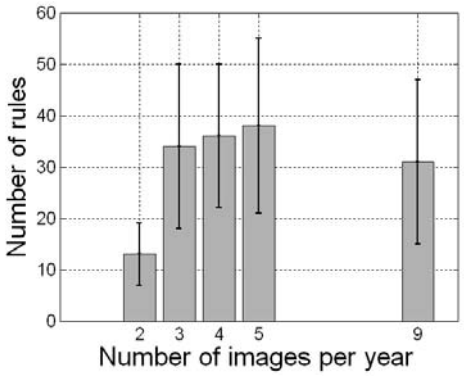


Figure 7: Number of induced rules according to the number of satellite images per year. Histogram bars represent the average values for the different training sets, and error bars the standard deviation.

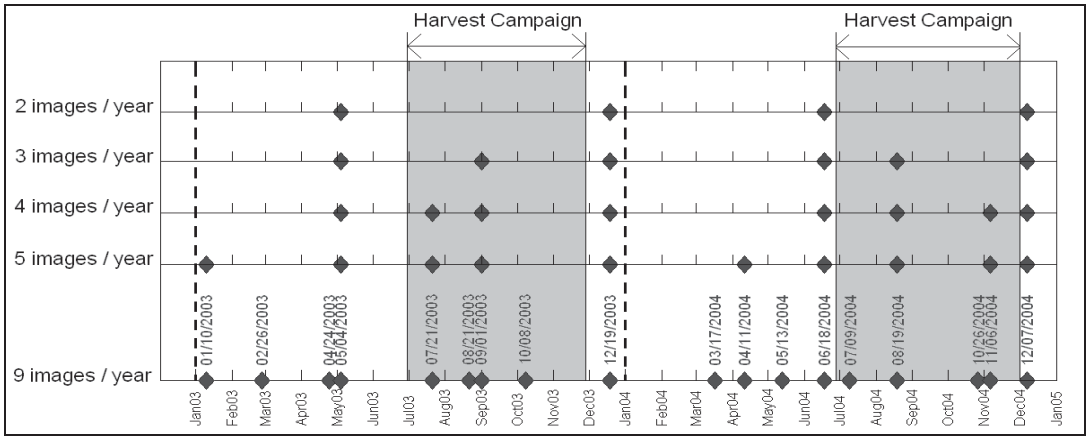


Figure 5: Acquisition dates selected at different stages of the assessment of system robustness with respect to the number of images per year.

ISBN: 978-989-95079-6-8

To assess the information source contribution three sets of input variables have been considered:

- SAT: the set of input variables that are related to the information extracted from the satellite images;
- MOD: the set of input variables built using the sugarcane growth model;
- EXP: the set of some input variables dealing with expert knowledge. Although the expert knowledge is difficult to isolate since it contributes to the design of the whole system, some variables are only related to this sources. They are encoding agronomic knowledge such as the nominal cycle length.

Four combinations have been tested and the results are reported in Fig.8. The best ones are the closest to the origin. As expected, all of them include SAT and MOD, and in most cases, the best one is SAT+MOD+EXP.

The sugarcane growth model contribution is interesting to analyze for each of the two classes "Harvested" and "Not Harvested". Fig.9 shows that it is higher in the harvested case. This result was expected: there is no need of a growth model to label a field as not harvested. The essential role of the model is to remove potential ambiguity concerning harvested fields.

6 Conclusions

This paper presented a novel approach for dealing with time series of optical satellite images used for crop monitoring. Data extracted from time series were combined with information from crop model output and expert knowledge, in order to make credible decisions. The description of the approach was done using an application example of sugarcane harvest monitoring with a SPOT5 time series. A decision system designed and implemented for automatic harvest detection was described.

Results obtained when evaluating the system were in substantial agreement with ground truth data; the overall accuracy reached 96.07%. The next step concerning the sugarcane application consists in examining the robustness of the automatically generated rules by testing the system at other sites and in other years.

The approach outlined in this paper is generic and very promising. Many models that simulate the growth of the main annual crops exist (e.g., STICS [22]), and expert knowledge about these crops could be obtained easily either from farmers or from agronomic knowledge bases [23].

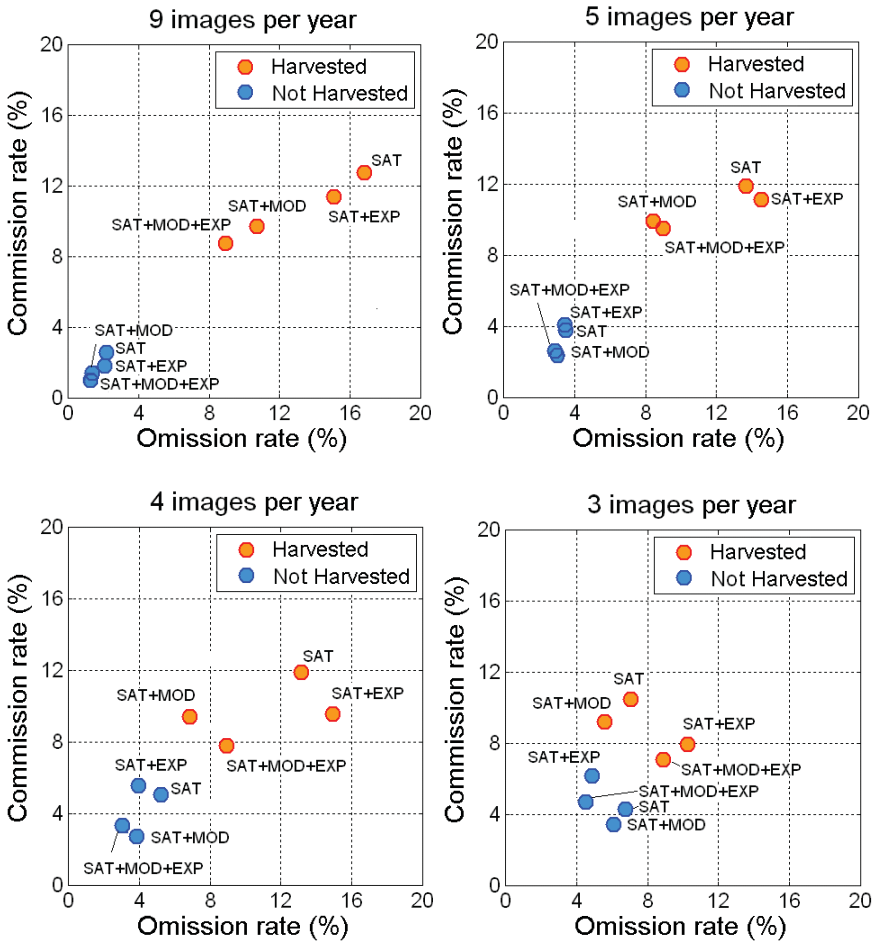


Figure 8: System performances obtained for different image frequencies by using different combinations of information sources.

ISBN: 978-989-95079-6-8

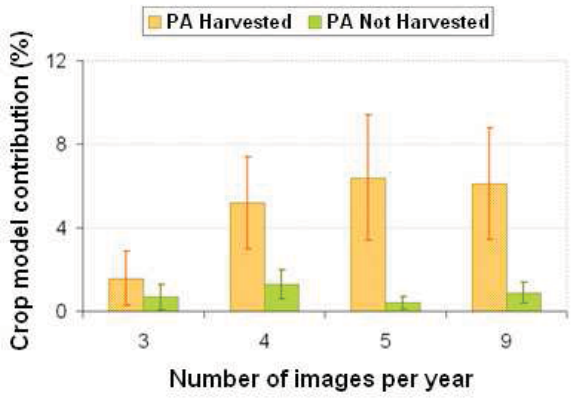


Figure 9: Crop growth model contribution in the system performances (PA Harvested and PA Not Harvested: producer's accuracy for the two classes "Harvested" and "Not Harvested" respectively). Histogram bars represent the average values for the different training sets, and error bars the standard deviation.

The combination of crop model outputs and expert knowledge with time series of high spatial-resolution satellite-images seems to be an excellent tool for crop monitoring.

Acknowledgment

We would like to thank Hélène de Boissezon and Sébastien Garrigues (CNES) for their help and support. We would also like to thank the CNES for funding the Kalideos database (http://kalideos.cnes.fr) and for generously providing the SPOT5 images.

References

[1] M. Pax-Lenney and C. E. Woodcock, "Monitoring agricultural lands in Egypt with multitemporal landsat TM imagery: How many images are needed?," *Remote Sensing of Environment*, vol. 59, pp. 522-529, 1997.

[2] G. A. Ippoliti-Ramilo, J. C. N. Epiphanio, and Y. E. Shimabukuro, "Landsat-5 Thematic Mapper data for pre-planting crop area evaluation in tropical countries," *International Journal of Remote Sensing*, vol. 24, pp. 1521-1534, 2003.

[3] C. S. Murthy, P. V. Raju, and K. V. S. Badrinath, "Classification of wheat crop with multi-temporal images: Performance of maximum likelihood and artificial neural networks," *International Journal of Remote Sensing*, vol. 24, pp. 4871-4890, 2003.

[4] M. D. Turner and R. G. Congalton, "Classification of multi-temporal SPOT-XS satellite data for mapping rice fields on a West African floodplain," *International Journal of Remote Sensing*, vol. 19, pp. 21-41, 1998.

[5] J. A. Martinez-Casasnovas, A. Martin-Montero, and M. A. Casterad, "Mapping multi-year cropping patterns in small irrigation districts from time-series analysis of Landsat TM images," *European Journal of Agronomy*, vol. 23, pp. 159-169, 2005.

[6] S. Panigrahy and S. A. Sharma, "Mapping of crop rotation using multidate Indian Remote Sensing Satellite digital data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 52, pp. 85-91, 1997.

[7] L. Bruzzone and P. Smits, "Analysis of Multi-temporal Remote Sensing Images : Multitemp 2001," in *Series in remote sensing*, vol. 2, A. P. Cracknell, Ed.: World Scientific Publishing Co. Pte. Ltd., 2002, pp. 440.

[8] P. Smits and L. Bruzzone, "Analysis of Multi-temporal Remote Sensing Images : MultiTemp 2003," in *Series in remote sensing*, vol. 3, A. P. Cracknell, Ed.: World Scientific Publishing Co. Pte. Ltd., 2004, pp. 387.

[9] C. Gers and E. Schmidt, "Using SPOT4 satellite imagery to monitor area harvested by small scale sugarcane farmers at Umfolozi," presented at 75th South African Sugar Technologists' Association (SASTA), 2001.

[10] E. Bappel, A. Bégué, P. Degenne, V. Lebourgeois, and B. Siegmund, "Real time sugarcane harvest monitoring using SPOT4&5 satellite data," presented at Proceedings of International Society of Sugar Cane Technologists (ISSCT), Guatemala, 2005.

[11] A. Bégué, P. Degenne, A. Pellegrino, P. Todoroff, and F. Baillarin, "Application of remote sensing technology to monitor sugar cane cutting and planting in Guadeloupe (French West Indies)," presented at Geomatica, Cuba, 2004.

[12] H. DeBoissezon and A. Sand, "Reference Remote Sensing Data Bases: Temporal series of calibrated and ortho-rectified satellite images for scientific use," presented at Proceedings of Recent Advances in Quantitative Remote Sensing, Valencia, Spain, 2006.

[13] http://kalideos.cnes.fr, "Images Spot: copyrigth CNES, Distribution Spot Image," 2007.

[14] M. El Hajj, A. Bégué, B. Lafrance, O. Hagolle, G. Dedieu, and M. Rumeau, "Relative radiometric normalization and atmospheric correction of a SPOT 5 time series," *Sensors*, vol.8, pp. 2774-2791, 2008.

[15] M. El Hajj, M. Rumeau, A. Bégué, O. Hagolle, and G. Dedieu, "Radiometric normalization of high spatial resolution multi-temporal imagery: A comparison between a relative method and atmospheric correction," presented at Proceedings of SPIE Europe Remote Sensing - The International Society for Optical Engineering, Florence, Italy, 2007.

[16] J. F. Martiné and P. Todoroff, "Le modèle de croissance Mosicas et sa plateforme de simulation Simulex: état des lieux et perspectives," *Revue Agricole et Sucrière de l'Ile Maurice*, vol. 81, pp. 133-147, 2002.

[17] M. El Hajj, A. Bégué, and S. Guillaume, "Multi-source information fusion: Monitoring sugarcane harvest using multi-temporal images, crop growth modelling, and expert knowledge," presented at MULTITEMP-2007, Fourth International Workshop on the Analysis of Multi-Temporal Remote Sensing Images, Provinciehuis Leuven, Belgium, 2007.

[18] M. El Hajj, A. Bégué, S. Guillaume, and J. F. Martiné, "Integrating SPOT-5 time series, crop growth modeling and expert knowledge to monitor agricultural practices - Case of sugarcane harvest on Reunion Island," *Accepted in Remote Sensing of Environment*, 2009.

[19] E. H. Mamdani and S. Assilian, "Experiment in linguistic synthesis with a fuzzy logic controller," *International Journal of Man-Machine Studies*, vol. 7, pp. 1-13, 1975.

[20] J.R. Quinlan, "Induction of decision trees," *Machine learning*, vol.1, pp. 81-106, 1986.

[21] S. Guillaume, B. Charnomordic, and J.-L. Lablée, "FisPro: An open source software for fuzzy inference systems," http://www.inra.fr/internet/Departements/MIA/M/fispro/, Ed.: INRA-Cemagref, 2002.

[22] INRA, "STICS : Simulateur mulTIdisciplinaire pour les Cultures Standard," http://www.avignon.inra.fr/stics/.

[23] G. Russell, R. I. Muetzelfeldt, K. Taylor, and J. M. Terres, "Development of a crop knowledge base for Europe," *European Journal of Agronomy*, vol. 11, pp. 187-206, 1999.

ISBN: 978-989-95079-6-8

Soft Computing Confronting Philosophical and Sociological Critiques to Classical AI

Verónica Sanz

Visiting Graduate Student
European Centre for Soft Computing
Mieres, Asturias (Spain)
E-mail: veronica.sanz@softcomputing.es

Abstract— *Artificial Intelligence has been one of the fields within Computer Science that has generated more interest and debates among philosophers. Later on, the most recent field of Science and Technology Studies (STS) also has shown some interest in AI. In both cases most of the authors have been quite critical about the promises, the practices and particularly the epistemological basis of Classical AI. The first part of the paper consists on an enumeration of the most important authors and their critiques to AI in Philosophy and STS studies. Since Soft Computing implies important changes with respect to traditional AI approaches like Symbolic AI, the second part of the paper will be devoted to confront Soft Computing with the critiques and challenges and to weight up to what extent Soft Computing could (or could not) answer differently than other AI approaches regarding the critiques received.*

Keywords— Epistemology, Philosophical critiques to AI, Science and Technology Studies (STS), Soft Computing.

1 Introduction

Since its origins Artificial Intelligence has been one of the fields within Computer Science that has generated more interest and debates among philosophers and cognitive scientists. Not in vain, Artificial Intelligence deals with some of the traditionally most debated issues in philosophy such as knowledge and reasoning. In concrete, some of the ambitious statements of the first AI researchers touch very sensitive "fibers" of philosophers. Authors like John Searle, Daniel Dennett, Hubert Dreyfus or Roger Penrose were among the first ones to cope with these issues [1, 2, 3, 4, 5].

The philosophical debate during the 1970s and 80s focused on Artificial Intelligence via Symbolic AI (the branch of artificial intelligence research that concerns itself with attempting to explicitly represent human knowledge in a declarative form -i.e. facts and rules-), since this was the approach followed by the main scientists in the field from the early days on. The main interest of philosophers has been traditionally centred in the question about whether computers can or cannot think, which generally implies doing it "in the sense humans do". There was an assumption among philosophers (though induced by claims made by some of the first AI researches) that the aim of AI is to create "an artificial mind". Despite the successes of some AI symbolic systems during the 60s, the enormous promises of early AI texts posed the debate in terms of the possible success or failure of the whole enterprise of creating an artificial mind.

Because of that the discussion was pursued in a very bi-polar manner with some philosophers (like Dennett) clearly optimistic about the question and others (like Searle and Dreyfus) very critical about it. The philosophical critique was in tune with a significant resistance in general culture, which tried to assure the specificity of human nature with regard to computational artifacts, often appealing to non-cognitive aspects of the mind (intentionality, emotion, perception, etc).

2 Philosophical Critiques

2.1 Turing

The British mathematician Alan Turing was not only the one who set the basic elements of computation, but also the one who posed the first stone of this debate about the possibility of "thinking machines", proposing a mental experiment known, since then, as the "Turing test" [6]. It consists on an interrogator asking questions (through a separating barrier) to a machine and a human, trying to guess who (which) is answering. If the interrogator cannot tell whether the responder is the human or the machine, then the intelligence of the machine would be proved.

The interest of Turing's idea lies on the practical way he is proposing the test as an "imitation game". He is not defining, at least *a priori*, what intelligence *is* but establishing a practice were the performance of human intelligence is showed. Whether the machine thinks "in the way humans do" or not is not addressed by the Turing's test. In fact, it is precisely this kind of philosophical question what he tries to avoid.

2.2 Searle

In his famous paper from 1980 [1] the American philosopher John Searle argued against the strong claim that computers can be intelligent *in the same way* humans are, which for him implies having cognitive or psychological states (i.e. conscious understanding and intentionality). He distinguishes two strands of AI. The former claim is called "Strong AI": intelligent machines would replicate not only intelligent *products,* but the intelligent *process*[1]. By contrast, "Weak AI" is the perspective that understand computers as useful tools that mimic some of the abilities and tasks of humans (included those that are considered intelligent), in part because they can simulate "smart actions"[2]. Weak AI does not make the claim that computers *actually understand* or *have* mental states.

Searle's critique is directed to Strong AI. To do that, he proposes a thought experiment known as "The Chinese room" in response to Alan Turing's suggestion to replace the question "Can machines think?" with the question of whether

ISBN: 978-989-95079-6-8

they can succeed in an "imitation game". In short, Searle's conclusion is that a native English speaker who knows no Chinese together with a book of instructions for manipulating the symbols would able to pass out Chinese symbols which are correct answers to the questions. The instructions enable him/her to pass the Turing Test but he does not understand a word of Chinese[3].

The axis of the argument relies also on the idea that *syntactic manipulation* (formal symbol manipulation or computation) does not imply understanding *semantics* (understanding of meanings). For Searle, we can not attribute intelligence to such a system since "real intelligence" is related with understanding. The background argument of the Chinese room is that one of the main features of human minds is intentionality (which is a mental state) and that is something that computers do not (and can not) have.

2.3 Dennet

Though many of the philosophers of AI have been very critical, Daniel Dennett is an exception [2]. He shared with Symbolic AI researchers what in philosophy of mind and cognitive sciences is called the "Computational (and Representational) Theory of Mind". Dennett has been positively linked to AI and he has participated in some AI projects. He disagrees with Searle in two important points:

a) The way that Searle understands intentionality is already assumed from the beginning as limited to humans. Against that, Dennett affirms that human consciousness and intentionality are not as specific as we use to think, which is showed in the way we use a "mentalist language" when describing the actions of animals, machines and so on.

b) The Chinese Room thought experiment is very ingenious but is very far from actual practices in everyday life. It involves a number of things that are quite impossible to do so it is very doubtful to be used as a proof of something[4].

2.4 Dreyfus

However, the philosopher who most energy and time has devoted to the critique of AI during the past 30 years has been the phenomenologist Hubert Dreyfus [3, 4]. Drawing upon Heidegger and existential phenomenology, his arguments are quite different from Searle's (though he also affirms that neither current machines nor future ones would be intelligent).

Dreyfus criticizes the epistemological assumption of Symbolic AI that all activity can be formalized in the form of propositional knowledge (which would allow computation in the form of "rule-based symbol manipulation") as it is done in many Symbolic AI based expert systems. For Dreyfus, many parts of human knowledge and competences cannot be reduced to an algorithmic procedure, i.e. the kind of knowledge one might employ when learning a skill like walking, riding a bike or ice skating. This learning is done through practice with our bodies: to learn how to do it one must get on and try it, it's not enough to follow written rules. This requires having a body and experimenting with it in a changing environment.

Though Dreyfus concedes that certain well-circumscribed experts systems for particular domains have succeeded, he predicts that many others will never fulfil the expectations (i.e. Lenat y Guha's "Cyc" [7]). This is not because of technical shortages (that could be overcome in the future) but because symbolic type of artificial systems are incapable of representing bodily-skills knowledge, which can not be captured in an algorithm.

3 Soft Computing confronting Philosophical Critiques

3.1 Not related to Strong AI

The problem with Searle's thought experiment is that most of practitioners in AI do not take issue with intentionality or creating an artificial mind. This debate is in fact a philosophical one (the old debate between Behaviorism and Mentalism[5]) and is far from the interests of Soft Computing (SC) practitioners, as far as I could observe. Although this debate was in part provoked by the excessively optimistic predictions of some of the founding fathers of AI, it can be said that the SC community is more focused on solving concrete problems by constructing machines that can perform "intelligent" tasks, which is more in tone with what Searle called "Weak AI". Because of that, these kind philosophical critiques do not apply to Soft Computing.

However in one of his last articles, Zadeh says that there is a possibility of achieving "human-level machine intelligence", though for that there is needed a paradigm shift [8].

3.2 Necesity of Semantics

The other axis of Searle's critique relied on the idea that *syntax* (formal symbol manipulation or computation) does not imply *semantics* (understand meanings). Searle applied it to say that, to have "real intelligence", syntax is not enough. One of the fundaments of Fuzzy Logic is precisely to deal with meaning in natural language. Fuzzy Logic agrees with Wittgenstein's definition of meaning as its use in language [9], that is, as used by concrete people in concrete contexts.

To model and operate with natural language and its non-static meanings would require tools which can deal with imprecision and uncertainty, which is precisely the main characteristic of Fuzzy Logic. The enormous importance of natural language for SC is shown in the turn Zadeh proposed in 1996 saying that Fuzzy Logic was equivalent to "Computing with Words*"* [10]. Computing with words (CW), Zadeh says, is a methodology in which words are used in place of numbers for computing and reasoning. As fuzzy sets are able to model some natural language's concepts, it would be possible to use these models to compute directly with them.

3.3 The "Bodily-skills knowledge" Problem

Dreyfus's main critique to symbolic type of artificial systems is its incapability of representing bodily-skills knowledge. Other types of AI systems have tried to avoid this problem, like Rodney Brooks' "bottom-up" strategies in *embedded* or *situated robotics* [11]. The work of Brooks and his colleagues focus on *embodiment,* the idea that intelligent behaviour is an *emergent* phenomenon resulting from the

ISBN: 978-989-95079-6-8

individual's interactions with its environment, which are more related to action than to representation.
What Soft Computing has to say about this debate? In general we could say not much. The problem of embodiment is typical in Robotics (where there are "embodied machines", not only software) but not necessarily in most of the branches of AI. Fuzzy Logic indeed deals with language and representation of meanings and not so much with actions. However, applications of Fuzzy Logic to concrete systems (fuzzy control) do actually deal with systems-in-action. One clear example is the "inverted pendulum" solved by Yamakawa [12] using a small set of simple fuzzy rules. Fuzzy control system design is based on empirical methods, basically an approach of trial-and-error, which is quite related to the way humans learn bodily-skills knowledge.

4 STS Critiques

Among the different non-technical disciplines not only philosophy has been interested in the study of AI. Within the so-called area of STS, at least some practitioners have turn around to Artificial Intelligence.
An STS methodology implies the analysis of the social and cultural contexts where all scientific and technological developments take part. This model resists classical accounts that science and technology are totally objective and neutral with respect to the context where they are made.
Regarding AI, STS takes a different stand on that philosophical critique. They don't take part in abstract discussions about the ultimate goal of AI. Their methodology consists on looking at real practices, including ethnographic observation of AI practitioners. From the STS view, the important thing is the way these engineers understand and use (or model) concepts such as knowledge and reasoning in their daily practice.

4.1 Harry Collins

Harry Collins' study of AI [13] is the most important contribution from part of STS to the study of AI. The field of Experts Systems has attracted the most attention from part of social scientists. His critique is in a way quite similar to Dreyfus' regarding the impossibility of formalizing certain types of knowledge.
Based on recent Sociology of Scientific Knowledge [14] and Laboratory Studies [15], in which is shown how experimenters need to share a culture in order to produce science, Collins maintains that skills (know-how knowledge) are principally acquired through enculturation and the acquisition of tacit capacities whose transmission is implicit (and cannot, therefore, fully be spelt out in formal algorithms). This kind of knowledge it is commonly known in STS as "tacit knowledge", which can be also understood as the "common knowledge" of a particular community (what is taken for granted within it)[6].
However, Collins understands "tacit knowledge" more widely than Dreyfus' embodied knowledge. Acquisition of skills-type knowledge in his view is related to socialization rather than embodiment. The underlying argument of Collins against Dreyfus is the importance of natural language. Collins affirms that one can learn to speak fluently about the world without being bodily involved in it. This model assumes that socialization (in a certain group) is necessary to communicate knowledge. In this way, Collins brings the discussion to the cultural arena
Another of the issues addressed by Collins is his focus on AI systems that are already being used and seem like functioning, not in possible future ones. The challenge is, then, to explain that "success". On the contrary of philosopher's understanding of computers as artificial brains, to him the role (and success) of AI is to create "social prostheses". And it is the ability of humans to interpret their actions and repairing their deficiencies which makes them successful. In a way, humans interpret machines actions as one kind of actions humans can perform (what Collins calls "machine-like action"), but this is only a little part of the whole kind of humans' actions. This human ability to interpret and repair normally remains invisible, but, for Collins, is here where the intelligence attributed to machines rests on.

4.2 Diana Forshyte

The work of Diana Forsythe [16] belongs to the field of Anthropology of Science[7] rooted in cultural anthropology's tradition. This normally implies a more or less long-term field research using the ethnographic method of participant observation. In her ethnographic study of an Expert Systems community in the US, she investigated what she called the "Culture of the Experts Systems Engineers", defined as the values and assumptions (explicit but mostly implicit) that constitute what they take for granted. Forshyte also refers to this as "common sense truths", which are those things 'everybody knows' within a given setting (this shared and tacit knowledge is indeed the main criterion to be considered a member of the community). The importance of the ethnographic method relies on the main thesis of cultural anthropology about the complex relation between beliefs and practices: usually what people do, what they think they do, and what they report they do are someway different.
Investigating the culture of the Expert Systems practitioners she found they share a particular notion of knowledge (a very restricted one), that has important implications for the products they make. Their epistemological stances are very important because they are going to be incorporated (encoded) into the expert systems produced. To prove this she focused on the first step of the process of building an expert system: 'knowledge acquisition' (also known as elicitation process), consisting in the "extraction" of the knowledge from human experts. The very use of the terms "extraction" and others such "store" (knowledge is presumed to be "stored" in the head of the experts) are metaphors which implies an understanding of knowledge as objective and susceptible of being formally transferred into a machine.
Despite that, the elicitation process is considered by knowledge engineers themselves a very persistent problem and is thought to be the cause of failures in the systems when encountering 'real-world situations' (situations that the systems builders did not anticipate). Knowledge engineers blame human experts interviewed for not being able to explicit clearly their expert knowledge, that is, in the way they can encode it (it is not casual that the kind of expert

ISBN: 978-989-95079-6-8

knowledge they prefer to work with are already those very formalized in well-defined narrow domains).
They would never think about the possibility that their implicit assumptions about knowledge can be the cause of their problems. Forshyte affirms that the knowledge encoded in experts systems is static, brittle and narrow, while in real life knowledge is continuously modified in relation to contextual factors and trough negotiations between different actors, and that is the real cause of the 'brittleness' of the systems[8].
In general terms, the kind of expert systems studied by Forshyte are included in the more general approach of Symbolic AI, since, in the end, knowledge "extracted" from human experts has to be written-down in rules-type propositional knowledge. With respect to this, the same critique from Dreyfus and Collins about the inability to model other types of knowledge (particularly social knowledge) is applicable relating Forshyte's work.

4.3 Alison Adam

One tradition within STS studies are those investigating the relations of gender factors with science and technology[9]. Notions of gender are constructed not only in relation to economy or moral values but also science and technology became in our culture more associated with the masculine than the feminine, reflecting cultural's view of what constitutes masculine or feminine. However this is not a fixed but a historical process[10].
Alison Adam is a British computer scientist and feminist who has pursued the most extended study about gender and Artificial Intelligence [17]. She accurately applies the insights of feminist epistemology (an alternative kind of epistemology which criticizes the classical conservative epistemologically) to the main traditions in AI as Symbolic AI, Expert Systems, and also more recent approaches such as Artificial Life and Situated Robotics. Adam's work is devoted to "uncover" the implicit (gendered) epistemological assumptions in AI, because on them depend how AI researchers understand some key concepts which very much influence their work such 'intelligence', 'knowledge', 'reasoning' and 'efficiency'. The idea, then, is not to look for the contributions of individual women in the history of AI, but rather to look if there can be gendered models of knowledge implicit in AI systems.
One of Adam's theses is that traditional AI systems are based on the Cartesian ideal of a disembodied mind and the over-valuation of mental (abstract) knowledge over corporeal (concrete) knowledge; the former has been historically associated with masculine realm while the latter has been largely associated with the feminine. For Adam as for other feminist epistemologists [18, 19], this is very much related to the "mathematization" of the Western Culture (started in ancient Greece and increased in the Modern Age) and the idea that the entire universe was governed by numbers. Although there was already among the Greeks a plea for other ways of knowing it, was this Platonic and then Aristotelian conception of the world which finally succeeded. Aristotle is considered "the father of Logic", and defines the correct way of reasoning in terms of a dichotomous logic (there are only true or false statements which imply only valid or invalid inferences). As we know, the rejection of this thesis was the starting point of Zadeh's Fuzzy Logic.
Following the work of the philosopher Lorrain Code [20], Adam shows that traditional AI is based on the rationalistic epistemological model of "S knows that p". 'S' refers to the knowledge subject (*who knows*), and 'p' the object of that knowledge (w*hat can be known*).
Regarding 'p', as we already said, propositional knowing (or *knowing that)* is considered to be superior knowledge while others (skill-knowledge or *knowing how*) are less important. Feminist epistemologies consider this stratification as an *epistemic discrimination* (which they affirm is not innocent of gender factors).
Regarding 'S', the ideal knower of Classical AI is an anonymous, universal and "knowing from nowhere" subject[11]. On the contrary, Adam emphasizes the "situatedness" of the knower. For example, who are the "knowledge subjects" of systems like "Cyc"? This project leaded by Douglas Lenat [7] tries to build a vast knowledge base which will include most of human common-sense knowledge. Above all, Lenat assumes that there is one *and only one* consensus reality available common to all humans. The development of STS and Anthropology is the last decades have shown that it is quite problematic to assume that hypothesis. We "see" the world according to our theories of the world, and these theories depend on our cultural location. At bottom, what CYC knowledge base would consist on is what Adam's calls "*The-World-As-The-Builders-Of-Cyc-Believe-It-To-Be",* who are a group of predominantly middle-class male US university professors, that, without realizing it or intending it, are privileging their "consensus reality" over other groups' [17].
As a way of summarizing the differences between the two kind of epistemologies we can make the following table:

Table 1: Contrasting Epistemologies

Symbolic AI Epistemology	**STS Epistemology**
Objective Knowledge *(context free, universal)*	Situated Knowledge *(contextual, local)*
Propositional knowing (*knowing that*)	Skills-type and Common-sense Knowledge
Abstract reasoning	Concrete reasoning
Epistemological Monism	Epistemological Pluralism
Lack of Reflexivity	Accountability and Responsibility

5 Soft Computing confronting STS Critiques

5.1 Contextualism

As we said before, workers in the field of Symbolic AI claim that cognition is the manipulation of internal symbols by logical rules, so, therefore, human knowledge is, to a large extent, *context free*. We have explained how this idea has been criticized by authors like Dreyfus or Collins.
Wittgenstein's pragmatist turn in *The Philosophical Investigations*, which is assumed by SC, implies that meaning can only be grasped when used by concrete people

ISBN: 978-989-95079-6-8

in concrete contexts. The philosopher, Wittgenstein says, must "look and see" the variety of uses to which the word is put in particular cases. In this way, previous empirical analysis of the context of use it is an obligatory step for SC developments.
This obligates SC to be an e*mpirically-based research,* not only a theoretical approach (this is also because the ultimate goal of SC is engineering, that is, constructing machines, which in the end has to be used by users).
In spite that the idea of contextualism may be implicit in SC practitioners, the problem about the "situatedness" of the knowledge subject regarding the difference or culture, class, nation, race or gender has not been addressed as such neither in theory nor in specific projects, as far as I know.

5.2 The "Common-sense knowledge" Problem

As we have seen, the apelation to the inability of Classical AI to deal with common-sense knowledge is an important part of STS critiques. Their argument is that this kind of knowledge is mostly tacit and took for granted, and, because of that, it can not be formalized.
On the contrary, many SC articles contain the affirmation that fuzzy logic could represent common-sense knowledge. The most explicit article from Zadeh about the topic [21] explain that common-sense knowledge is expressed in *dispositions*, as opposed to propositions. A disposition in philosophy is a kind of belief that is stored in the mind but is not currently being considered (in a way we can say is a kind of tacit knowledge). For the purpose of representation Zadeh defines a disposition as a proposition with implicit fuzzy quantifiers, so in that way fuzzy logic can be used to deal with that kind of knowledge.
Although this strategy is only in its first steps, I think is a very promising way of dealing with common-sense knowledge, at least much better than Guha and Leant kind of propositional knowledge data base. However, there are many issues that are not addressed or solved yet in SC. A very important one has to do with the very definition of "common-sense knowledge" and its interrelations with concepts as "tacit knowledge", "skill-knowledge", "situated knowledge" and so on. Is it all common-sense knowledge tacit or just "taken for granted"? The difference between common-sense and skill-knowledge is that one can be formulated in *dispositions* and the other one not? What Zadeh exactly means when he talks about "world knowledge" –defined as "the knowledge humans acquire though experience, communication and education"?

5.3 Pluralism and Hybridization

Though we have not explicitly say it before, one of the main proposals from STS and feminist epistemologies is the appeal to alternative ways of knowing and doing science. Lucy Suchman, regarding computer science, talks about **"**expanding the frame of basic modeling assumptions" [22], which means expanding the very foundational paradigms of the discipline (particularly its assumptions about knowledge) and introduce alternative paradigms. This, regarding theory of knowledge, is what is called *Epistemological Pluralism*.
I think SC can be a little bit closer to these suggestions than other approaches. On one hand the basic characteristic of SC is precisely the combination of methods. Quoting Zadeh: *"SC is not a single methodology; rather is a consortium of methodologies which are aimed at exploiting the tolerance for imprecision (...) The constituent methodologies in SC are for the most part complementary and synergistic rather than competitive* [22]*".* So we can say pluralism and hybridization are at the very core of what SC is, but… do they mean with that the same as STS scholars when they mention it?
This question leads us to the following point.

5.4 Issues of Accountability and Responsibility

Pluralism in STS and feminist epistemology is related to issues of accountability and responsibility, which means that the ultimate goal is to introduce (some) changes in science and technology to make them more inclusive and socially just. Authors as Lucy Suchman [23] and Philip Agre [24] whose work is devoted to Artificial Intelligence, appeal to epistemological and methodological pluralism as a way of accommodating diversity. Agre calls this possible way of doing technology a *critical technical practice*, which consists on identifying core metaphors of the field, noticing what (when working with these metaphors) is marginalized, and inverting dominant metaphors bringing alternative perspectives from the margins of the discipline to the center.
As far as I could observe, Soft Computing practitioners have not introduce the issues of accountability and responsibility explicitly in their practice (on the other hand like most part of computer science and technology except but some particular approaches in software engineering like participatory design or user-centred design). However my suggestion is that, because of some particular characteristics (that we already pointed out), Soft Computing is more suitable to include this issues than other more abstract approaches such, for example, Symbolic AI.

6 Conclusion Remarks

Along the article I have enumerated various critiques made to traditional approaches in Artificial Intelligence from part of Philosophy and, later, from Science and Technology Studies. Soft Computing distinguishes itself very much from traditional AI approaches. Because of that, it appears to me as a very interesting topic to question how Soft Computing could cope with these critiques and challenges.
Although this research is still in a preliminary state, I think the issues treated in this article can provide a good frame for future work in the new field of Soft Computing in Humanities and Social Sciences. In addition to that, it could be also of some use for Soft Computing itself. Soft Computing have to face many challenges in order to improve other AI methodologies and try to achieve what they weren't able to. Having a clear picture of these challenges will be very useful. Philosophical and sociological studies of science could help to identify problems and challenges, and, although far from engineering practice, can inspire and suggest new ways of facing the most durable challenges.

ISBN: 978-989-95079-6-8

Acknowledgments

The author wants to thank Sergio Guadarrama for his comments and fruitful discussions about the topic of the paper.

In addition, I want to thank the Foundation for the Advancement of Soft Computing for providing me the best place and environment to proceed with my research in this topic.

[1] Strong AI is would be, then, the materialization of the "Computational Theory of Mind" (the philosophical view that holds that the mind is *literally* a digital computer and reasoning consists on formal manipulations of symbolic representations of the world). Through this thesis the debate became not only a philosophical one but also a psychological one: it provides a working model of human intelligence.

[2] In fact this is in tune with the most quoted definition of AI: *"AI is the ability of a computer or other machine to perform those activities that are normally thought to require intelligence"* [25]

[3] For a detailed description see *Stanford Encyclopedia of Philosophy's* entry "The Chinese Room Argument".

[4] Actually Dennett elaborated this critique in general against all kind of "thought experiments".

[5] That is, whether or not there are "internal mental states" in the mind that can be studied by psychologists, or only observable behavior.

[6] Very much related with a very important concept in STS known as "epistemic cultures"

[7] The pioneers of the application of Anthropology to the study of science are the so-called laboratory studies of Latour and Woolgar [15], Karin Knorr-Cetina, and Michael Lynch.

[8] As an anthropologist whose background understanding about knowledge is precisely the opposite (in Social Sciences knowledge is a highly problematic issue, mostly tacit and unconscious and whose "locus" is the a social group or community) is very shocking to confront herself with such a different approach.

[9] We should explain here what we mean with "gender" in gender studies. It is not about the gender of the particular individuals involved in science and technology (particular men or women) not even men and women in general, but the gender systems (defined as gender structures, gender identities and gender symbolisms) in which these science and technology is developed (in our case Western gender system), and how this system influence its process, for example gendered models of knowledge, which is the main topic of the so-called Feminist Epistemology.

[10] However there is no intrinsic reason why a technology may become associated with the masculine or the feminine gender, this may change over time. It is precisely in the contingence of this process where feminism locates its hope in change towards a more egalitarian relation between women and technology.

[11] This epistemology considers logic as the paradigm of valid knowledge. Logic characterizes precisely by depersonalizing the subject and being the paradigm of context-free knowledge.

References

[1] John Searle. Minds, brains and programs, *Behavioral and Brain Sciences* 3:417-457, 1980

[2] Daniel Dennett. *The Intentional Stance*, Cambridge: MIT Press, 1987

[3] Hubert L. Dreyfus. *What Computers Can't Do: The Limits of Artificial Intelligence.* New York: Harper and Row, 1972.

[4] Hubert L. Dreyfus. *What Computers Still Can't Do: A Critique Of Artificial Reason ,* MIT Press, Cambridge, MA, 1992

[5] Roger Penrose. *The Emperor's New Mind* Oxford: OUP, 1989.

[6] Alan Turing. Computing Machinery and Intelligence, *Mind* 59:433-460, 1950

[7] Douglas B. Lenat and R.V. Guha, Building Large Knowledge-Bused Systems: *Representation and Inference in the Cyc Project.* Reading, Mass.: Addison Wesley, 1990

[8] Lotfi A. Zadeh. Toward Human Level Machine Intelligence- Is It Achievable? The Need for a Paradigm Shift. *IEEE Computational Intelligence Magazine*, August 2008:11-22

[9] Ludwig Wittgenstein. *Philosophical Investigations (PI)*, G.E.M. Anscombe and R. Rhees (eds.), Oxford: Blackwell, 1953.

[10] Lotfi A. Zadeh. Fuzzy Logic = Computing with Words. *IEEE Trans Fuzzy Systems*, 4(2):103-111, 1996

[11] Rodney A. Brooks. Intelligence without representation, Artificial Intelligence 47:139–159, 1991.

[12] Takeshi Yamakawa., Stabilization of an inverted pendulum by a high-speed fuzzy logic controller hardware system, *Fuzzy Sets and Systems* 32:161–180, 1989.

[13] Harry M. Collins, *Artificial Experts: Social Knowledge and Intelligent Machines. Cambridge*, Mass.: MIT Press, 1990.

[14] David Bloor. *Knowledge and Social Imagery*, London, Routledge, 1976.

[15] Bruno Latour and Steve Woolgar. *Laboratory Life,* Beverly Hills: Sage Publications, 1979.

[16] Diana Forshyte "Engineering Knowledge: The Construction of Knowledge in Artificial Intelligence", *Social Studies of Science*, Vol. 23, No. 3: 445-477, 1993

[17] Alison Adam. *Artificial Knowing. Gender and the Thinking Machine*. London and New York, Routledge, 1998

[18] Sandra Harding. *The Science Question in Feminism,* Ithaca, Cornell University Press, 1986.

[19] Evelyn Fox Keller. *Reflections on Gender and Science.* New Haven: Yale University Press, 1985.

[20] Lorraine Code, *What Can She Know? Feminist Theory and the Construction of Knowledge*. Ithaca, N.Y.: Cornel1 Univ. Press, 1999

[21] Lotfi A. Zadeh. Commonsense Knowledge Representation based on Fuzzy Logic, *IEEE*: 61-65, October 1983

[22] Lotfi A. Zadeh. Soft computing and fuzzy logic. IEEE Software 11(6):48-56, 1994.

[23] Lucy Suchman. *Human-Machine Reconfigurations. Plans and Situated Action* 2nd Edition, Cambridge: Cambridge University Press, 2006.

[24] Philip Agre. *Computation and Human Experience,* Cambridge: Cambridge University Press, 1997.

[25] Enric Trillas. *La inteligencia artificial. Máquinas y personas.* Madrid, Ed. Debate, 1998

Optimizing Fuzzy Flip-Flop Based Neural Networks by Bacterial Memetic Algorithm

Rita Lovassy[1,2] László T. Kóczy[1,3] László Gál[1,4]

1 Faculty of Engineering Sciences, Széchenyi István University Győr, Hungary
2 Inst. of Microelectronics and Technology, Kandó Kálmán Faculty of Electrical Engineering, Budapest Tech, Budapest, Hungary
3 Dept. of Telecommunication and Media Informatics, Budapest University of Technology and Economics, Hungary
4 Dept. of Technology, Informatics and Economy
Szombathely University of West Hungary
Email: lovassy.rita@kvk.bmf.hu, koczy@sze.hu, gallaci@ttmk.nyme.hu

Abstract—In our previous work we proposed a Multilayer Perceptron Neural Networks (MLP NN) consisting of fuzzy flip-flops (F^3) based on various operations. We showed that such kind of fuzzy-neural network had good learning properties. In this paper we propose an evolutionary approach for optimizing fuzzy flip-flop networks (FNN). Various popular fuzzy operation and three different fuzzy flip-flop types will be compared from the point of view of the respective fuzzy-neural networks' approximation capability.

Keywords— Bacterial Memetic Algorithm, feedbacked fuzzy J-K and fuzzy D flip-flops, Multilayer Perceptron Neural Networks

1 Introduction

Fuzzy set theory, artificial neural networks, bacterial evolutionary algorithms and their hybrid combination have been the subject of intense study and application, especially in the last decades. These were developed with the aim to deal with problems which were hard to solve using traditional techniques. Fuzzy systems are transparent and interpretable, neural networks possess the property of auto-adaptability, while evolutionary, especially bacterial algorithms have been used for the approximation of the optimal structure. To approximate various test functions we use a combination of the above mentioned three main branches of Computational Intelligence. Although some paper report about Support Vector Machines (SVM) outperforming traditional Multilayer Perceptron (MLP) in classification task, but in prediction and regression problems the MLP gives smaller errors with lower network complexity [12]. The paper is organized as follows. After the Introduction Section 2 defines triangular norms and co-norms; and then highlight the five well known fuzzy operations (algebraic, Yager, Dombi, Hamacher and Frank). Section 3 presents the concept of a single fuzzy J-K and D flip-flop, using the fundamental equation as it was proposed in [13]. A comparative study of several types of fuzzy J-K flip-flops based on the five fuzzy operations is made. Comparisons of fuzzy J-K flip-flops with feedback and of two different interpretations of fuzzy D flip-flops are presented. Section 4 is devoted to the investigation of the F^3 based neurons and the MLP [11] constructed from them. We proposed the Fuzzy Flip-Flop Neural Network (FNN) architecture that can be used for learning and approximating various simple transcendental functions. In Section 5 we present Bacterial Memetic Algorithm with Modified Operator Execution Order (BMAM) [5] for FNN variables optimization. Experimental results and comparison between different types of FNNs trained with Levenberg-Marquardt (LM) method and the BMAM are discussed in Section 6, followed by a brief conclusion in the last Section.

2 Fuzzy operations

Klement, Mesiar and Pap enumerate and give the basic definitions and properties of the most general triangular norms in [8], including also graphical illustrations, comparisons. In general a triangular norm [8] (t-norm) is a binary operation T on the unit interval [0, 1], i.e., a function $T:[0,1]^2 \rightarrow [0,1]$, such that for all $a, b \in [0,1]$ the commutativity, associativity, monotonicity and boundary conditions are satisfied. Triangular co-norms (s-norm) were introduced [8] as dual operations of t-norms.

Table 1 shows various t-norms and s-norms selected by us, (as possible bases of the FNN), all five can be used as basic operators in the definition of the fundamental formulas of the fuzzy flip-flop introduced in the next Section. (Many other did not produce good enough neural network properties, cf. [10].) For simplicity we denote the t-norm with i, and the s-norm with u, where the subscripts refer to the initial of the name of the norm: e.g., in case of Yager norms: $i_Y(a,b) = a\ i_Y\ b$ and $u_Y(a,b) = a\ u_Y\ b$.

3 Fuzzy flip-flops

3.1 Fuzzy J-K Flip-Flops

The fuzzy J-K flip-flop is an extended form of the binary J-K flip-flop. In this approach the truth table for the J-K flip-flop is fuzzified, where the binary NOT, AND and OR operations are substituted by their respective fuzzy counterparts, i.e. fuzzy negation, t-norm, and co-norm respectively.

ISBN: 978-989-95079-6-8

Table 1: Some t-norms and co-norms.

Fuzzy operation	t-norm; i (a, b)	s-norm; u (a, b)
Algebraic (A)	ab	$a+b-ab$
Yager (Y)	$1-\min\left[1,((1-a)^w+(1-b)^w)^{1/w}\right]$	$\min\left[1,(a^w+b^w)^{1/w}\right]$
Dombi (D)	$\frac{1}{1+\left[(1/a-1)^\alpha+(1/b-1)^\alpha\right]^{1/\alpha}}$	$\frac{1}{1+\left[(1/a-1)^{-\alpha}+(1/b-1)^{-\alpha}\right]^{-1/\alpha}}$
Hamacher (H)	$\frac{ab}{\gamma+(1-\gamma)(a+b-ab)}$	$\frac{a+b-(2-\gamma)ab}{1-(1-\gamma)ab}$
Frank (F)	$\log_s\left[1+\frac{(s^a-1)(s^b-1)}{s-1}\right]$	$1-\log_s\left[1+\frac{(s^{1-a}-1)(s^{1-b}-1)}{s-1}\right]$

Parameters w, α, γ and s lie within the open interval $(0, \infty)$. The next state $Q(t+1)$ of a J-K flip-flop is characterized as a function of both the present state $Q(t)$ and the two present inputs $J(t)$ and $K(t)$. (For simplicity (t) will be omitted in the next.) The so called fundamental equation of the J-K type fuzzy flip-flop [13] is

$$Q(t+1) = (J \vee \neg K) \wedge (J \vee Q) \wedge (\neg K \vee \neg Q) \quad (1)$$

where $\neg, \wedge, \vee$ denote fuzzy operations (e.g. $\neg K = 1-K$). As a matter of course, it is possible to substitute the standard operations by any other reasonable fuzzy operation triplet (e.g. De-Morgan triplet), thus obtaining a multitude of various fuzzy flip-flop pairs.

In [9] we studied the behavior of fuzzy J-K flip-flops based on various fuzzy operations, and illustrated their behavior by the graphs belonging to the next states of fuzzy flip-flops for typical values of Q, J and K.

3.2 Fuzzy D Flip-Flops

Connecting the inputs of the fuzzy J-K flip-flop in a particular way, namely, by applying an inverter in the connection of the input J to K, case of $K = 1-J$, a fuzzy D flip-flop is obtained. Substituting $D = \neg K = J$ in equation (1), the fundamental equation of fuzzy D flip-flop will be

$$Q(t+1) = (D \vee D) \wedge (D \vee Q) \wedge (D \vee \neg Q) \quad (2)$$

As an alternative approach, Choi and Tipnis [4] proposed an equation which also exhibits the characteristics of a fuzzy D flip-flop, as follows

$$Q(t+1) = D \wedge (D \vee Q) \wedge (\neg Q \vee D) \quad (3)$$

We will refer to this type of fuzzy D flip-flop as Choi (type fuzzy) D flip-flop (because of the first author). Comparing the characteristical equation of the fuzzy D flip-flop (2), with expression (3), there is essential difference between the two fuzzy flip-flops.

As $D = D \wedge D$, and $D = D \vee D$ i.e. the t-norm (T) and co-norm (S) are idempotent [1], hold only in the exceptional case ($T(x, x) = x$ and $S(x, x) = x$ for all $x \in [0,1]$); when

T = min, and S = max.

For example, using the algebraic norm

$$u_A(a,a) = a + a - a \cdot a = 2 \cdot a - a^2 = a \quad (4)$$

is satisfied only in the two borderline cases, when $a = 0$, or $a = 1$. It is surprising how much the satisfaction of idempotence influences the behavior of the fuzzy D flip-flops, as it will be shown.

In the next Section we will give an overview of the different type fuzzy J-K, D and Choi D F^{3}s, based on familiar norms listed in Table 1.

3.3 Fuzzy Flip-Flops Based on Various Fuzzy Operations

The behavior of fuzzy flip-flops based on algebraic, Yager, Dombi, Hamacher and Frank t-norms, combined with the standard negation $c(a)$=1-a in every case has been analyzed and compared [9] , in order to investigate, whether and to what degree they present more or less sigmoidal (S-shaped) $J \to Q(t+1)$ transfer characteristics in the particular cases, when K=1-Q, K=1-J, with a fixed value of Q. F^3 based algebraic type t-norm presents non-sigmoidal behavior, with piecewise linear characteristics and several breakpoints, but having the advantage of the hardware implementation of F^3 [14], this type was also studied for comparison.

3.4 Fuzzy J-K, D and Choi D Flip-Flops Based on Some Classes of Fuzzy Set Unions and Intersections

Using a triplet consisting of the standard negation, and a dual pair of fuzzy operations, i.e., using Yager t-norm and s-norm, the maxterm form in the unified equation (1) can be rewritten as:

$$Q_Y(t+1) = \left(J \; u_Y \left(1-K\right)\right) \; i_Y \; \left(J \; u_Y \; Q\right) \; i_Y \; \left(\left(1-K\right) \; u_Y \left(1-Q\right)\right) \quad (5)$$

This is the fundamental equation of the Yager type fuzzy J-K flip-flop.

By substituting the above mentioned triplet into equations (2) and (3) we defined the fundamental equations of the Yager type fuzzy D (6) and Yager type Choi D (7) flip-flops.

$$Q_Y(t+1) = (D \; u_Y \; D) \; i_Y (D \; u_Y \; Q) \; i_Y \; (D \; u_Y \; (1-Q)) \quad (6)$$

$$Q_Y(t+1) = D \; i_Y (D \; u_Y \; Q) \; i_Y \; ((1-Q) \; u_Y \; D) \quad (7)$$

In a very similar way, in the cases of algebraic, Dombi, Hamacher and Frank operation triplets we defined the corresponding fundamental equations of the respective fuzzy J-K, D and Choi D flip-flops.

ISBN: 978-989-95079-6-8

The parameter values in the Yager, Dombi, Hamacher and Frank norms strongly influence the $J \to Q(t+1)$ characteristics curvature. We have compared the characteristics for various typical parameter values and choose w=2, α=2, γ=10 and s=100, where we obtained more or less S-shaped $J \to Q(t+1)$ characteristics. A change of the t-norms in the characteristical equations of fuzzy J-K, D and Choi D flip-flops leads to the modification of the slope of the transfer function, which will affect the learning rate in the implementation of neural networks.

4 Fuzzy flip-flop based neurons

In this section we propose the use of the fuzzy flip-flops discussed above as neurons in a MLP. In the next, we study the effect of applying the mentioned five t-norms and co-norms in the investigation of the F^3 based neurons and the MLPs constructed from them. An important aspect of these F^3s is that they all have a certain convergent behavior when their input J is excited repeatedly. This convergent behavior guarantees the learning property of the networks constructed this way.

4.1 Function Approximation by Fuzzy Flip-Flop Network Trained with Levenberg – Marquardt Algorithm

A fuzzy flip-flop based supervised feedfordward backpropagation network is applied in order to approximate test functions. This function is represented by a set of 100 input/output data sets. All the input and output signals are distributed in the unit interval. In general, two trainable layer networks with sigmoid transfer functions in the hidden layer and linear transfer functions in the output layer have good approximation and interpolation properties [7], analogously the neural system model proposed is based on two hidden layers constituted from fuzzy flip-flop neurons.

The nonlinear characteristics exhibited by fuzzy neurons are represented by quasi sigmoid transfer functions given by fuzzy J-K, D and Choi D flip-flops based on algebraic, Yager, Dombi, Hamacher and Frank operations. The proposed network activation function is the same at each hidden layer, from unit to unit. For simplicity we did not apply activation function or threshold to the output layer, and considered our model with only one output. The number of neurons was chosen after experimenting with different size hidden layers. Smaller neuron numbers in the hidden layer result in worse approximation properties, while increasing the neuron number results in better performance, but longer simulation time. In our approach the weighted input values are connected to input J of the fuzzy flip-flop based on a pair of t-norm and t-conorm, having quasi sigmoid transfer characteristics. The output signal is then computed as the weighted sum of the input signals, transformed by the transfer function.

During network training, the weights and thresholds are first initialized to small, random values and the network was trained with Levenberg-Marquardt algorithm with 100 maximum numbers of epochs as more or less sufficient. First we fixed the activation function, the number of layers and the number of units in each layer. The chosen target activation function was the *tansig* (hyperbolic tangent sigmoid transfer function). This function is well suited to the demands of backpropagation.

5 Bacterial Memetic Algorithm with Modified Operator Execution Order (BMAM)

The Bacterial Memetic Algorithm with Modified Operator Execution Order (BMAM) [5] is a very recent soft computing tool. It combines a special kind of Genetic Algorithm [6] - Pseudo-Bacterial Genetic Algorithm (PBGA) - and the Levenberg-Marquardt (LM) method. We use this algorithm for training fuzzy flip-flop based neural network (FNN) to improve function approximation performance. The Bacterial Memetic Algorithm (BMA) resulted in better approximation properties in fuzzy modeling problems than the Bacterial Evolutionary Algorithm (BEA) (which outperformed the traditional Genetic Algorithms [2], [3]).

The core of BMAM contains the bacterial mutation, which is inspired by the biological bacterial cell model. Its basic idea is to improve the parts of chromosomes contained in each bacterium. The bacterial mutation mechanism is used by the bacteria which can transfer genes to other bacteria. To find the optimal approximation for our network we encoded our FNN weighs, biases, Q and fuzzy operation parameter values in a bacterium (chromosome). Therefore a procedure is working on changing the variables, testing the model obtained in this way and selecting the best models.

5.1 Function Approximation by Fuzzy Flip-Flop Network Trained with BMAM

The learning of the FNN is formulated as a parameter optimization problem, using the mean square error as the fitness evaluation set-up. The basic steps followed by the algorithm embrace the bacterial mutation operation and the LM method.

In the initial population a number of individuals are randomly created and evaluated. Next, an evolutionary cycle is started by applying the bacterial mutation operation for each individual. A number of copies (clones) of the bacterium are generated. Then the same part or parts of the chromosome is choose and mutate randomly, except one single clone that remains unchanged during this mutation cycle. The LM method nested into evolutionary algorithm is applied for a few times for each individual. The selection of the best clone is made and transfers its parts to the other clones. The part choosing-mutation-LM method-selection-transfer cycle is repeated until all the parts are mutated, improved and tested. The best individual is remaining in the population, and all other clones are deleted. This process is repeated until all the individuals have gone through the modified bacterial mutation.

6 Simulation results

The combination of two sine wave forms with different period lengths as test function was,

$y = \sin(c_1 * x) * \sin(c_2 * x)/2+0.5$,

ISBN: 978-989-95079-6-8

where the input vector x generated a sinusoidal output y. The values of constants c_1 and c_2 were selected to produce a frequency proportion of the two components 1:0.35.
The function approximation goodness depended on the flexible Q and fuzzy operation parameter values, on the fuzzy flip-flop types, hidden layer neuron numbers and in addition on the fuzzy operations themselves. To check the goodness of training, we used the Mean Squared Error (MSE) as a measure of the error made by the FNN. The simulation results are summarized in the next two tests.

6.1 Test 1

In one of our earlier paper [9] we optimized the value of Q, for every combination of J-K, D and Choi D type F^3s with all five fuzzy operation pairs, the parameters of the norms chosen for suitable values, solving the problem of the selection of the value of Q with the fine tuning of FNN.
The fuzzy flip-flop based FNNs are based on algebraic, further parametric Yager, Dombi, Hamacher and Frank norms. Table 2 shows the experimental results obtained with LM optimization [9].

Table 2: Q optimums obtained with LM algorithm.

	Fuzzy neuron type		
Fuzzy operation	J-K	D	Choi D
Algebraic (A)	**0.21**	**0.15; 0.91**	**0.09; 0.81**
Yager (Y)	0.06	0.20	0.17
Dombi (D)	**0.11**	0.92	0.06
Hamacher (H)	0.32	0.57	0.38
Frank (F)	0.31	0.45	0.55

The $J \to Q(t+1)$ fuzzy D and Choi D flip-flop characteristics based on algebraic norm present about the same sigmoidal character in multiple points of the domain [7], which fact can leads to rather different optimum results, corresponding to minimal median MSE values. In Test 1 we propose a new method to find the optimal Q – fuzzy operation parameter pair for every combination of J-K, D and Choi D type F^3s by training a 1-8-8-1 FNN with bacterial memetic algorithm. The advantage of this network size lies in good learning capability. In our application a population with 29 parameters - according to the network size - was initialized. During simulations 20 generations of 5 individuals with 5 clones were chosen to obtain the best fitting variables. Table 3 shows the unique optimal Q values found by BMAM algorithm, which are approximately equal to the respective optimum values listed in Table 2. We obtained almost the same results after every run, although in short term the training is not always successful.

Table 3: Q optimums obtained with BMAM algorithm.

	J-K	D	Choi D
Algebraic	0.25	0.12	0.13

The experimental results in case of J-K F^3 type FNN based on Dombi norm further indicated that the proposed network had multiple optimal Q – α parameter pairs. The test revealed, for example, the Q = 0.12, α = 2.9 values as a result that was close to the optimum solution mentioned above.

6.2 Test 2

In the next we will compare the function approximation performance of FNN trained in two different ways. In the first one the network training function updates weight and bias values according to Levenberg-Marquardt optimization (section 4.1). In the second approach the FNN variables - weights, bias, for fixed Q and fuzzy operation parameter values - optimization was made by the bacterial memetic algorithm with the modified operation execution order (section 5.1). During the simulations we covered all 15 possible combinations of fuzzy J-K, D and Choi D type FNNs with all five fuzzy operation pairs to approximate the above mentioned test function.
By changing the number of the layers, we used a 1-4-3-1 FNN in order to have a network that approximated accurately and also fast enough the sine waves, to emphasize the difference between various FNNs.
The parameter of Dombi, Yager, Hamacher and Frank norms were fixed α=2, w=2, γ=10 and s=100, which values provided good learning and convergence properties.
In our approach we fixed the value of Q, the present state value belonging to each fuzzy flip-flop, according to the values in Table 2. Tables 4, 5 and 6 present the 30 runs average approximation goodness, by indicating the minimum, median, mean and maximum mean squared error of the training values for each of the *tansig*, algebraic, Yager, Dombi, Hamacher and Frank types of FNNs.
Figures 1 and 2 present the graphs of the simulations in case of fuzzy J-K flip-flop with feedback based neural network trained with LM and respectively BMAM algorithms. In both of cases the algebraic F^3 provides a fuzzy neuron with rather bad learning ability. Comparing the median MSE values, considering them as the most important indicators of trainability, the Yager and Dombi types FNNs performed best, they can be considered as rather good function approximators. Our simulations have shown that BMAM performs better in every case than the original LM technique.
Figures 3, 4 and 5, 6 compare the behavior of fuzzy D flip-flop and Choi type fuzzy D flip-flop based NNs. It is interesting that according to the numerical illustrations (Tables 5 and 6) the average of 30 run mean squared errors in these cases the best results after the idealistic *tansig* function are given by the Hamacher and Yager F^3, which are followed by the Dombi and finally by the algebraic one. The Hamacher and Yager types FNNs have excellent approximation properties. It is surprising how much the satisfaction of idempotence influences the behavior of the fuzzy D flip-flop based NN. Comparing the simulation results belonging to the two types of fuzzy D flip-flop with the same norms, it can be seen that, for the same value of Q, the value of the MSE differs, which fact leads to a rather different behavior in the applications.
Using BMAM algorithm in the FNN variable optimization task, we obtained better function approximation capability with lower error than in case of FNNs trained with LM method.

ISBN: 978-989-95079-6-8

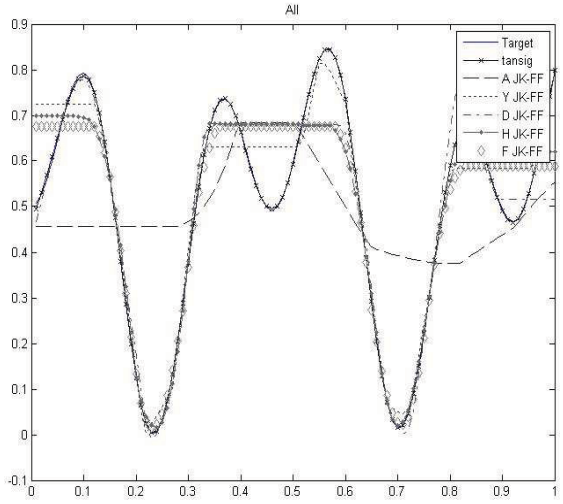


Figure 1: Simulation results of J-K FNN trained with LM algorithm.

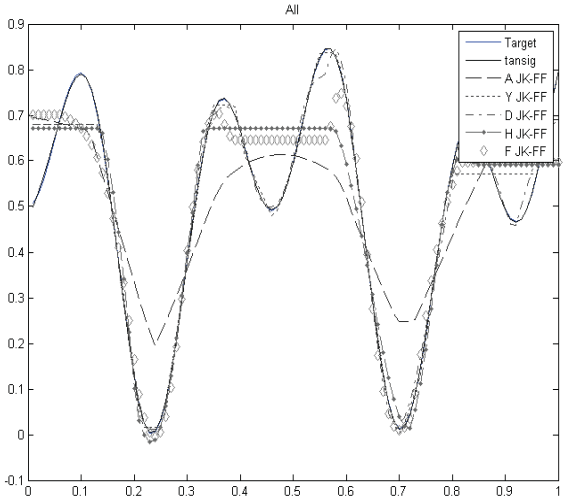


Figure 2: Simulation results of J-K FNN trained with BMAM algorithm.

Table 4: Mean squared error values case of JK FNN.

	LM				BMAM			
	min	median	mean	max	min	median	mean	max
tansig	6.6912×10^{-6}	**0.0056865**	0.0185562	0.0600578	4.8485×10^{-6}	**7.5218×10^{-5}**	0.0006223	0.0041448
A	0.04402231	**0.0548533**	0.0555179	0.0818562	0.00888158	**0.0170071**	0.0195847	0.0368526
Y	0.00286463	**0.0534354**	0.0474332	0.0615071	0.00133805	**0.0031207**	0.0310288	0.0064944
D	0.00420610	**0.0512862**	0.0439112	0.0691943	9.2216×10^{-5}	**0.0022534**	0.0024855	0.0052513
H	0.00551096	**0.0385112**	0.0400807	0.0720822	0.00351781	**0.0071140**	0.0081815	0.0256592
F	0.05552121	**0.0485408**	0.0425236	0.0660629	0.00228814	**0.0066725**	0.0076724	0.0312439

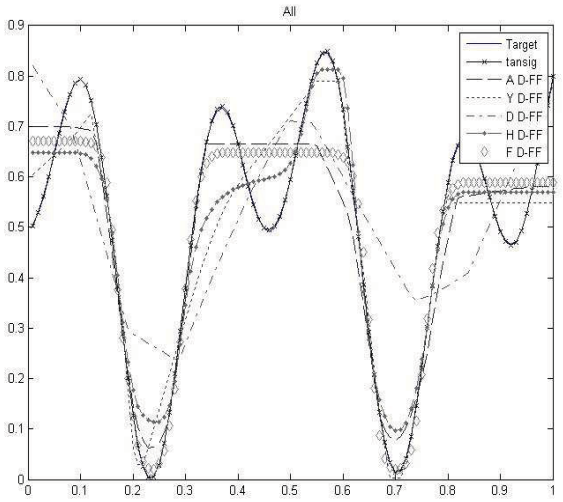


Figure 3: Simulation results of D FNN trained with LM algorithm.

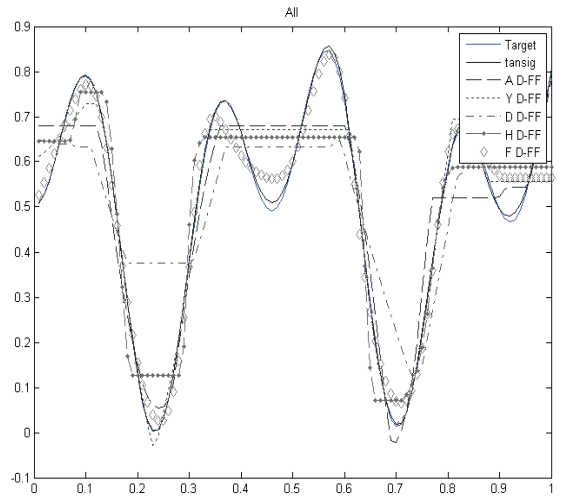


Figure 4: Simulation results of D FNN trained with BMAM algorithm.

Table 5: Mean squared error values case of D FNN.

	LM				BMAM			
	min	median	mean	max	min	median	mean	max
tansig	4.3611×10^{-6}	**0.0068415**	0.0203523	0.0565211	9.6899×10^{-7}	**4.7893×10^{-5}**	0.0006056	0.0033605
A	0.00771386	**0.0531832**	0.0468789	0.0838021	0.00490354	**0.00918654**	0.0114988	0.0330397
Y	0.00562192	**0.0515206**	0.0448304	0.0653909	0.00031811	**0.00335034**	0.0033998	0.0065352
D	0.03089988	**0.0502911**	0.0502276	0.0692478	0.01059031	**0.02699423**	0.0255351	0.0365076
H	0.00511140	**0.0485498**	0.0425834	0.0679154	0.00017240	**0.00586916**	0.0073385	0.0271431
F	0.00694345	**0.0506654**	0.0458813	0.0640876	0.00051204	**0.00681308**	0.0065404	0.0229583

ISBN: 978-989-95079-6-8

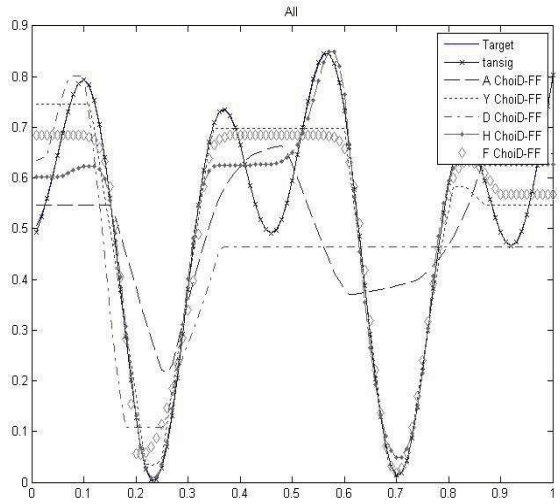


Figure 5: Simulation results of Choi D FNN trained with LM algorithm.

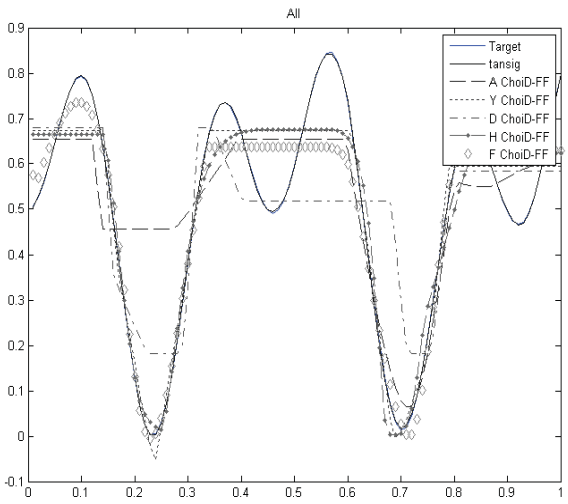


Figure 6: Simulation results of Choi D FNN trained with BMAM algorithm.

Table 6: Mean squared error values case of Choi D FNN.

	LM				BMAM			
	min	median	mean	max	min	median	mean	max
tansig	1.1454×10^{-6}	**0.0291153**	0.0251398	0.0624285	1.9737×10^{-7}	$\mathbf{2.7722 \times 10^{-5}}$	0.0005026	0.0024156
A	0.03693712	**0.0567222**	0.0557563	0.0878214	0.00777224	**0.0265908**	0.0251665	0.0363809
Y	0.00660108	**0.0508917**	0.0483636	0.0641307	0.00328527	**0.0077983**	0.0087765	0.0254508
D	0.03222194	**0.0534598**	0.0537578	0.0861916	0.01436651	**0.0235421**	0.0247559	0.0347627
H	0.00459423	**0.0497143**	0.0446552	0.0737876	0.00546122	**0.0077690**	0.0111944	0.0286523
F	0.00668745	**0.0518677**	0.0445716	0.0727711	0.00293771	**0.0092323**	0.0126012	0.0310469

7 Conclusions

In this paper we proposed the use of BMAM to optimizing fuzzy flip-flop based neural network (FNN). We compared the function approximation performance of five different types of FNNs, which depends from the choice of different fuzzy flip-flop types and from the training algorithm. The result of learning as well as its performance might differ quite significant under these two learning modes. The simulations have shown that the bacterial memetic algorithm can improve the function approximation capability by optimizing the FNN variables values.

Acknowledgment

This paper was supported by the Széchenyi University Main Research Direction Grant 2009, National Scientific Research Fund Grant OTKA T048832 and K75711, and National Office for Research and Technology.

References

[1] R. Bellman and M. Giertz, "On the analytic formalism of the theory of fuzzy sets", Information Science, vol. 5, 1973, pp. 149-156.

[2] J. Botzheim, C. Cabrita, L. T. Kóczy, A. E. Ruano, **"**Fuzzy rule extraction by bacterial memetic algorithms" International Journal of Intelligent Systems, 2009, pp. 312-339.

[3] J. Botzheim, C. Cabrita, L. T. Kóczy, A. E. Ruano, "Genetic and bacterial programming for B-spline neural networks design", Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol. 11., No. 2, 2007, pp. 220-231.

[4] B. Choi and K. Tipnis, "New Components for Building Fuzzy Logic Circuits", Proc. of the 4th Int. Conf. on Fuzzy Systems and Knowledge Discovery, vol. 2, 2007, pp. 586-590.

[5] L. Gál, J. Botzheim and L. T. Kóczy, „Improvements to the Bacterial Memetic Algorithm used for Fuzzy Rule Base Extraction", Computational Intelligence for Measurement Systems and Applications, CIMSA 2008, Istanbul, pp.38-43.

[6] J. H. Holland, "Adaptation in Nature and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence", The MIT Press Cambridge, 1992.

[7] K. M. Hornik, M. Stinchcombe and H. White, "Multilayer feedforward networks are universal approximators", Neural Networks, vol.2, no.5, 1989, pp. 359-366.

[8] E. P. Klement, R. Mesiar and E. Pap, "Triangular Norms", Series: Trends in Logic , vol.8, 2000.

[9] R. Lovassy, L. T. Kóczy and L. Gál, "Function Approximation Capability of a Novel Fuzzy Flip-Flop Based Neural Network", IJCNN 2009 Atlanta- accepted.

[10] R. Lovassy, L. T. Kóczy and L. Gál, "Analyzing Fuzzy Flip-Flops Based on Various Fuzzy Operations", Acta Technica Jaurinensis Series Intelligentia Computatorica vol. 1, no. 3, 2008, pp. 447-465.

[11] R. Lovassy, L. T. Kóczy and L. Gál, "Multilayer Perceptron Implemented by Fuzzy Flip-Flops", IEEE World Congress on Computational Intelligence, WCCI 2008, Hong Kong, pp. 1683-1688.

[12] S. Osowski, K. Siwek and T. Markiewicz, "MLP and SVM Networks – a Comparative Study", Proceedings of the 6th Nordic Signal Processing Symposium, Finland, 2004, pp. 37-40.

[13] K. Ozawa, K. Hirota and L. T. Kóczy, "Fuzzy flip-flop", In: M. J. Patyra, D. M. Mlynek, eds., Fuzzy Logic. Implementation and Applications, Wiley, Chichester, 1996, pp. 197-236.

[14] K. Ozawa, K. Hirota, L. T. Kóczy and K. Omori, "Algebraic fuzzy flip-flop circuits", Fuzzy Sets and Systems 39/2, North Holland, 1991, pp.215-226.

ISBN: 978-989-95079-6-8

Bi-criteria Genetic Selection of Bagging Fuzzy Rule-based Multiclassification Systems

Krzysztof Trawiński, Arnaud Quirin, Oscar Cordón

European Centre for Soft Computing. Mieres, Spain
Email: {krzysztof.trawinski, arnaud.quirin, oscar.cordon}@softcomputing.es

Abstract— *Previously we proposed a scheme to generate fuzzy rule-based multiclassification systems by means of bagging, mutual information-based feature selection, and a multicriteria genetic algorithm (GA) for static component classifier selection guided by the ensemble training error. In the current contribution we extend the latter component by the use of two bi-criteria fitness functions, combining the latter error measure with the selected ensemble likelihood. A study on four popular UCI datasets with different dimensionalities is conducted in order to analyze the accuracy-complexity trade-off obtained by the two GAs, the initial fuzzy ensemble and a single fuzzy classifier.*

Keywords— Bagging, feature selection, fuzzy rule-based multiclassification systems, genetic selection of individual classifiers, multicriteria genetic algorithm.

1 Introduction

Multiclassification systems (MCSs) are promising data mining tools dealing with complex classification problems, especially when the number of dimensions or the size of the data are really large [1]. They usually combine decision trees [2] or neural networks [3], but also more recently fuzzy classifiers [4, 5, 6].

In a previous study [7], we described a methodology in which classical MCS design approaches such as bagging [8] and random subspace [2] are used to generate fuzzy rule-based multiclassification systems (FRBMCSs) using a basic heuristic fuzzy classification rule generation method [9], as well as a classifier selection technique based on a GA driven by a multicriteria fitness function [10]. Later, we improved our methodology in [10] using a feature selection approach based on the Battiti's method [11] and the GRASP procedure [12]. Finally, in [13], we extended the genetic selection method by considering additional error measures.

We drew the conclusion that a feature and instance selection procedure combined with a simple grid partitioning FRBS is a good approach to overcome the curse of dimensionality problem in large datasets while using FRBMCS, mainly due to the fact that these kind of classifiers are instable enough. Nevertheless, once a set of classifiers has been trained, we still need to deal with the high number of rules and the correlations between individual classifiers. This is why a selection of the classifiers is so crucial. As said before, we already proposed a GA guided by several single-criteria fitness functions, based on the training error [10], the likelihood [7], or the Out-Of-Bag error [13]. This methodology, quite novel in this topic, lead us to the generation of a compact sets of rules, while still preserving its accuracy, in a single GA run, without resorting on a Pareto-based multi-objective optimization technique. However, many experimentations suggested the choice of the fitness function is very dependent of the problem being solved. For instance, when using the training error, the accuracies of two FRBMCSs can be similar or even null, making difficult for the GA to discriminate between them in order to improve the generalization ability. On the contrary, using the likelihood alone seems to give bad results on many datasets. This suggests the fact that a combination of some criteria could be a good idea to overcome this issue, producing better results than any criterion in isolation.

The aim of the current contribution is to propose a solution by exploring two new fitness functions based on a combination of the training error and the likelihood measures. By doing so, we will try to combine them using the two most simple ways: weighted average and lexicographic order (i.e. considering the optimization of a single criterion, and using the second in case of tie). Introducing such elaborated method we hope that it will allow the FRBMCS to deal with high dimensional data.

We aim to check if the new GA fitness functions will perform better in terms of accuracy than the previous ones for some datasets, while still being competitive for the others datasets. A preliminary study will be conducted on small and medium size datasets from the UCI machine learning repository to test the two new fitness functions in comparison to a single classifier, the original FRBMCS, and the GA-selected FRBMCSs using the said fitness functions. Several parameter settings for the global approach (e.g. different granularity levels as well as different feature selection methods) will be tested and compared regarding the accuracy and the size of the rule base obtained by a single classifier and the original FRBCS ensemble.

This paper is set up as follows. In the next section, existing GA-based methods to select MCSs are reviewed. Sec. 3 recalls our approach for designing FRBMCSs considering bagging and feature selection, while Sec. 4 describes the proposed multicriteria GA for component classifier selection. The experiments developed and their analysis are shown in Sec. 5. Finally, Sec. 6 collects some concluding remarks and future research lines.

2 Related work on genetic selection of MCSs

In general, the selection of a subset of classifiers is done using the *overproduce-and-choose strategy* (OCS) [14], in which a large set of classifiers is produced and then selected to extract the best performing subset. GAs are a popular technique within this strategy. In the literature, performance, complexity and diversity measures are usually considered as search criteria. Complexity measures are employed to increase the interpretability of the system whereas diversity measures are used to avoid overfitting.

ISBN: 978-989-95079-6-8

Among the different genetic OCS proposals, we can remark the following ones. In [15], a hierarchical multi-objective GA (MOGA) algorithm, performing feature selection at the first level and classifier selection at the second level, is presented which outperforms classical methods for two handwritten recognition problems. The MOGA allows both performance and diversity to be considered for MCS selection. In [16] a GA is used to select from seven diversity heuristics for k-means cluster-based ensembles and the ensemble size is also encoded in the genome. In the study of Martínez-Munoz et al. [17], a GA is compared to five other techniques for ensemble selection. Even if the performance of the GA was the worst obtained, they showed that while selecting a small subset of classifiers, the generalization error was significantly decreased. In [18], the authors developed a multidimensional GA to optimize two weight-based models, in which the weights are assigned to each classifier or to each class. They applied their system to 6 different classifiers (only linear and quadratic classifiers are explored), but on only two small datasets and without comparing to the results obtained on a single classifier. Finally, our own previous studies [10, 7] also consider a multicriteria GA for the ensemble selection in an OCS fashion, with performance (training error) and complexity as criteria to guide the GA.

In general, the performance obtained with the initial MCS is outperformed by the ensemble selected by the GA, while simplifying the system. In our current contribution, we will confirm this conclusion by the study of two improved fitness functions mixing the two most used criteria: the accuracy and the complexity of the classifiers. The fitness function will directly incorporate one or two accuracy criteria (i.e., the training error and the likelihood), while the MCS complexity will be implicitly optimized by the considered coding scheme.

3 Bagging and feature selection-based FRBMCSs

In this section we will both detail how the individual classifiers and the FRBMCSs are designed. A normalized dataset is split into two parts, a training set and a test set. The training set is submitted to an instance selection and a feature selection procedure in order to provide individual training sets (the so-called *bags*) to train simple FRBCSs (through the method described in section 3.1). The instance selection and the feature selection procedures are described in section 3.2. After the training, we got an initial FRBMCS, which is validated using the training and the test errors (*Ensemble Training Error* and *Ensemble Test Error*), as well as a measure of complexity based on the total number of rules in the FRBCSs. This ensemble is selected using a multicriteria GA (described in the next section) guided by two accuracy-based fitness functions. The final FRBMCS is validated using different accuracy (Training Error, Test Error) and complexity measures (number of classifiers, total number of rules).

3.1 Individual FRBCS composition and design method

The FRBCSs considered in the ensemble will be based on fuzzy rules R_j with a class C_j and a certainty degree CF_j in the consequent: If x_1 is A_{j1} and ... and x_n is A_{jn} then Class C_j with CF_j, $j = 1, 2, \ldots, N$, and they will take their decisions by means of the single-winner method [9]. This fuzzy reasoning method has been selected due to its high simplicity and interpretability. The use of other more advanced ones [19] is left for future works.

To derive the fuzzy knowledge bases, one of the heuristic methods proposed by Ishibuchi et al. in [9] is considered. The consequent class C_j and certainty degree CF_j are statistically computed from all the examples located in a specific fuzzy subspace $D(A_j)$. C_j is computed as the class h with maximum confidence according to the rule compatible training examples $D(A_j) = \{x_1, \ldots, x_m\}$: $c(A_j \Rightarrow Class\ h)$ = $|D(A_j) \bigcap D(Class\ h)|/|D(A_j)|$. CF_j is obtained as the difference between the confidence of the consequent class and the sum of the confidences of the remainder (called CF_j^{IV} in [9]).

This method is good for our aim of designing FRBCS ensembles since it is simple and quick. However, it carries two drawbacks: its low accuracy and the generation of large fuzzy rule bases. We aim to consider more advanced techniques in the future.

3.2 FRBMCS design approaches

The generation of the FRBMCSs is performed by means of a bagging approach combined with a feature selection method [10]. Three different feature selection methods, random subspace and two variants of Battiti's MIFS, greedy and GRASP, are considered.

As said before, *random subspace* [2] is a method in which we select randomly a set of features from the original dataset. The greedy Battiti's MIFS method [11] is based on a forward greedy search using the Mutual Information measure [20], with regard to the class. This method selects the set S of the most informative features about the output class which cannot be predicted with the already selected features. It uses a coefficient, β, to set up the penalization on the information brought by the already selected features.

The MIFS-GRASP variant is an approach where the set is generated by iteratively adding features randomly chosen from a Restricted Candidate List (RCL) composed of the best τ percent decisions according to the Battiti's quality measure. Parameter τ is used to control the amount of randomness injected in the MIFS selection. With $\tau = 0.5$, we get an average amount of randomness, while still preserving the quality-based ordering of the features.

For the bagging approach, the bags are generated with the same size as the original training set, as commonly done. In every case, all the classifiers will consider the same fixed number of features.

Finally, no weights will be considered to combine the outputs of the component classifiers to take the final FRBMCS decision, but a pure voting approach will be applied: the ensemble class prediction will directly be the most voted class in the component classifiers output set.

4 A multicriteria GA-based MCS selection method

In our previous studies, we used a multicriteria GA, which is able to obtain a list of possible MCS designs ranked by their quality *from a single chromosome* thanks to its novel coding scheme. However, the fitness function considered was based on a single criterion, either the likelihood (L) [7], the *training*

ISBN: 978-989-95079-6-8

error (TE) [10, 13], or the out-of-bag error [13]. Although the TE-based GA provided better overall results, the L-based GA outperformed it in some of the cases. That led us to the idea of combining the both measures, which moreover show complementary characteristics.

4.1 Multicriteria genetic optimization

The GA searches for an optimal sequence of the classifiers, in the way that the most significant classifiers have the lowest indexes, while those redundant members, which can be safely excluded, are in the last positions. The coding scheme is thus based on an order-based representation, a permutation $\Pi = \{j_1, j_2, \ldots, j_l\}$ of the l originally generated individual classifiers. In this way, each chromosome encodes l different solutions to the problem, based on considering a "basic" MCS comprised by a single classifier, that one stored in the first gene, then another one composed of two classifiers, those in the first and the second genes, and so on.

So, the computation of the evaluation criteria for the whole ensemble is obtained in a *cumulative* way, defined as a vector containing the measured values of the first classifier; the subset formed by the first and the second; and so on. The fitness function is thus using the values of a multicriteria vector, being composed of an array of l values, $L^i = L'_{\{j_1,j_2,\ldots,j_i\}}$, corresponding to the cumulative measure-value of the l mentioned MCS designs.

At the end of the GA, the best chromosome is that member in the population overcoming the others using the considered criterion. Then, the final design encoded in this chromosome is the MCS comprising the classifiers from the first one to the one having the the best cumulative measured value (although any other design not having the optimal accuracy but, for example, showing a lowest complexity can also be directly extracted). In this way, an implicit use of a complexity criterion is also made.

To increase its convergence rate, the GA works following a steady-state approach. The initial population is composed of randomly generated permutations. In each generation, a tournament selection of size 3 is performed, and the two winners are crossed over to obtain a single offspring that directly substitutes the loser. In this study, we have considered OX crossover and the usual exchange mutation [21].

4.2 The two used evaluation criteria

For the definition of the fitness functions, we use TE and L as the evaluation criteria.

The TE is computed as follows. Let $h_1(\mathbf{x}), \ldots, h_l(\mathbf{x})$ be the outputs of the component classifiers of the selected ensemble for an input value $\mathbf{x} = (x_1, \ldots, x_n)$. For a given sample $\{(\mathbf{x}^k, C^k)\}_{k\in\{1\ldots m\}}$, the TE of that MCS is:

$$TE = \frac{1}{m} \cdot \#\{k \mid C^k \neq \arg\max_{j\in\{1\ldots M\}} h_j(\mathbf{x}^k)\} \qquad (1)$$

Fitness evaluation using TE alone was already studied in one of our previous publications [10]. We will call it *Training Error-based Fitness Function* (TEFF).

The L is computed as follows. Let the classes $h_1(\mathbf{x}), h_2(\mathbf{x}), \ldots, h_l(\mathbf{x})$ be the decisions of the component classifiers of the selected ensemble S for an input value $\mathbf{x} = (x_1, \ldots, x_n)$. We will assume that the fraction of the members of S that agree on the class of $\mathbf{x}$ is an estimate of the conditional probability of that class:

$$P_S(C|\mathbf{x}) = \frac{1}{|S|} \cdot \#\{i \in S \mid h_i(\mathbf{x}) = C\}.$$

The L of the subset S, to be maximized, is:

$$L_S = \prod_k P_S(C^k|\mathbf{x}^k).$$

As the small values of L_S may produce numerical instabilities, we use instead a bounded log-likelihood:

$$L'_S = \sum_k \log(P_S(C^k|\mathbf{x}^k) + \epsilon),$$

where the value ϵ foresees that case for which none of the members of the subset has found the true class of the pattern. In [7], we endowed the fitness function with the L, as it allows us to discern differences between ensembles with the same TE (specifically, between those with null error!). A learning process using only the TEFF will automatically end up with the learning, while L will go on improving the estimations of the probability distributions for each class, thus reducing the chances of overfitting the training data.

4.3 The two new bi-criteria fi tness functions

In this contribution we propose two approaches for the fitness function combining the L with the TE measure, the *Lexicographical Order-based Fitness Function* (LOFF) and the *Weighted Combination Fitness Function* (WCFF).

Notice that, working in this way, we are introducing a second multi-criteria optimization level in our algorithm. On the one hand, a multi-criteria optimization is made by means of the considered coding scheme and the cumulative evaluation of the possible MCS designs (see Sect. 4.1). On the other hand, a higher level is introduced when evaluating the latter possible designs by means of a bi-criteria fitness function.

In the first one, the LOFF, we use the lexicographical order to deal with the multicriteria optimization. When comparing two chromosomes, one is better than the other if it takes a better (lower) minimum value of the TE. In case of tie, the L measure is considered. The ordering scheme gives priority to TE, as it provided better results in our previous study, while taking the L only in the last resort in the case of the frequent ties encountered by the system.

In the second approach, the WCFF, we propose objective function scalarization by a weighted combination of both measures:

$$WC = factor_0 * \alpha * TE + (1 - \alpha) * L \qquad (2)$$

where α is a weight in [0,1] and $factor_0 = L_0/TE_0$ is a first evaluation-based normalization using L_0 and TE_0, the L and the TE from the initial FRBMCS. The fitness function is to be minimized.

5 Experiments and analysis of results

In this section, we discuss the performance obtained by a single FRBCS, the initial FRBMCS, and three different GA-selected FRBMCSs including our two new fitness functions on four chosen datasets.

ISBN: 978-989-95079-6-8

5.1 Experimental setup

To evaluate the performance of the generated FRBMCSs, we have selected four datasets from the UCI machine learning repository (see Table 1). In order to compare the accuracy of the considered classifiers, we used Dietterichs 5×2-fold cross-validation (5×2-cv), which is considered to be superior to paired k-fold cross validation in classification problems [22].

Table 1: Data sets considered

Data set	#attr.	#examples	#classes
Pima	8	768	2
Glass	9	214	7
Vehicle	18	846	4
Sonar	60	208	2

Three different granularities, 3, 5 and 7, are tested for the single FRBCS derivation method, for feature sets of size 5 selected by means of three approaches: the greedy Battiti's MIFS filter feature selection method [11], the Battiti's method with GRASP (with τ equal to 0.5, see section 3.2), and random subspace [2]. Battiti's method has been run by considering a discretization of the real-valued attribute domains in ten parts and setting the β coefficient to 0.1.

The FRBMCSs generated are initially comprised by 50 classifiers. The GA for the component classifier selection works with a population of 50 individuals and runs during 50 generations. The mutation probability considered is 0.05. The weights of WCFF were set to 0.8 for TE and 0.2 for L as our aim was to allow a small influence of the L in the cases in which the TE gives similar values. The other values for the weights will not improve the results significantly.

All the experiments have been run in a cluster at the University of Granada on Intel quadri-core Pentium 2.4 GHz nodes with 2 GBytes of memory, under the Linux operating system.

5.2 Comparison of the three fi tness functions

The statistics (5×2-cv error, number of rules, and run time required for each run, expressed in seconds) for the genetically selected FRBCS ensembles using LOFF, WCFF and TEFF are collected in Tables 2, 3 and 4 respectively. There are three subtables for each of the feature selection method considered. The best results for a given feature selection method are shown in bold and the best values overall are outlined.

Comparing the three fitness functions, we can see how the WCFF approach is able to outperform the TEFF and LOFF considering the individual test error 9 times (+2 draw). The best individual improvement was observed on the sonar dataset (-8% regarding LOFF, -4% regarding TEFF) with greedy and 5 labels. We observed that in 4 out of 9 cases, WCFF outperforms other approaches on the sonar and pima datasets. The best overall result was obtained on the pima dataset with GRASP and 5 labels (draw with LOFF).

The FRBMCSs based on LOFF are better than TEFF and WCFF in 12 of the 36 cases (+2 draw). The best individual improvement was observed on the sonar dataset with random subspace and 5 labels (-9% regarding WCFF, -12% regarding TEFF). The best overall result was obtained on the pima dataset with greedy and 5 labels (draw with WCFF) and on the glass dataset with GRASP and 5 labels. We can observe that in 5 out of 9 cases, LOFF outperforms the other genetic approaches on the glass dataset.

Comparing the two new fitness functions, the LOFF provides better results than WCFF considering individual test error in 21 cases (+ 2 draw). We can also observe that LOFF outperforms the other approach on the glass dataset in 7 out 9 cases (+1 draw) giving a good performance for the vehicle (7 out 9 cases) datasets. However, the WCFF is better in 6 out of 9 cases on the pima dataset and in 5 out of 9 cases (+1 draw) on the sonar dataset. In general, LOFF is better than WCFF since it gives a lower influence to L, whereas the weighted combination is better on the sonar dataset due to the many ties with the TE on this dataset, making the use of the L more likely.

The TEFF-based FRBMCSs outperforms the LOFF and the WCFF considering individual test error for 13 of the 36 times. The best overall result was obtained on the sonar dataset with random subspace and 3 labels, and on the vehicle with random subspace and 7 labels. We may conclude that the LOFF and WTEL are competitive with TEFF. The LOFF got best results 12 times (+2 draw), WTEL 9 (+2 draw) and TEFF 13 times. However, in the direct comparison, the use of LOFF improves the single TEFF performance in 18 out of 36 cases (+3 draw) and the WTEL improves the single TEFF performance in 19 out of 36 cases, which indicates that L does not produce overfitting. Thus, the use of L as a secondary criterion is only useful in some of the cases.

On 36 cases, the number of classifiers is lower in 12 cases (+1 draw) with the LOFF, in 9 cases with the WCFF and in 15 cases (+1 draw) with the TEFF. The two new fitness functions generate a higher number of classifier, since they are more conservative due to the use of L. Such fitness functions could be viewed as a proper way to improve performance of the datasets having the larger size.

5.3 Genetically selected FRBMCSs vs. single FRBCS/original FRBMCSs

The results of the single FRBCSs are presented in Table 5 while those of the original FRBMCSs are presented in Table 6. In all the 36 cases, the generated FRBMCSs improve the performance of the single FRBCS.

Although the main goal of the genetic selection is to reduce the complexity of the generated FRBMCS, the accuracy results obtained from that process are also improved in most of the cases, showing the potential of the approach. In only 10 of the 36 cases (+1 draw) the original FRBMCS outperforms the best genetically designed one in terms of accuracy. Comparing the best overall TE values of genetically selected FRBMCSs with those of the original FRBMCSs, the GA improves the results on the sonar dataset (-2% regarding TEFF) and glass (-10% regarding LOFF), showing a slight increase for the other problems (+2% for pima regarding LOFF/WTEL, +2% for vehicle regarding TEFF).

5.4 Statistical signifi cance of the results

Table 7 shows the results of the statistical tests performed to check if the performance of the initial FRBMCSs and the performance of the GA selected FRBMCSs outperform significantly the performance of the single classifier. The Wilcoxon signed-rank test [23] has been used for this purpose. The best

ISBN: 978-989-95079-6-8

Table 2: Results for the FRBCS ensembles selected by the GA using the LOFF

		Bagging + Greedy				Bagging + Random Subspace				Bagging + GRASP $\tau = 0.50$			
		Pima	Glass	Vehicle	Sonar	Pima	Glass	Vehicle	Sonar	Pima	Glass	Vehicle	Sonar
	5×2-cv	0.257	**0.363**	0.468	**0.246**	0.261	**0.375**	0.423	**0.219**	0.253	0.367	0.440	**0.227**
3 labels	#classifiers	3.8	10.1	9.5	8.5	4.0	13.0	14.2	19.9	4.2	9.1	11.7	12.9
5 #attr.	#rules	659.7	1275.9	1298.6	1287.2	668.4	1450.0	2336.3	3363.0	742.4	1162.3	1795.1	2110.9
	avg. #rules	176.6	126.5	138.3	151.0	167.6	112.5	165.3	169.9	175.2	125.9	152.0	163.6
	time	599.57	127.00	665.84	164.46	436.40	111.97	482.74	117.36	534.17	134.80	591.25	144.59
	5×2-cv	**0.238**	0.374	0.391	0.259	**0.257**	0.387	0.374	0.223	**0.240**	**0.358**	0.399	0.234
5 labels	#classifiers	12.9	8.5	14.0	10.1	12.9	14.2	14.2	14.1	10.9	16.0	13.4	13.0
5 #attr.	#rules	7491.1	2211.3	6711.9	5696.9	7160.0	3439.0	10374.3	9273.3	6535.1	4190.4	7610.7	8291.5
	avg. #rules	582.0	268.1	482.1	569.4	550.8	247.5	740.3	680.7	599.5	271.7	582.1	644.4
	time	434.14	121.72	487.14	117.84	599.22	167.95	665.79	162.04	370.50	102.98	405.34	100.62
	5×2-cv	0.258	0.387	**0.375**	0.258	0.270	0.393	**0.347**	0.275	0.258	0.375	**0.355**	0.250
7 labels	#classifiers	14.4	9.5	14.0	7.3	15.8	13.4	17.8	6.2	13.3	9.1	15.5	8.5
5 #attr.	#rules	18851.1	3786.4	16266.7	7609.4	19619.0	5169.0	29596.5	7690.1	17892.8	3738.7	20978.9	10210.4
	avg. #rules	1308.1	400.2	1146.7	1056.8	1240.5	392.6	1686.0	1234.8	1338.9	409.5	1386.4	1214.6
	time	369.73	103.94	405.34	100.18	366.53	103.25	407.70	99.94	498.07	140.94	552.16	135.59

Table 3: Results for the FRBCS ensembles selected by the GA using the WCFF

		Bagging + Greedy				Bagging + Random Subspace				Bagging + GRASP $\tau = 0.50$			
		Pima	Glass	Vehicle	Sonar	Pima	Glass	Vehicle	Sonar	Pima	Glass	Vehicle	Sonar
	5×2-cv	0.254	**0.364**	0.488	0.256	0.258	0.384	0.430	**0.224**	0.255	**0.370**	0.459	**0.226**
3 labels	#classifiers	6.2	8.8	15.6	6.8	6.3	10.5	17.7	22.2	7.5	13.9	10.3	10.4
5 #attr.	#rules	1081.6	1139.8	2209.2	1030.0	1020.8	1190.4	2944.8	3683.5	1331.4	1722.5	1587.2	1677.3
	avg. #rules	174.3	125.4	140.2	153.2	161.2	113.1	166.6	171.0	172.4	126.3	153.3	167.6
	time	401.63	111.35	441.11	108.23	382.65	105.19	419.06	102.40	392.34	108.28	428.73	105.53
	5×2-cv	**0.243**	0.380	0.395	**0.238**	**0.256**	**0.382**	0.372	0.244	**0.238**	0.386	0.401	0.234
5 labels	#classifiers	17.4	12.9	14.7	14.9	16.3	9.2	16.0	19.1	14.6	13.0	15.4	22.4
5 #attr.	#rules	10118.3	3438.4	7258.3	8202.4	9104.2	2347.5	11739.3	12502.4	8686.7	3457.0	8947.0	13497.0
	avg. #rules	589.8	264.5	486.4	566.8	550.0	260.3	737.4	674.3	593.8	270.3	593.6	622.2
	time	383.37	106.09	420.65	102.93	401.48	111.36	441.42	108.02	371.67	102.42	411.51	100.52
	5×2-cv	0.250	0.402	**0.368**	0.256	0.272	0.393	**0.353**	0.275	0.254	0.383	**0.363**	0.243
7 labels	#classifiers	13.8	13.4	12.6	8.0	16.2	15.1	10.3	6.2	14.8	11.7	12.1	9.7
5 #attr.	#rules	17992.3	5231.7	14853.7	8248.2	20343.7	5923.8	18443.5	7690.1	19726.7	4728.6	17199.3	11271.4
	avg. #rules	1323.8	399.7	1164.6	1053.8	1258.5	411.1	1814.7	1234.8	1343.0	402.2	1435.8	1173.8
	time	372.49	102.26	407.72	100.36	368.65	103.92	408.42	99.82	386.31	107.54	425.35	104.44

Table 4: Results for the FRBCS ensembles selected by the GA using the TEFF

		Bagging+Greedy				Bagging+GRASP $\tau = 0.50$				Bagging + Random Subspace			
		Pima	Glass	Vehicle	Sonar	Pima	Glass	Vehicle	Sonar	Pima	Glass	Vehicle	Sonar
	5×2-cv	0.257	**0.360**	0.461	**0.235**	0.254	0.372	0.449	**0.237**	**0.256**	**0.381**	0.428	**0.216**
3 labels	#classifiers	4.1	7.3	10.3	12.3	14.4	10.2	12.9	13.9	4.2	13.7	13.4	20.1
5 #attr.	#rules	696.5	904.3	1431.0	1842.1	763.0	1317.9	1991.6	2252.6	703.4	1546.0	2239.5	3376.7
	avg. #rules	171.5	125.4	138.3	148.3	174.3	126.0	155.9	161.7	168.1	113.1	168.9	168.3
	time	94.06	26.35	103.26	25.32	93.37	26.49	102.09	25.18	92.77	26.39	103.24	25.08
	5×2-cv	**0.242**	0.383	0.392	0.247	**0.239**	**0.363**	0.399	0.252	0.263	0.392	0.378	0.249
5 labels	#classifiers	11.5	15.9	15.5	10.4	10.9	14.7	12.0	7.8	11.9	13.7	13.0	9.4
5 #attr.	#rules	6744.9	4233.1	7338.4	5757.7	6497.4	3986.7	7227.3	4893.9	6680.0	3312.2	9455.9	6208.8
	avg. #rules	592.8	268.7	481.9	567.0	593.5	282.0	611.3	630.0	555.8	245.0	734.3	668.8
	time	93.48	26.10	103.48	25.17	92.58	26.16	103.75	24.86	91.47	26.18	104.81	24.83
	5×2-cv	0.258	0.393	**0.374**	0.258	0.256	0.395	**0.356**	0.257	0.265	0.393	**0.337**	0.267
7 labels	#classifiers	12.7	8.9	14.6	6.3	16.4	10.3	13.2	6.7	17.0	15.5	17.5	6.4
5 #attr.	#rules	16614.3	3524.3	16102.3	6427.0	21836.6	4140.6	18296.2	7767.8	21289.5	5980.6	28854.2	7655.2
	avg. #rules	1313.9	404.5	1115.7	1040.9	1346.2	401.9	1386.5	1148.7	1248.4	386.2	1680.2	1203.7
	time	92.87	26.50	102.90	24.85	92.49	26.18	102.93	25.31	92.31	26.08	103.52	25.19

Table 5: Results for the single FRBCSs with feature selection

		Greedy				GRASP $\tau = 0.50$				Random Subspace			
		Pima	Glass	Vehicle	Sonar	Pima	Glass	Vehicle	Sonar	Pima	Glass	Vehicle	Sonar
3 labels	5×2-cv	0.266	0.446	0.549	**0.261**	0.267	0.447	0.546	0.316	0.265	0.457	0.512	**0.319**
5 #attr.	#rules	178.50	135.30	136.40	146.60	179.50	137.00	135.80	169.00	161.80	109.50	154.50	174.50
	time	0.08	0.04	0.12	0.08	0.09	0.04	0.12	0.09	0.07	0.03	0.12	0.08
5 labels	5×2-cv	**0.246**	**0.376**	0.430	0.287	**0.246**	**0.375**	0.425	**0.314**	**0.262**	0.435	0.460	0.329
5 #attr.	#rules	682.70	291.00	437.60	615.20	682.70	293.50	418.90	752.70	604.20	259.60	587.80	773.60
	time	0.42	0.25	0.65	0.16	0.39	0.26	0.63	0.17	0.36	0.24	0.67	0.17
7 labels	5×2-cv	0.262	0.414	**0.402**	0.291	0.266	0.423	**0.399**	0.317	0.276	**0.418**	**0.415**	0.340
5 #attr.	#rules	1600	431.20	1021	1218	1599	437.20	907.50	1470	1432	410.90	1266	1536
	time	1.75	1.32	3.27	0.52	1.71	1.34	3.25	0.55	1.66	1.32	3.37	0.63

Table 6: Results for the FRBCS ensembles

		Bagging+Greedy				Bagging+GRASP $\tau = 0.50$				Bagging + Random Subspace			
		Pima	Glass	Vehicle	Sonar	Pima	Glass	Vehicle	Sonar	Pima	Glass	Vehicle	Sonar
	5×2-cv	0.261	0.463	0.525	0.255	0.262	0.464	0.494	0.246	0.299	0.450	0.453	0.250
3 labels	#rules	8578	6208	6843	7282	8609	6289	7362	7951	7936	5671	8008	8174
5 #attr.	avg. #rules	171.55	124.16	136.87	145.65	172.18	125.77	147.24	159.03	158.71	113.42	160.16	163.47
	time	3.43	1.51	4.87	2.52	3.45	1.53	4.91	2.57	3.34	1.49	5.06	2.58
	5×2-cv	**0.235**	**0.396**	0.400	**0.240**	**0.234**	**0.405**	0.399	**0.220**	**0.260**	0.430	0.378	**0.221**
5 labels	#rules	29405	12877	22177	26769	29748	13302	25578	30068	27199	11998	30799	31824
5 #attr.	avg. #rules	588.11	257.54	443.55	535.37	594.95	266.04	511.56	601.36	543.97	239.96	615.97	636.47
	time	17.93	12.11	31.21	6.66	18.05	12.23	32.79	6.96	17.64	11.94	33.91	7.13
	5×2-cv	0.243	0.430	**0.375**	0.262	0.247	0.425	**0.353**	0.242	0.263	**0.402**	**0.330**	0.241
7 labels	#rules	64891	18633	48479	49587	65802	19272	54721	54684	59824	17999	67936	57298
5 #attr.	avg. #rules	1298	372.66	969.58	991.74	1316	385.45	1094	1094	1196	359.98	1359	1146
	time	84.70	67.36	166.51	24.72	85.27	68.27	170.48	25.49	82.12	66.06	174.24	25.57

Table 7: Statistical test for the comparison of the single FRBCS and the different FRBMCSs methodology. For each dataset, the best result is marked ('*') and the others are compared to it

		Best single classifier (app./labels)	Best ensemble (app./labels)	Best ens. selected (app./labels/fitness)	Best ens. TEFF (app./labels)
Pima	Approach	GRASP/5	GRASP/5	Greedy/5/LOFF	GRASP/5
	$\mu \pm \sigma$	0.246 ±0.00991	0.234 ±0.019	0.238 ±0.0167	0.240 ±0.0159
	Symbol	=	*	=	=
Glass	Approach	GRASP/5	Greedy/5	Greedy/3/LOFF	GRASP/3
	$\mu \pm \sigma$	0.375 ±0.0526	0.396 ±0.0568	0.358 ±0.0382	0.360 ±0.0507
	Symbol	=	=	*	=
Vehicle	Approach	GRASP/7	Random/7	Random/7/LOFF	GRASP/7
	$\mu \pm \sigma$	0.399 ±0.0262	0.330 ±0.0179	0.347 ±0.0196	0.337 ±0.0168
	Symbol	+	*	=	=
Sonar	Approach	Greedy/3	GRASP/5	Random/5/WCFF	GRASP/3
	$\mu \pm \sigma$	0.261 ±0.0463	0.220 ±0.0445	0.219 ±0.0264	0.216 ±0.0265
	Symbol	+	=	=	*

ISBN: 978-989-95079-6-8

result for each dataset is marked with a star '*'.

The best results (in average) are always obtained by the initial or the selected ensembles. Even if they are only significant for two datasets, they correspond to those with the largest dimension. Notice that, on sonar, the GA outperforms significantly the single classifier. The initial ensemble outperforms the best genetically selected FRBMCSs in one of the four cases (vehicle), whereas TEFF achieves the best result for the sonar dataset and WCFF for the sonar.

In the direct comparison, TEFF is the best choice two times for the vehicle and sonar dataset. The new fitness functions outperforms one of the four cases for pima with LOFF and glass dataset with WCFF.

Thus, combining bagging and the GA selection process to design FRBMCSs performs better for high dimensional problems with a large number of attributes, producing a smaller rule base while reducing the test errors in some cases, which was our original goal. When combining these two techniques with an advanced feature selection process we also get an improvement of the accuracy for datasets with higher dimensions (glass, vehicle and sonar, see Table 7).

6 Conclusions and future works

In this study, we extended our previously developed methodology in which a bagging approach together with a feature selection technique are used to train FRBMCSs, at a later stage selected by a multicriteria GA. Two new fitness functions were tested, the LOFF and the WCFF, based on one or two accuracy criteria (i.e., the training error and the likelihood). The generated selected FRBCS ensembles are performing correctly on classification problems with a significant number of features. By using abovementioned techniques, we would like to obtain FRCMCS dealing with high dimensional data.

One of the next steps we will consider in the future line is the design of a generic framework to define the multicriteria fitness function. At least two different information levels will be studied: the chromosome and the objective level. Furthermore, we would like to extend this study on larger data sets (more than 1,000 examples), to study the influence of other parameters (the GA parameters, the weighting coefficient in the WCFF, etc.), and to design more advanced genetic MCS selection techniques (e.g. the use of Pareto-based algorithms). Analysis of the different fuzzy rule genreation techniques and introduction a diversity criterion in the algorithm are another important points for future research.

References

[1] L. Kuncheva. *Combining Pattern Classifiers: Methods and Algorithms*. Wiley, 2004.

[2] T. Ho. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):832–844, 1998.

[3] D. Optiz and R. Maclin. Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research*, 11:169–198, 1999.

[4] J. Canul-Reich, L. Shoemaker, and L. O. Hall. Ensembles of fuzzy classifiers. In *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–6, London, 2007.

[5] W. Pedrycz and K.C. Kwak. Boosting of granular models. *Fuzzy Sets and Systems*, 157(22):2934–2953, 2006.

[6] S.J. Verzi, G.L. Heileman, and M. Georgiopoulos. Boosted ARTMAP: Modifications to fuzzy ARTMAP motivated by boosting theory. *Neural Networks*, 19(4):446–468, 2006.

[7] O. Cordón, A. Quirin, and L. Sánchez. A first study on bagging fuzzy rule-based classification systems with multicriteria genetic selection of the component classifiers. In *IEEE International Workshop on Genetic and Evolving Fuzzy Systems (GEFS)*, pages 11–16, Germany, 2008.

[8] L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.

[9] H. Ishibuchi, T. Nakashima, and M. Nii. *Classification and Modeling With Linguistic Information Granules*. Springer, 2005.

[10] O. Cordón, A. Quirin, and L. Sánchez. On the use of bagging, mutual information-based feature selection and multicriteria genetic algorithms to design fuzzy rule-based classification ensembles. In *Int. Conf. on Hybrid Intelligent Systems*, pages 549–554, Barcelona, 2008.

[11] R. Battiti. Using mutual information for selecting features in supervised neural net learning. *IEEE Transactions on Neural Networks*, 5(4):537–550, 1994.

[12] T. A. Feo and M. G. C. Resende. Greedy randomized adaptive search procedures. *J. of Global Optimization*, 6:109–133, 1995.

[13] O. Cordón and A. Quirin. Comparing two genetic overproduce-and-choose strategies for fuzzy rule-based multiclassification systems generated by bagging and mutual information-based feature selection. Technical Report #AFE 2008-05, European Centre for Soft Computing, Mieres, 2008.

[14] D. Partridge and W.B. Yates. Engineering multiversion neural-net systems. *Neural Computation*, 8(4):869–893, 1996.

[15] L.S. Oliveira, M. Morita, R. Sabourin, and F. Bortolozzi. Multi-objective genetic algorithms to create ensemble of classifiers. *Lecture Notes in Computer Science*, 3410:592–606, 2005.

[16] S.T. Hadjitodorov and L.I. Kuncheva. Selecting diversifying heuristics for cluster ensembles. *Lecture Notes in Computer Science*, 4472:200–209, 2007.

[17] G. Martínez-Munoz, D. Hernández-Lobato, and A. Suárez. Selection of decision stumps in bagging ensembles. *Lecture Notes in Computer Science*, 4668:319–328, 2007.

[18] B. Gabrys and D. Ruta. Genetic algorithms in classifier fusion. *Applied Soft Computing*, 6(4):337–347, 2006.

[19] O. Cordón, M.J. del Jesus, and F. Herrera. A proposal on reasoning methods in fuzzy rule-based classification systems. *International J. of Approximate Reasoning*, 20:21–45, 1999.

[20] C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. University of Illlinois Press, 1949.

[21] Z. Michalewicz. *Genetic Algorithms + Data Structures = Evolution Programs*. Springer, 1996.

[22] T.G. Dietterich. Approximate statistical test for comparing supervised classification learning algorithms. *Neural Computation*, 10(7):1895–1923, 1998.

[23] E. Lehmann. *Nonparametric Statistical Methods Based on Ranks*. McGraw-Hill, New-York, 1975.

ISBN: 978-989-95079-6-8

A new Fuzzy Noise-rejection Data Partitioning Algorithm with Revised Mahalanobis Distance

M.H. Fazel Zarandi, Milad Avazbeigi[1] I.B. Turksen[2]

1 Department of Industrial Engineering, Amirkabir University of Technology
Tehran, Iran
2 Department of Industrial Engineering, TOBB Economy and Technology University
Ankara, Turkey
Email: zarandi@cic.aut.ac.ir, miladavazbeigi@cic.aut.ac.ir, bturksen@etu.edu.tr

Abstract—*Fuzzy C-Means (FCM) and hard clustering are the most common tools for data partitioning. However, the presence of noisy observations in the data may cause generation of completely unreliable partitions from these clustering algorithms. Also, application of the Euclidean distance in FCM only produces spherical clusters. In this paper, a new noise-rejection clustering algorithm based on Mahalanobis distance is presented which is able to detect the noise and outlier data and also ellipsoidal clusters. Unlike the traditional FCM, the proposed clustering tool provides much efficient data partitioning capabilities in the presence of noise and outliers. For validation of the proposed model, the model is applied to different noisy data sets.*

Keywords— Cluster Validity Index (CVI), Fuzzy C-Means (FCM), Possibilistic C-means (PCM), Revised Gustafson-Kessel (GK), Revised Mahalanobis Distance.

1 Introduction

Clustering methods have been extensively used in computer vision and pattern recognition. Fuzzy clustering methods have shown spectacular ability to detect not only volume clusters, but also clusters which are actually thin shells, i.e. curves and surfaces.

Most analytic fuzzy clustering approaches are derived from the fuzzy C-means (FCM) algorithm. FCM uses the probabilistic constraint that the membership of a data point across classes sums to 1. The constraint is used to generate the memberships update equations for an iterative algorithm. The memberships resulting from FCM and its derivative however, do not always correspond to the intuitive concept of belonging or compatibility. Moreover, the algorithms have considerable trouble in noisy environments. Reference [1] summarizes the main problems of classic FCM as follows:

- In order to get the optimal partition, initial locations of the cluster centers should be assigned. The FCM algorithm always converges to a local extreme. Different choices of initial cluster centers may lead to a different extrema.
- The scientific basis for the choice of m, the weighting exponent, is still not clear.
- The optimum number of clusters in the data is assigned a priori. There should be a criterion to assign the optimal number of clusters.

To overcome these problems and drawbacks, first, [2] introduces a new method for fuzzy clustering called possibilistic fuzzy clustering. Their approach differs from the previous clustering methods in that the resulting partition of the data can be interpreted as a possibilistic partition, and the membership values may be interpreted as degrees of possibility of the points belonging to the classes, i.e., the compatibilities of the points with the class prototypes. They construct an appropriate objective function whose minimum will characterize a good possibilistic partition of the data, and derive the membership and prototype update equations from necessary conditions for minimization of the related criterion function. Next nominated work is [1]. Melek et al. in [1] show how their PCM addresses the mentioned problems of the FCM for some example cases. However, their method has some limitations which are the concentration of this paper.

The Rest of the paper is organized as follows: Section 2 discusses the drawbacks of the previous PCM's. Then, in section 3 the proposed method is presented in order to overcome these limitations. Section 4 applies the method for different data sets to verify and validate the proposed method.

2 Limitations of the traditional PCM's

Reference [1] uses Euclidean distance in all data partitioning steps. However, clearly, Euclidean distance usually fails to recognize the appropriate shape of the clusters in complex data sets. Especially, when data include ellipsoidal shapes, Euclidean distance loses sight to those data which should be considered in a cluster from data which should not be included in that cluster.

Moreover, the cluster validity index (*CVI*) applied in [1], suffers from linear behaviour of the Euclidean distance. There are many samples of data sets which the applied *CVI* is not able to recognize the number of clusters correctly.

Another limitation is about noise identification procedure. Application of Euclidean distance can mislead the noise rejection procedure. This problem is discussed in the related section of the proposed algorithm.

While Euclidean distance implies the above mentioned limitations, Mahalanobis distance can mitigate or in some cases can overcome the cited limitations completely. Another advantage of the Mahalanobis distance is that the

ISBN: 978-989-95079-6-8

Mahalanobis distance can identify both spherical and ellipsoidal clusters correctly.

3 Improvements to the traditional PCM's

As quoted in introduction, the proposed method applies the Mahalanobis distance for noise rejection problem. However, when applying mahalanobis distance, some difficulties may occur with covariance matrix.

In this section, first, the problem with calculation of the Covariance Matrix in Mahalanobis distance is discussed. An estimation of Covariance Matrix is introduced to mitigate the problem. Then, new Mahalanobis distance and revised Gustafson-Kessel Clustering are presented based on the estimation. Next, the revised Gustafson-Kessel Clustering method with the new Mahalanobis distance is integrated into the PCM which is presented in [1].

Moreover, a new *CVI* is applied which is based on proximity of two fuzzy sets. The new CVI is independent from the distance type and therefore can be integrated to the method with Mahalanobis distance. This *CVI* is used in [3].

3.1 Revised GK with revised Mahalanobis

In statistics, Mahalanobis distance is a distance measure introduced by P. C. Mahalanobis in 1936. It is based on correlations between variables by which different patterns can be identified and analyzed. It is a useful way of determining similarity of an unknown sample set to a known one. It differs from Euclidean distance in that it takes into account the correlations of the data set and is scale-invariant, i.e. not dependent on the scale of measurements.

Mahalanobis distance can be defined as dissimilarity measure between two random vectors $\vec{x}, \vec{y}$ of the same distribution with the covariance matrix S:

$$D_M(\vec{x},\vec{\mu}) = \sqrt{(\vec{x}-\vec{\mu})^T S^{-1}(\vec{x}-\vec{\mu})}$$

The eigenvalues and eigenvectors of the covariance matrix describe the shape and orientation of the clusters in GK method of clustering which applies mahalanobis distance instead of Euclidean one. When an eigenvalues is zero or when the ration between the maximal and minimal eigenvalues, i.e., the condition number of F, is very large the matrix is nearly singular. In such a case, the inverse of covariance matrix cannot be calculated. Also the normalization to a fixed volume fails, as the determinant (the volume of the covariance matrix) becomes zero and the following formula thus cannot be applied in GK method [4]:

$$\det(F_i)^{1/n} F_i^{-1} \quad (1)$$

A straightforward way to avoid numerical problems is to constrain the ratio between the maximal and minimal eigenvalues such that it is smaller than some predefined threshold. When this threshold exceeds, the minimal eigenvalues is increased such that the ratio equals to the threshold and the covariance is reconstructed by [4]:

$$F = \Phi \Lambda \Phi^{-1} \quad (2)$$

where, Λ is the diagonal matrix containing the limited eigenvalues and Φ is a matrix whose columns are the corresponding eigenvectors.

Fig. 1 shows an example of a data set which cannot be clustered with standard GK. Using new estimation for covariance matrix, the numerical problems would be resolved and GK can appropriately find the correct number of clusters (example from [4]).

The above modification prevents the GK algorithm from running into numerical problems. However, as a result one can get clusters that are extremely long in the direction of the largest eigenvalues and have little relationship with real distribution of data. This can cause over fitting of the data and consequently one obtains a poor model [4].

This problem occurs mainly when the number of data points in a cluster becomes too low. In such a case, the computed covariance matrix is not a reliable estimate of the underlying data distribution [7]. One way to tackle this problem is to

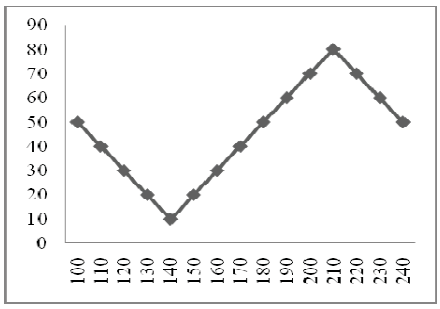


Figure 1: Linear Clusters

limit the ratio between maximal and minimal eigenvalues even further than described in the previous section. This will prevent the extreme elongation of the clusters. Another way is to add a scaled identity matrix to the covariance matrix. Reference [7] and [8] describe several different methods to improve the covariance estimation. Inspired by these methods, [4] proposes the following estimate for the GK algorithm and calculation of Mahalanobis distance:

$$F_i^{new} = (1-\gamma)F_i + \gamma \det(F_0)^{1/n} I \quad (3)$$

where $\gamma \epsilon [0,1]$ is the tuning parameter and F_0 is the covariance matrix of the whole data set. Depending on γ, the clusters are forced to have a more or less equal shape. When γ is 1, all covariance matrices are equal and have the same size, which of course limits the possibility of the algorithm to properly identify clusters.

For the complete description of the revised GK, readers can refer to [4]. The main steps of the algorithm are as follows [4]:

Repeat for *l = 1, 2, …, Ite#*

Step 1: Compute cluster prototypes (means)

$$v_i^{(l)} = \frac{\sum_{k=1}^{N} (\mu_{ik}^{(l-1)})^m Z_k}{\sum_{k=1}^{N} (\mu_{ik}^{(l-1)})^m}, 1 <= i <= K \quad (4)$$

Step 2: Compute the cluster covariance matrices

$$F_i = \frac{\sum_{k=1}^{N} (\mu_{ik}^{(l-1)})^m (Z_k - v_i^{(l)})(Z_k - v_i^{(l)})^T}{\sum_{k=1}^{N} (\mu_{ik}^{(l-1)})^m}, 1 \le i \le K \quad (5)$$

Add a scaled identity matrix:

$$F_i = (1-\gamma)F_i + \gamma \det(F_0)^{1/n} I, 1 \le i \le K \quad (6)$$

Extract eigenvalues λ_{ij} and eigenvectors ϕ_{ij} from F_i. Find $\lambda_{imax}=max_j\lambda_{ij}$ and set:

$$\lambda_{ij} = \frac{\lambda_{i\max}}{\beta} \; \forall j \; for \; which \; \frac{\lambda_{ij}}{\lambda_{i\max}} > \beta \quad (7)$$

Reconstruct F_i by

$$F_i = [\phi_{i1} \dots \phi_{in}] diag(\lambda_{i1}, \dots \lambda_{in}) [\phi_{i1} \dots \phi_{in}]^{-1}$$

Step 3: Compute the distances

$$D_{ikA_i}^2 = (Z_k - v_i^{(l)})^T [\rho_i \det(F_i)^{1/n} F_i^{-1}](Z_k - v_i^{(l)}) \quad (8)$$

where *1<=i<=k* and *1<=k<=N*.

Step 4: Update the partition matrix

$$for \; 1 \le k \le N$$

$$if \; D_{ikA_i} > 0 \; for \; 1 \le i \le K,$$

$$\mu_{ik}^{(l)} = \frac{1}{\sum_{j=1}^{K} (D_{ikA_j} / D_{ikA_i})^{\frac{2}{m-1}}} \quad (9)$$

Otherwise

$$\mu_{ik}^{(l)} = 0 \; if \; D_{ikA_i} > 0, \; and \; \mu_{ik}^{(l)} \in [0,1] \quad (10)$$

with $\sum_{i=1}^{K} \mu_{ik}^{(l)} = 1 \; otherwise.$

Until $\left\| U^{(l)} - U^{(l-1)} \right\| < \varepsilon$.

Improved GK clustering does not have the problems that standard GK may face in some cases. In the next section, applied cluster validity index is described.

3.2 Cluster Validity Index (CVI)

This paper uses a cluster validity index proposed by [9].

Definition1: The relative similarity $S_{rel}(x_j{:}A_p,A_q)$ between two fuzzy sets A_p and A_q at x_j is defined as:

$$\frac{f(x_j{:}A_p \cap A_q)}{f(x_j{:}A_p \cap A_q) + f(x_j{:}A_p - A_q) + f(x_j{:}A_q - A_p)} \quad (11)$$

while $f(x_j{:}A_p \cap A_q) = u_{A_p}(x_j) \wedge u_{A_q}(x_j)$ and $\wedge$ is the minimum operator. Moreover for function of deference:

$$f(x_j{:}A_p - A_q) = Max(0, u_{A_P}(x_j) - u_{A_q}(x_j)) \quad (12)$$

Definition2: The relative similarity between two fuzzy sets A_p and A_q is defined as:

$$S_{rel}(A_p,A_q) = \sum_{j=1}^{n} S_{rel}(x_j : A_p,A_q) h(x_j) \quad (13)$$

where

$$h(x_j) = -\sum_{p=1}^{c} u_{A_p}(x_j) \log(u_{A_p}(x_j)) \quad (14)$$

Here, h is the entropy of datum x_j and $u_{A_p}(x_j)$ is the membership value of x_j to cluster A_p.

Definition3: The cluster validity index is as follows:

$$V(U,V;X) = \frac{2}{c(c-1)} \sum_{p \ne q}^{c} S_{rel}(A_p,A_q) \quad (15)$$

The optimal number of clusters is obtained by minimizing V over the range of c values (Number of clusters).

Moreover, some other popular CVI's such as "Xie and Beni" CVI [5] and "Known" CVI [6] are evaluated. Results show that the new CVI has better results and is able to find the number of clusters correctly in many complex situations.

The inputs to the CVI used in the proposed model are obtained by the revised GK.

4 The proposed method

In this section, we present our method base on the modifications presented in section 3.

For the selection of the weight exponent (m), it is suggested to be chosen far from its both extremes so as to ensure that the cluster validity index shows the optimum number of fuzzy clusters. A fuzzy total scatter matrix is defined in [10] as:

$$S_T = \sum_{k=1}^{N} (\sum_{i=1}^{K} (u_{ik})^m)(x_k - \overline{v})(x_k - \overline{v})^T \quad (16)$$

The trace of the fuzzy total scatter matrix decreases monotonically from a constant value z to zero as m varies from one to infinity. For data partitioning, a suitable value for m is that which gives a value for trace (s_T) equal to z/2 [11]. The constant value z is defined as:

$$z = trace(\sum_{k=1}^{N} [(x_k - \frac{1}{N}\sum_{k=1}^{N} x_k)(x_k - \frac{1}{N}\sum_{k=1}^{N} x_k)^T]) \quad (17)$$

Using the trace value, we determine the value of (m) which is the degree of fuzziness of the system. Next, using the introduced *CVI* and achieved (m), the appropriate number of clusters would be obtained.

After determination of m and number of clusters, we repeat the following procedure, iteratively:

We first implement the revised GK on the data set. To choose the initial cluster centers of the revised GK, we apply simple FCM for initialization of our clustering method. Other methods like [1] have applied AHC. The GK is sensitive to the initial cluster centers and initial membership function values. Our experiments show that unlike FCM, AHC is not a good method for the initialization of the revised GK.

Next, in order to find the data points that are ''too far'' from all cluster centers, [13] proposes the following index for each data point x_j:

$$W_j = \sum_{i=1}^{C} \left\| x_j - v_{hi} \right\|_A \quad (18)$$

where, *j = 1, 2, ... , N*, c is the number of clusters, and N is the number of data. The index W_j is the summation of the distance of the data point x_j to all cluster centers. This gives a measure of how far each data point is from the different cluster centers assigned in the first step of the algorithm. The noise is identified through the data points that have large

ISBN: 978-989-95079-6-8

values of W_j and therefore, a threshold X is assigned to trim these outliers from the data set. The value of the threshold depends on the range of the input data to the algorithm. While [1] applies (18) to find the W_j, we proposed the following equation in order to find the W_j:

$$W_j = \underset{i=1}{\overset{C}{Mean}}(WGHT_i \times \left\| x_j - \upsilon_{hi} \right\|_A) \quad (19)$$

where, "$WGHT_i$" is a weight assigned to the *i-th* distance after that we sorted the distances of x_j from all cluster centers (c). Reference [1] does not consider such weight. It is clear that simply the summation of distances can not provide a good measure for detection of the outliers. Instead the proposed model can assign the largest weight to the smallest distance and the smallest weight to the largest distance vice versa using (19). Weights are the input to the model. After choosing the threshold, [1] computes:

$$z = \frac{\eta_n}{N} \quad (20)$$

where, η_n is the number of noise points and N is the total number of data. The percentage of ''good'' data points, i.e., inliers can then be calculated as:

$$\hat{z} = 1 - z \quad (21)$$

After identifying the percentage of inliers in the data, we compute the corresponding chi-square data distribution value [13]. Then, we calculate the cut-off distance:

$$u_{FC}cut^2 = \upsilon_i \chi^2 \quad (22)$$

where, υ_i is a resolution parameter that depends on the number of clusters, and χ^2 is the chi-square value computed by (21). By knowing the new cut-off distance, the optimum number of clusters, the degree of fuzziness, and the initial location of the clusters centers, we calculate the membership matrix through (23):

$$u_{ij} = \frac{1}{1 + \{\frac{d^2(x_j, \upsilon_i)}{\upsilon_i}\}^{\frac{1}{m-1}}} \quad (23)$$

It should be noted that (20), (21), (22) and (23) are presented in [1]. In the proposed PCM, we use these equations with revised Mahalanobis distance.

5 Results of Experiments

In this section, the proposed model is applied for handling four different cases. The parameters of the proposed model for these cases are summarized in table 1:

- *NOC*: Number of Clusters
- *m* is the degree of fuzziness of system obtained through 16 and 17.
- *Ω* is the cut-off distance
- *Beta and Gamma* are the parameters of (3) and (7).
- *#Itr*: The number of iteration which PCM goes on.
- *W: WGHT* in (19).
- *Error:* If the changes in the value of membership functions were smaller than "*Error*", the algorithm stops.

In all cases, *Beta* is 1.00E+16, *#Itr* is 7 and *Error* is 1.00E-07. It should be noted that the data of the cases and their related configuration are very similar to [9] and [3]. However, we regenerated the data by ourselves.

Table 1: Parameters of the algorithm for different cases

Parameter	***NOC***	***m***	***Ω***	***Gamma***	***W***
Case 1	4	3	210	1.00E-07	[6 2 2 1]
Case 2	4	2	375	1.00E-09	[8 3 2 2]
Case 3	5	2	315	1.00E-01	[8 3 2 2 1]
Case 4	5	3	4590	1.00E-07	[8 4 3 2 1]
Case 5	5	3	805	1.00E-01	[8 4 3 2 1]

It should be mentioned that the experiments can't be applied for the previous noise rejection methods in the literature. This is because of the ellipsoidal forms of data which can't be handled using Euclidean distance. To make this clear, in Fig. 2, the clustering methods based on Euclidean distance apparently fail to recognize the shapes of clusters and noises correctly. That's why, only experiments are applied for the proposed method.

5.1 Case Study I

Every data cloud includes 500 data and 50 data are randomly generated as noise. Main data and noises both have Gaussian distribution.

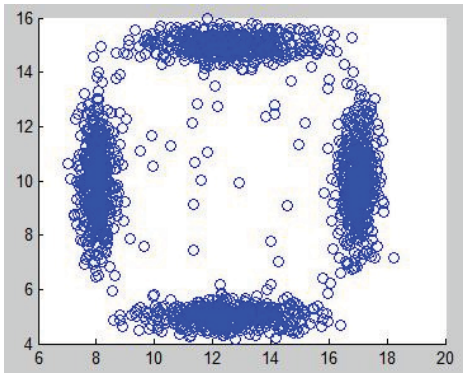


Figure 2: Data before clustering

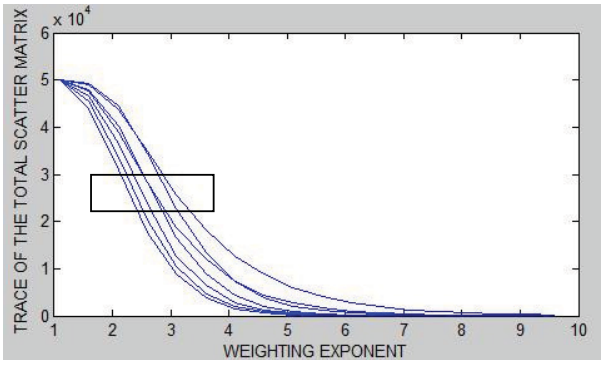


Figure 3: Finding the appropriate *m* in case study I

To determine the exponent of fuzziness (*m*), the total scatter matrix is calculated for different values of m and different

ISBN: 978-989-95079-6-8

number of clusters as is shown in (16). Then using z in (17), the appropriate m is chosen.
The trace value calculated as 5.8315e+004. We use half of trace value (2.9157e+004) to determine the optimum m (Fig. 3).

To choose the correct number of clusters, we need to calculate the CVI for different number of clusters. Fig. 4 shows the applied CVI for case study I. As the figure shows, the optimum number of clusters is equal to four.
Base on Fig. 3 and Fig. 4, we consider fuzziness exponent (m) equal to 2.5 and the number of clusters equal to 4. The result of the clustering is shown in Fig. 5. The result shows that the proposed method can identify the noises from the main data accurately. The expansion of the eclipse shape clusters can be limited by Ω as an input to the model.

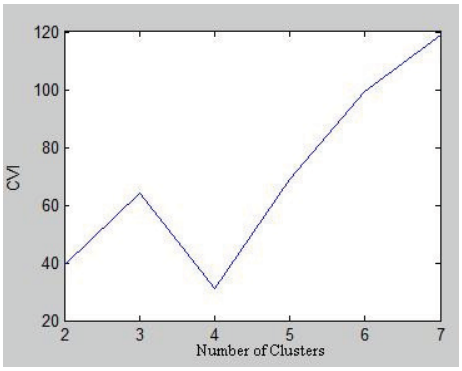
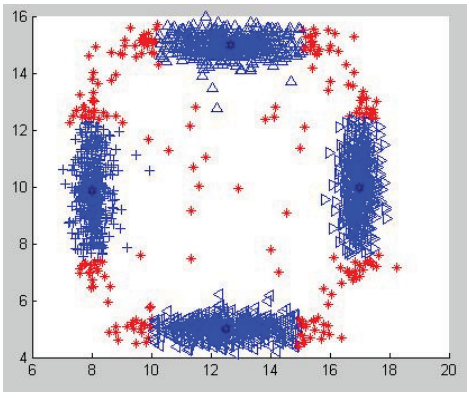
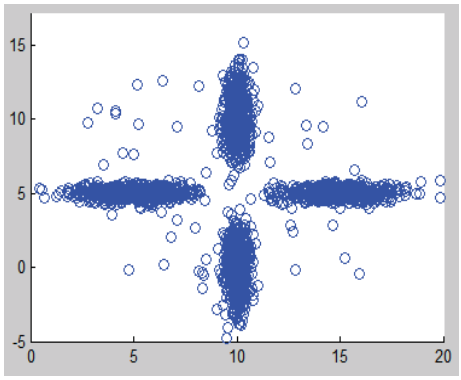


Figure 4: Identification of the optimum number of clusters


Figure 5: Data after Clustering

It is clear that, the method presented in [1] can not identify the correct shape of the clusters because of the spherical behaviour of the data. In Fig. 5, data which are shown with star "*", have a maximum membership function value lower than 0.2, hence are identified as noise.
In the other cases, each data cloud includes 500 data and 50 data are randomly generated as noise. Main data and noises both have Gaussian distribution. Also only the figures related to clustering and choosing the appropriate number of clusters are presented.

5.2 *Case Study II*


Figure 6: Data before clustering

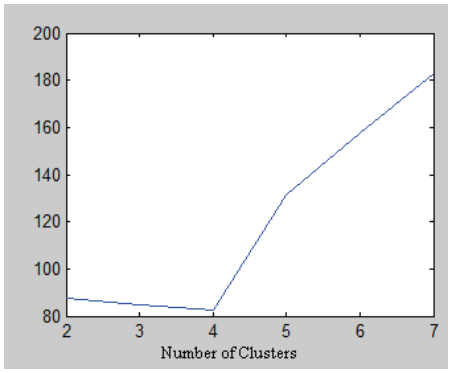
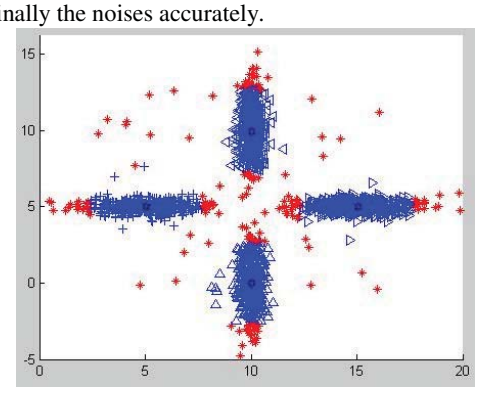


Figure 7: Identification of the optimum number of clusters

This case is presented to show the strength of the used CVI and the effect of Ω. As Figure 8 shows, the combination of these two elements in the proposed model, enable the model to identify the number of clusters, the shapes of the clusters and finally the noises accurately.


Figure 8: Data after Clustering

5.3 *Case Study III*

Fig 9 shows the data set and noises.

ISBN: 978-989-95079-6-8

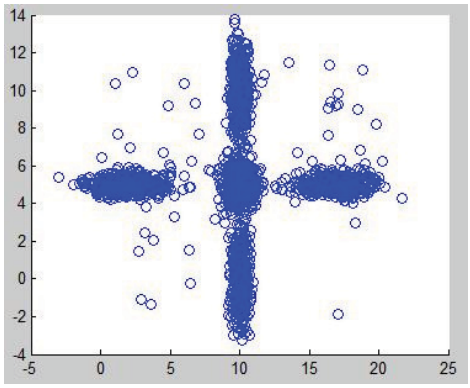


Figure 9: Data before clustering

As Fig. 10 shows, the optimum number of clusters is equal to five. This case is presented to validate that whether the method is able to find the spherical and ellipsoidal shapes simultaneously or is not. As Figure 11 shows, the method could find the circle at the center of the figure and four eclipses around the circle accurately. Also the noises are detected accurately as shown in Figure 11.

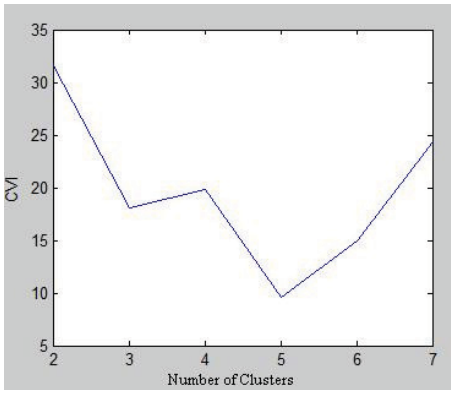


Figure 10: Identification of the optimum number of clusters

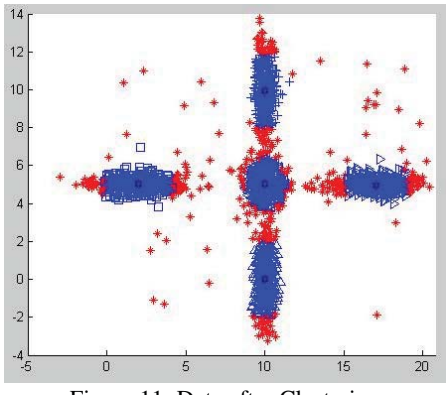


Figure 11: Data after Clustering

5.4 Case Study IV

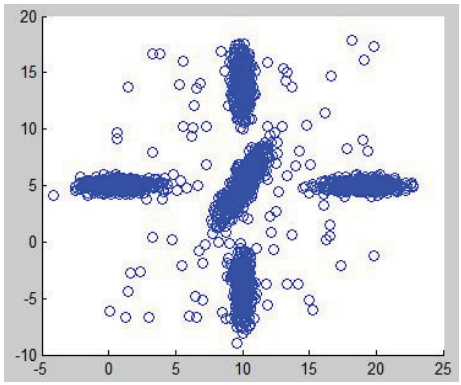


Figure 12: Data before clustering

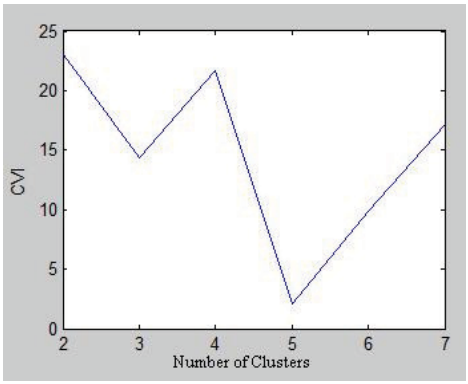


Figure 13: Identification of the optimum number of clusters

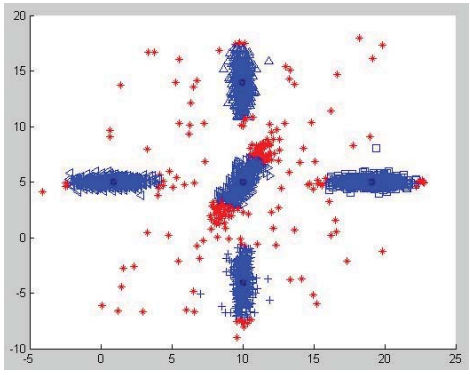


Figure 14: Data after Clustering

In this case, five eclipses are spread in different directions as shown in Figure 12. The clustering results show that the method can find the complex behaviors of data in different directions accurately as well as simple cases.

6 Conclusions

In this paper, a new noise rejection clustering method base on Mahalanobis distance and a different cluster validity index is presented. The new method can be distinguished from the existing methods in the literature from the following points:

ISBN: 978-989-95079-6-8

- The noise rejection clustering methods existing in the literature mainly use Euclidean distance in their clustering method. The proposed method, applies Mahalanobis distance which enables the method to identify the ellipsoidal behavior of data besides spherical behavior.
- In order to find the appropriate number of clusters, the method uses a well-defined cluster validity index which is independent from distance type.
- A new weighting system is attached to the noise rejection method, which helps for better detection of noises.

Method is applied to different cases and the results show that the method is capable of identification of noise and outliers within spherical and ellipsoidal data.

References

[1] W.W. Melek, A.A. Goldenberg, M.R. Emami, A fuzzy noise-rejection data partitioning algorithm, *International Journal of Approximate Reasoning*, 38: 1–17, 2005.

[2] R. Krishnapuram, J.M. Keller, A Possibilistic Approach to Clustering, *IEEE Transactions on Fuzzy Systems,* 2 (1): 98-110, 1993.

[3] M.H. Fazel Zarandi, B. Rezaee, I.B. Turksen, E. Neshat, A type-2 fuzzy rule-based expert system model for stock price analysis. *Expert Systems with Application*, 36:139-154, 2009.

[4] R. Babuska, R.J. van der Veen, U. Kaymak, Improved Covariance Estimation for Gustafson-Kessel Clustering. *IEEE*, 1081-1085, 2002.

[5] X.L. Xie, G. Beni, A validity measure for fuzzy clustering, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8): 841–847, 1991.

[6] S.H. Kwon, Cluster validity index for fuzzy clustering. *Electronics Letters*, 34(22): 2176–2177, 1998.

[7] S. Tadjudin and D.A. Landgrebe. Covariance estimation with limited training samples. *IEEE Transactions on Geosciences and Remote Sensing*, 37(4): 2113-2118, July 1999.

[8] J.F. Friedman, Regularize discriminate analysis, *J.R. Statist. Soc.*, 84:17-42, 1989.

[9] Kim, Y.I., Kim, D.W., Lee, D., Lee, K.H., A cluster validation index for GK cluster analysis based on relative degree of sharing, *Information Science*, 168:225-242, 2004.

[10] M. Sugeno, T. Yasukawa, A fuzzy-logic-based approach to qualitative modeling, *IEEE Transactions on Fuzzy Systems,* 1(1), 1993.

[11] M.R. Emami, I.B. Turksen, A.A. Goldenberg, Development of a systematic methodology of fuzzy logic modeling, *IEEE Transactions on Fuzzy Systems,* 6 (3): 346–361, 1998.

[12] J.H. Ward, Hierarchical grouping to optimize an objective function, *Journal of American Statistics Association,* 58: 236–244, 1963.

[13] W.W. Melek, Neurofuzzy Control of Modular and Reconfigurable Robots, PhD. Dissertation, University of Toronto, Canada, 2002.

ISBN: 978-989-95079-6-8

Determining OWA weights by maximizing consensus

José Luis García-Lapresta Bonifacio Llamazares Teresa Peña

PRESAD Research Group, Dept. of Applied Economics
University of Valladolid, Spain
Email: {lapresta, boni, maitepe}@eco.uva.es

Abstract— *In this paper we propose a method for generating OWA weighting vectors from the individual assessments on a set of alternatives in such a way that these weights maximize the consensus among individual assessments with respect to the outcome provided by the OWA operator.*

Keywords— OWA operators; consensus; distances; mathematical programming.

1 Introduction

In 1988 Yager [9] introduced OWA operators as a tool for aggregating numerical values in multi-criteria decision making. An OWA operator is similar to a weighted mean, but with the values of the variables previously ordered in a decreasing way. Thus, contrary to the weighted means, the weights are not associated with specific variables and, therefore, they are anonymous. Moreover, they satisfy other interesting properties, such as monotonicity, unanimity, continuity and compensativeness.

Initially, the weights of an OWA operator may be fixed taking into account the importance we want to give to the assessments. So, the outcome of an OWA operator may be the maximum, the minimum, the average or a median of the individual assessments, among a large number of possibilities.

It is important to note that the determination of the weights of OWA operators is a relevant issue since the origins of the theory of OWA operators. In this way, Yager [9] proposes to use linguistic quantifiers for generating the OWA weights; O'Hagan [6] generates the OWA weights by maximizing their entropy whenever a degree of orness has been fixed; Filev and Yager [3] consider an exponential smoothing approach for generating the OWA weights. After these seminal papers, a large variety of techniques have been proposed in the literature (see, for instance, Wang and Parkan [7] and Xu [8]).

When a group of individuals provides assessments on an alternative and these values are aggregated, it is relevant to know the degree of agreement or consensus among the individual assessments with respect to the aggregated value. In fact, it is desirable that the aggregation function used to obtain the collective value reflects the opinions of as many agents as possible. Using a specific OWA operator for aggregating individual assessments does not necessarily ensure such consensus for every opinion situation.

In our proposal, we do not fix the OWA weighting vector, but we generate an OWA operator for each profile of individual assessments, just one that maximizes the consensus (or equivalently, minimizes the disagreement) in the group with respect to the outcome provided by the OWA operator. More concretely, once the agents opinions are known, we first calculate the distances among individual assessments on the alternatives and the collective assessments generated by an arbitrary OWA operator. Secondly, we use an aggregation operator for obtaining a representative measure of disagreement from the individual assessments to the collective one. By solving a mathematical program, we obtain the weighting vector(s) that maximize(s) the consensus among individual and collective opinions.

The paper is organized as follows. Section 2 is devoted to introduce notation, basic notions and our proposal for generating an OWA operator for each profile of individual assessments. Section 3 contains some illustrative examples. Finally, Section 4 shows some open problems and further research.

2 A model for generating OWA weights

Consider a set of agents (experts or voters) $V = \{1, \dots, m\}$ ($m \geq 2$) who show their opinions on a set of alternatives $A = \{a_1, \dots, a_n\}$ ($n \geq 2$) through numbers in the interval $[0, 1]$.

A *profile* is a $m \times n$ matrix

$$P = \begin{pmatrix} a_1^1 & \cdots & a_j^1 & \cdots & a_n^1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_1^i & \cdots & a_j^i & \cdots & a_n^i \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_1^m & \cdots & a_j^m & \cdots & a_n^m \end{pmatrix}$$

where $a_j^i \in [0, 1]$ is the assessment that agent i assigns to alternative a_j. The set of profiles is denoted by $\mathcal{P}$.

Given a weighting vector $\boldsymbol{w} = (w_1, \dots, w_m) \in [0, 1]^m$ such that $\sum_{i=1}^m w_i = 1$, the *OWA operator associated with* $\boldsymbol{w}$ is the mapping $F_{\boldsymbol{w}} : [0, 1]^m \longrightarrow [0, 1]$ defined by

$$F_{\boldsymbol{w}}(x_1, \dots, x_m) = \sum_{i=1}^m w_i \cdot y_i$$

where y_i is the i-th greatest number of $\{x_1, \dots, x_m\}$. We use the following notation:

$$\mathcal{W} = \left\{ \boldsymbol{w} \in [0, 1]^m \;\middle|\; \sum_{i=1}^m w_i = 1 \right\}.$$

Given $\boldsymbol{w} \in \mathcal{W}$, from the opinions given by the agents on an alternative a_j, $\{a_j^1, \dots, a_j^m\}$, we generate a collective assessment on a_j:

$$v_j(\boldsymbol{w}) = F_{\boldsymbol{w}}\left(a_j^1, \dots, a_j^m\right).$$

We denote $\boldsymbol{v}(\boldsymbol{w}) = (v_1(\boldsymbol{w}), \dots, v_n(\boldsymbol{w}))$ the vector that contains the collective assessments on the alternatives of A generated by the OWA operator $F_{\boldsymbol{w}}$.

ISBN: 978-989-95079-6-8

By $\boldsymbol{a}^i = \left(a_1^i, \ldots, a_n^i\right)$ we denote the vector that contains the assessments of individual $i \in \{1, \ldots, m\}$ on the alternatives of A.

In our context, an *aggregation operator* is a continuous mapping $A : [0,1]^m \longrightarrow [0,1]$ that satisfies the following conditions:

1. *Monotonicity*, i.e., $A(x_1, \ldots, x_m) \leq A(y_1, \ldots, y_m)$ for all $(x_1, \ldots, x_m), (y_1, \ldots, y_m) \in [0,1]^m$ such that $x_i \leq y_i$ for every $i \in \{i, \ldots, m\}$.
2. *Unanimity*, i.e., $A(x, \ldots, x) = x$ for every $x \in [0,1]$.

It is easy to see that every aggregation operator is *compensative*, i.e.,

$$\min\{x_1, \ldots, x_m\} \leq A(x_1, \ldots, x_m) \leq \max\{x_1, \ldots, x_m\},$$

for every $(x_1, \ldots, x_m) \in [0,1]^m$.

On aggregation operators, see Fodor and Roubens [4], Grabisch, Orlovski and Yager [5], Calvo, Kolesárova, Komorníková and Mesiar [2] and Beliakov, Pradera and Calvo [1], among others.

2.1 The general model

In order to present our general model, it is necessary to fix two ingredients:

- A distance $d : [0,1]^n \times [0,1]^n \longrightarrow [0,1]$.
- An aggregation operator $A : [0,1]^m \longrightarrow [0,1]$.

Given a profile $P \in \mathcal{P}$, we propose to find weighting vector(s) $\boldsymbol{w} \in \mathcal{W}$ being solution(s) of the following mathematical program

$$\begin{array}{ll} \text{Min} & A\Big(d(\boldsymbol{a}^1, \boldsymbol{v}(\boldsymbol{w})), \ldots, d(\boldsymbol{a}^m, \boldsymbol{v}(\boldsymbol{w}))\Big) \\ \text{s. t. :} & \boldsymbol{w} \in \mathcal{W} \end{array} \tag{1}$$

Notice that from continuity of A and compactness of $\mathcal{W}$, the existence of solution(s) in (1) is always guaranteed.

Among the large variety of distances and aggregation operators that we may use in (1), we present with more detail those cases where Manhattan and Chebyshev distances are used, and the aggregation operators are the arithmetic mean and the maximum.

The Manhattan distance is defined by

$$d_1\big((x_1, \ldots, x_n), (y_1, \ldots, y_n)\big) = \sum_{i=1}^{n} |x_i - y_i|.$$

The Chebyshev distance is defined by

$$\begin{array}{l} d_\infty\big((x_1, \ldots, x_n), (y_1, \ldots, y_n)\big) = \\ = \max\{|x_1 - y_1|, \ldots, |x_n - y_n|\}. \end{array}$$

2.2 Using the arithmetic mean as aggregation operator

If we consider the arithmetic mean as aggregation operator, then (1) is equivalent to find the weighting vector that minimizes the sum of distances between the individual assessments and the collective assessments generated by the OWA operator associated with that weighting vector, i.e.

$$\begin{array}{ll} \text{Min} & \displaystyle\sum_{i=1}^{m} d\Big((a_1^i, \ldots, a_n^i), (v_1(\boldsymbol{w}), \ldots, v_n(\boldsymbol{w}))\Big) \\ \text{s. t. :} & \boldsymbol{w} \in \mathcal{W} \end{array} \tag{2}$$

1. If we use the Manhattan distance, then it is necessary to solve the following mathematical program:

$$\begin{array}{l} \text{Min} \displaystyle\sum_{i=1}^{m} \Big(|a_1^i - v_1(\boldsymbol{w})| + \cdots + |a_n^i - v_n(\boldsymbol{w})|\Big) \\ \text{s. t. :} \ w_1 \geq 0, \ldots, w_m \geq 0, \ w_1 + \cdots + w_m = 1 \end{array}$$

2. If we use the Chebyshev distance, then it is necessary to solve the following mathematical program:

$$\begin{array}{l} \text{Min} \displaystyle\sum_{i=1}^{m} \max\Big\{|a_1^i - v_1(\boldsymbol{w})|, \ldots, |a_n^i - v_n(\boldsymbol{w})|\Big\} \\ \text{s. t. :} \ w_1 \geq 0, \ldots, w_m \geq 0, \ w_1 + \cdots + w_m = 1 \end{array}$$

In the first case, i.e., when the Manhattan distance is used, it is possible to give the analytical solution of the problem.

Proposition 1 *The solution of Problem (2) for the Manhattan distance is the median operator.*

Proof

When the Manhattan distance is used in Problem (2), we obtain the following mathematical program:

$$\begin{array}{l} \displaystyle\min_{\boldsymbol{w} \in \mathcal{W}} \sum_{i=1}^{m} \sum_{j=1}^{n} |a_j^i - v_j(\boldsymbol{w})| = \\ \displaystyle\min_{\boldsymbol{w} \in \mathcal{W}} \sum_{j=1}^{n} \sum_{i=1}^{m} |a_j^i - v_j(\boldsymbol{w})|. \end{array} \tag{3}$$

On the other hand, it is known that given $x_1, \ldots, x_m \in \mathbb{R}$, the median operator is the solution of the following problem:

$$\min_{x \in \mathbb{R}} \frac{1}{m} \sum_{i=1}^{m} |x_i - x|.$$

Therefore, it is also the solution of

$$\min_{\boldsymbol{w} \in \mathcal{W}} \sum_{i=1}^{m} |a_j^i - v_j(\boldsymbol{w})|.$$

for all $j \in \{1, \ldots, n\}$, and, consequently it is the solution of Problem (3). □

2.3 Using the maximum as aggregation operator

We now consider the maximum as aggregation operator. Thus, with (4) we look for the weighting vector that minimizes the maximum distance between the individual assessments and the collective assessments generated by the OWA operator associated with that weighting vector, i.e.

$$\begin{array}{ll} \text{Min} & \displaystyle\max_{i=1,\ldots,m} \Big\{d\Big((a_1^i, \ldots, a_n^i), (v_1(\boldsymbol{w}), \ldots, v_n(\boldsymbol{w}))\Big)\Big\} \\ \text{s. t. :} & \boldsymbol{w} \in \mathcal{W} \end{array} \tag{4}$$

1. If we use the Manhattan distance, then it is necessary to solve the following mathematical program:

$$\begin{array}{l} \text{Min} \displaystyle\max_{i=1,\ldots,m} \Big\{|a_1^i - v_1(\boldsymbol{w})| + \cdots + |a_n^i - v_n(\boldsymbol{w})|\Big\} \\ \text{s. t. :} \ w_1 \geq 0, \ldots, w_m \geq 0, \ w_1 + \cdots + w_m = 1 \end{array}$$

ISBN: 978-989-95079-6-8

2. If we use the Chebyshev distance, then it is necessary to solve the following mathematical program:

$$\text{Min} \max_{i=1,\ldots,m} \max \Big\{ |a_1^i - v_1(\boldsymbol{w})|, \ldots, |a_n^i - v_n(\boldsymbol{w})| \Big\}$$
$$\text{s. t.: } w_1 \geq 0, \ldots, w_m \geq 0,\ w_1 + \cdots + w_m = 1$$

In this last case, i.e., when the Chebyshev distance is used, the solution of the mathematical program generates the mid-range OWA operator, as we show in the following proposition.

Proposition 2 *The solution of Problem (4) for the Chebyshev distance is the weighting vector* $\boldsymbol{w}^*$ *given by*

$$w_i^* = \begin{cases} 0.5, & \textit{if } i \in \{1, m\}, \\ 0, & \textit{otherwise.} \end{cases}$$

Proof

When the Chebyshev distance is used in Problem (4), we obtain the following mathematical program:

$$\min_{\boldsymbol{w}\in\mathcal{W}} \max_{i=1,\ldots,m} \max_{j=1,\ldots,n} \Big\{ |a_j^i - v_j(\boldsymbol{w})| \Big\} =$$
$$\min_{\boldsymbol{w}\in\mathcal{W}} \max_{j=1,\ldots,n} \max_{i=1,\ldots,m} \Big\{ |a_j^i - v_j(\boldsymbol{w})| \Big\}. \qquad (5)$$

On the other hand, it is obvious that for all $j \in \{1, \ldots, n\}$, the weighting vector given by

$$w_i^* = \begin{cases} 0.5, & \text{if } i \in \{1, m\}, \\ 0, & \text{otherwise,} \end{cases}$$

is the solution of the following problem:

$$\min_{\boldsymbol{w}\in\mathcal{W}} \max_{i=1,\ldots,m} \Big\{ |a_j^i - v_j(\boldsymbol{w})| \Big\}.$$

Therefore, $\boldsymbol{w}^*$ is also the solution of Problem (5). □

2.4 Restricted models

The general model (1) does not impose any restriction to the OWA weighting vectors that maximize the consensus among the individual assessments and the collective one. However, in some cases it could be interesting to consider some requirements for these weights by following a desired pattern, through a subclass $\mathcal{W}^* \subset \mathcal{W}$. For instance, we may restrict the search of the weighting vectors within one of the following cases.

1. Weighting vectors with a fixed *orness* or *attitudinal character* $\alpha \in (0, 1)$ (Yager [9]):

$$\mathcal{W}_\alpha^1 = \left\{ \boldsymbol{w} \in \mathcal{W} \;\middle|\; \frac{1}{m-1} \sum_{i=1}^{m} (m-i) w_i = \alpha \right\}.$$

2. Symmetric weights:

$$\mathcal{W}^2 = \left\{ \boldsymbol{w} \in \mathcal{W} \mid w_i = w_{m+1-i}\ \forall i \in \left\{ 1, \ldots, \left[\frac{m}{2} \right] \right\} \right\}.$$

3. Centered OWAs (after Yager [10]):

$$\mathcal{W}^3 = \left\{ \boldsymbol{w} \in \mathcal{W}^2 \mid w_1 \leq w_2 \leq \cdots \leq w_{\left[\frac{m+1}{2}\right]} \right\}.$$

4. Trimmed OWAs:

$$\mathcal{W}_1^4 = \{ \boldsymbol{w} \in \mathcal{W} \mid w_1 = w_m = 0 \},$$
$$\mathcal{W}_2^4 = \{ \boldsymbol{w} \in \mathcal{W} \mid w_1 = w_2 = w_{m-1} = w_m = 0 \},$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

It is worth noting that $\mathcal{W}^3 \subseteq \mathcal{W}^2 \subseteq \mathcal{W}_{0.5}^1 \subseteq \mathcal{W}$.

Analogously to the general model, consider a distance $d : [0,1]^n \times [0,1]^n \longrightarrow [0,1]$ and an aggregation operator $A : [0,1]^m \longrightarrow [0,1]$. Given a profile $P \in \mathcal{P}$, we now propose to find weighting vector(s) $\boldsymbol{w} \in \mathcal{W}^*$ being solution(s) of the following mathematical program

$$\begin{array}{ll} \text{Min} & A\Big(d(\boldsymbol{a}^1, \boldsymbol{v}(\boldsymbol{w})), \ldots, d(\boldsymbol{a}^m, \boldsymbol{v}(\boldsymbol{w}))\Big) \\ \text{s. t.:} & \boldsymbol{w} \in \mathcal{W}^* \end{array} \qquad (6)$$

Notice that from continuity of A and compactness of $\mathcal{W}^*$ in the previous cases, the existence of solution(s) in (6) is always guaranteed.

3 Some illustrative examples

The following matrix shows the opinions of four experts on three alternatives.

$$P = \begin{pmatrix} 0.7 & 0.6 & 0.1 \\ 0 & 0.5 & 0.8 \\ 0.6 & 0.1 & 1 \\ 0.6 & 0.7 & 0 \end{pmatrix}$$

Case 1. If we use Model (2) with the Chebyshev distance, the following mathematical programming must be solved:

Min
$$\max\{|0.7 - v_1(\boldsymbol{w})|, |0.6 - v_2(\boldsymbol{w})|, |0.1 - v_3(\boldsymbol{w})|\} +$$
$$\max\{|0 - v_1(\boldsymbol{w})|, |0.5 - v_2(\boldsymbol{w})|, |0.8 - v_3(\boldsymbol{w})|\} +$$
$$\max\{|0.6 - v_1(\boldsymbol{w})|, |0.1 - v_2(\boldsymbol{w})|, |1 - v_3(\boldsymbol{w})|\} +$$
$$\max\{|0.6 - v_1(\boldsymbol{w})|, |0.7 - v_2(\boldsymbol{w})|, |0 - v_3(\boldsymbol{w})|\}$$

$$\text{s. t.: } w_1 \geq 0,\ w_2 \geq 0,\ w_3 \geq 0,\ w_4 \geq 0,$$
$$w_1 + w_2 + w_3 + w_4 = 1,$$

where

$$v_1(\boldsymbol{w}) = 0.7w_1 + 0.6w_2 + 0.6w_3,$$
$$v_2(\boldsymbol{w}) = 0.7w_1 + 0.6w_2 + 0.5w_3 + 0.1w_4,$$
$$v_3(\boldsymbol{w}) = 1w_1 + 0.8w_2 + 0.1w_3.$$

The previous problem is a non-smooth optimization problem. However, it can be easily replaced by an equivalent smooth linear problem by using some auxiliary variables. We have used LINGO software to solve this linear problem. The solution obtained is showed in Table 1. Likewise, that table summarizes the solutions obtained when some restrictions are imposed to the OWA weighting vectors in the problem.

ISBN: 978-989-95079-6-8

Table 1: Solutions for Case 1.

	w_1	w_2	w_3	w_4
$\mathcal{W}$	0.4706	0	0	0.5294
$\mathcal{W}^1_{0.75}$	0.75	0	0	0.25
$\mathcal{W}^1_{0.5}$	0.5	0	0	0.5
$\mathcal{W}^1_{0.25}$	0	0.375	0	0.625
$\mathcal{W}^2$	0.5	0	0	0.5
$\mathcal{W}^3$	0.25	0.25	0.25	0.25
$\mathcal{W}^4_1$	0	0.1429	0.8571	0

Case 2. If we consider Model (4) with the Manhattan distance, the mathematical programming to solve is:

$$\begin{aligned}\text{Min} \quad & \max\Big\{ \\ & |0.7 - v_1(\boldsymbol{w})| + |0.6 - v_2(\boldsymbol{w})| + |0.1 - v_3(\boldsymbol{w})|, \\ & |0 - v_1(\boldsymbol{w})| + |0.5 - v_2(\boldsymbol{w})| + |0.8 - v_3(\boldsymbol{w})|, \\ & |0.6 - v_1(\boldsymbol{w})| + |0.1 - v_2(\boldsymbol{w})| + |1 - v_3(\boldsymbol{w})|, \\ & |0.6 - v_1(\boldsymbol{w})| + |0.7 - v_2(\boldsymbol{w})| + |0 - v_3(\boldsymbol{w})|\Big\} \\ \text{s. t. :} \quad & w_1 \geq 0,\ w_2 \geq 0,\ w_3 \geq 0,\ w_4 \geq 0, \\ & w_1 + w_2 + w_3 + w_4 = 1,\end{aligned}$$

where

$$\begin{aligned} v_1(\boldsymbol{w}) &= 0.7w_1 + 0.6w_2 + 0.6w_3, \\ v_2(\boldsymbol{w}) &= 0.7w_1 + 0.6w_2 + 0.5w_3 + 0.1w_4, \\ v_3(\boldsymbol{w}) &= 1w_1 + 0.8w_2 + 0.1w_3. \end{aligned}$$

As in Case 1, this problem can be also substituted by an equivalent smooth linear problem by introducing some auxiliar variables. Table 2 shows the solution obtained by using LINGO software for different requirements on the weights.

Table 2: Solutions for Case 2.

	w_1	w_2	w_3	w_4
$\mathcal{W}$	0.667	0	0.222	0.111
$\mathcal{W}^1_{0.75}$	0.673	0	0.2307	0.0963
$\mathcal{W}^1_{0.5}$	0	0.722	0.056	0.222
$\mathcal{W}^1_{0.25}$	0	0.375	0	0.625
$\mathcal{W}^2$	0	0.5	0.5	0
$\mathcal{W}^3$	0	0.5	0.5	0
$\mathcal{W}^4_1$	0	0.834	0.166	0

4 Further research

It is worth noting that the solutions obtained in the above described models might not be unique. Even more, depending the weighting vector we choose, the outcomes provided by the corresponding OWA operator could be different. In this way, it would be necessary to provide an appropriate procedure for choosing a single weighting vector among the set of multiple solutions. This open problem constitutes part of our further research.

Acknowledgment

This work is partially supported by the Junta de Castilla y León (Consejería de Educación y Cultura, Projects VA092A08 and VA002B08), the Spanish Ministry of Education and Science (Project SEJ2006-04267/ECON), and ERDF.

References

[1] G. Beliakov, A. Pradera and T. Calvo. *Aggregation Functions: A Guide for Practicioners*. Springer, Heidelberg, 2007.

[2] T. Calvo, A. Kolesárova, M. Komorníková and R. Mesiar. Aggregation operators: Properties, classes and construction methods, in *Aggregation Operators: New Trends and Applications*, eds. T. Calvo, G. Mayor and R. Mesiar. Physica-Verlag, Heidelberg, 2002, pp. 3–104.

[3] D. Filev and R.R. Yager. On the issue of obtaining OWA operator weights. *Fuzzy Sets and Systems*, 94:157–169, 1998.

[4] J. Fodor and M. Roubens. *Fuzzy Preference Modelling and Multicriteria Decision Support*. Kluwer Academic Publishers, Dordrecht, 1994.

[5] M. Grabisch, S.A. Orlovski and R.R. Yager. Fuzzy aggregation of numerical preferences, in *Fuzzy Sets in Decision Analysis, Operations and Statistics*, ed. R. Slowinski. Kluwer Academic Publishers, Boston, 1998, pp. 31–68.

[6] M. O'Hagan. Aggregating template rule antecedents in real-time expert systems with fuzzy set logic. *Proc. 22nd Annual IEEE Asilomar Conference on Signals, Systems and Computers*. Pacific Grove, California, 1988, pp. 681–689.

[7] Y.M. Wang and C. Parkan. A minimax disparity approach for obtaining OWA operator weights. *Information Sciences*, 175:20–29, 2005.

[8] Z. Xu. An overview of methods for determining OWA weights. *International Journal of Intelligent Systems*, 20:843–865, 2005.

[9] R.R. Yager. On ordered weighted averaging operators in multicriteria decision making. *IEEE Transactions on Systems, Man and Cybernetics*, 18:183–190, 1988.

[10] R.R. Yager. Centered OWA operators. *Soft Computing*, 11:631–639, 2007.

ISBN: 978-989-95079-6-8

Weighted Trapezoidal Approximations of Fuzzy Numbers

Przemysław Grzegorzewski[1,2] Karolina Pasternak-Winiarska[2]

1.Systems Research Institute, Polish Academy of Sciences
Newelska 6, 01-447 Warsaw, Poland
2.Faculty of Mathematics and Information Science, Warsaw University of Technology
Plac Politechniki 1, 00-661 Warsaw, Poland
Email: pgrzeg@ibspan.waw.pl, karolcia.pw@gmail.com

Abstract— *Fuzzy number approximation by trapezoidal fuzzy numbers which preserves the expected interval is discussed. New algorithms for calculating the proper trapezoidal approximation of fuzzy numbers with respect to the distance based on bi-symmetrical weighted functions are proposed. It is shown that the adequate approximation operator is chosen with respect both to the global spread of a fuzzy number and the size of possible asymmetry between the spread of the left-hand and right-hand part of a fuzzy number.*

Keywords— fuzzy number, expected interval, trapezoidal approximation, width, weighted distance.

1 Introduction

Trapezoidal approximation of fuzzy numbers was considered by many authors (see, e.g. [1, 2, 3, 5, 11, 12, 9, 10, 15]). In [11] a list of criteria which trapezoidal approximation operators should possess was formulated. Then [18] considered the trapezoidal approximation under weighted distance. Some gaps in that paper were removed in [16] and [17] where so-called extended trapezoidal fuzzy numbers were introduced and applied.

The most common weighing applied to fuzzy numbers is a linearly increasing one (see, e.g. [5, 18]). However, it seems that a slightly modified weighing would be more natural and convenient in many situations. Such weighted functions, called regular bi-symmetrical weighted functions, were introduced and the nearest trapezoidal approximation operator preserving the expected interval, with respect to the distance based on such weighted functions, was suggested.

The paper is organized as follows. Firstly we recall some basic notions related to fuzzy numbers and present a few principle ideas connected with trapezoidal approximation. Then in Sec. 3 we introduce the notion of bi-symmetrical weighted functions. Finally, in Sec 4, we show trapezoidal approximation operators based on the weighted distance utilizing bi-symmetrical weighted functions and discuss some properties of the above mentioned operators.

2 Basic concepts

Let A denote a fuzzy number, i.e. such fuzzy subset A of the real line $\mathbb{R}$ with membership function $\mu_A : \mathbb{R} \to [0,1]$ which is (see [6]): normal (i.e. there exist an element x_0 such that $\mu_A(x_0) = 1$), fuzzy convex (i.e. $\mu_A(\lambda x_1 + (1-\lambda)x_2) \geq \mu_A(x_1) \wedge \mu_A(x_2)$, $\forall x_1, x_2 \in \mathbb{R}$, $\forall \lambda \in [0,1]$), μ_A is upper semicontinuous, supp$A$ is bounded, where supp$A = cl(\{x \in \mathbb{R} : \mu_A(x) > 0\})$, and cl is the closure operator. A space of all fuzzy numbers will be denoted by $\mathbb{F}(\mathbb{R})$.

Moreover, let $A_\alpha = \{x \in \mathbb{R} : \mu_A(x) \geq \alpha\}$, $\alpha \in (0,1]$, denote an α-cut of a fuzzy number A. As it is known, every α-cut of a fuzzy number is a closed interval, i.e. $A_\alpha = [A_L(\alpha), A_U(\alpha)]$, where $A_L(\alpha) = \inf\{x \in \mathbb{R} : \mu_A(x) \geq \alpha\}$ and $A_U(\alpha) = \sup\{x \in \mathbb{R} : \mu_A(x) \geq \alpha\}$.

The expected interval $EI(A)$ of a fuzzy number A is given by (see [7, 13])

$$\begin{aligned} EI(A) &= [EI_L(A), EI_U(A)] \\ &= \left[\int_0^1 A_L(\alpha)d\alpha, \int_0^1 A_U(\alpha)d\alpha\right]. \end{aligned} \tag{1}$$

The middle point of the expected interval given by

$$EV(A) = \frac{1}{2}\left(\int_0^1 A_L(\alpha)d\alpha + \int_0^1 A_U(\alpha)d\alpha\right) \tag{2}$$

is called the *expected value* of a fuzzy number and it represents the typical value of the fuzzy number A (see [7, 13]). Sometimes its generalization, called *weighted expected value*, might be interesting. It is defined as

$$EV_q(A) = (1-q)\int_0^1 A_L(\alpha)d\alpha + q\int_0^1 A_U(\alpha)d\alpha, \tag{3}$$

where $q \in [0,1]$ (see [8]).

Another useful parameter characterizing a fuzzy number is called the *width* of a fuzzy number (see [4]) and is defined by

$$w(A) = \int_{-\infty}^{\infty} \mu_A(x)dx = \int_0^1 (A_U(\alpha) - A_L(\alpha))d\alpha. \tag{4}$$

Suppose that for a certain reason or just for simplicity we want to find a suitable approximation of a fuzzy number under study. It seems that sufficiently effective simplification of the fuzzy number shape can be reached by the piecewise linear curves leading to triangle, trapezoidal or orthogonal membership functions. Since these three mentioned shapes are particular cases of the trapezoidal one, further on we will consider just the trapezoidal approximation of fuzzy numbers. It means that we want to substitute given fuzzy number A by the trapezoidal fuzzy number $T(A)$, i.e. by a fuzzy number with a following membership function

$$\mu_{T(A)}(x) = \begin{cases} 0 & \text{if} \quad x < t_1, \\ \frac{x-t_1}{t_2-t_1} & \text{if} \quad t_1 \leq x < t_2, \\ 1 & \text{if} \quad t_2 \leq x \leq t_3, \\ \frac{t_4-x}{t_4-t_3} & \text{if} \quad t_3 < x \leq t_4, \\ 0 & \text{if} \quad t_4 < x. \end{cases} \tag{5}$$

ISBN: 978-989-95079-6-8

Since we can do this in many ways we need some additional constraints which guarantee that our approximation would be reasonable. One of the most natural idea is to construct $T(A)$ that is the closest to the original fuzzy number A with respect to given distance d. Sometimes we add additional requirements which warrant that our approximation would possess some desired properties, like preservation fixed parameters or relations, continuity, etc. This problem was considered by many authors (see, e.g. [1, 2, 3, 5, 11, 12, 9, 10, 15, 16]). For example, it was suggested in [11] to consider *the nearest trapezoidal approximation operator preserving the expected interval*, i.e. the approximation operator T which produces a trapezoidal fuzzy number $T(A)$ that is the closest with respect to distance

$$d(A,T(A)) = \left(\int_0^1 [A_L(\alpha)-T(A)_L(\alpha)]^2 d\alpha + \int_0^1 [A_U(\alpha)-T(A)_U(\alpha)]^2 d\alpha\right)^{1/2} \quad (6)$$

to given original fuzzy number A among all trapezoidal fuzzy numbers having identical expected interval as the original one, i.e. satisfying a following condition

$$EI(T(A)) = EI(A). \quad (7)$$

It is worth noting the invariance of the expected interval assures many other properties (for more details we refer the reader to [11] where the broad list of desired requirements that the approximation operator should possess is also given). The research on this operator was continued in [3, 12, 9, 10].

Whatever trapezoidal approximation is considered the goal reduces to finding such real numbers $t_1 \le t_2 \le t_3 \le t_4$ that characterize $T(A) = T(t_1,t_2,t_3,t_4)$. It is so because any trapezoidal fuzzy number is completely described by four real numbers that are borders of its support and core. Let us mention that mathematical formulae for these points corresponding to the nearest trapezoidal approximation operator preserving the expected interval are given in [9].

In some applications other distances than (6) are more suitable. It is easily seen that all α-cuts in (6) are treated evenly. This feature is sometimes criticized by authors who claim that elements belonging to α_1-cut should be treated with the higher attention that those from α_2-cut if $\alpha_1 > \alpha_2$ because the membership degree for the first group is higher and so they are less uncertain. Such reasoning in trapezoidal approximation can be found in [18] devoted to weighted trapezoidal approximation. More precisely, the authors consider there a trapezoidal approximation with respect to the weighted distance

$$d_{ZL}(A,T(A)) = \left(\int_0^1 \alpha[A_L(\alpha)-T(A)_L(\alpha)]^2 d\alpha + \int_0^1 \alpha[A_U(\alpha)-T(A)_U(\alpha)]^2 d\alpha\right)^{1/2} \quad (8)$$

with increasing weighting function.

3 A bi-symmetrical weighted distance

Although such increasing weighting might be useful in some occasions, another weighted distances would be more interesting in general. This is a straightforward conclusion from the fact that the least informative α-cut is not zero but 0.5. Actually, situation $\mu_A(x) = 1$ leads to perfect information that x surely belongs to A. If $\mu_A(x)$ is close to 1 we'll say that x rather belongs to A. And conversely, $\mu_A(x) = 0$ shows that x surely does not belong to A (and belongs to $\neg A$) which is also a perfect information. Similarly, x such that $\mu_A(x)$ is close to 0 is interpreted as a point that rather does not belong to A. However, if $\mu_A(x) = 0.5$ we dot know how to classify x because it belong to A and to it's completion $\neg A$ with the same degree. The same happens if $\mu_A(x)$ is close to 0.5. Thus, to sum up, degrees of membership both high (close to 1) and low (close to 0) are much more informative than those close to 0.5. Hence, if we try to incorporate this quite obvious remark into practice we have to consider a so-called bi-symmetrical weighted distance suggested below. Before defining this distance we will introduce the notion of the bi-symmetrical weighted function.

Definition 1
A function $\lambda : [0,1] \to [0,1]$ symmetrical around $\frac{1}{2}$, i.e. $\lambda(\frac{1}{2}-\alpha) = \lambda(\frac{1}{2}+\alpha)$ for all $\alpha \in [0,\frac{1}{2}]$, which reaches its minimum in $\frac{1}{2}$, is called the bi-symmetrical weighted function. Moreover, the bi-symmetrical weighted function is called regular if

(a) $\lambda(\frac{1}{2}) = 0$,

(b) $\lambda(0) = \lambda(1) = 1$,

(c) $\int_0^1 \lambda(\alpha)d\alpha = \frac{1}{2}$.

Definition 2
For two arbitrary fuzzy numbers A and B with α-cuts $[A_L(\alpha), A_U(\alpha)]$ and $[B_L(\alpha), B_U(\alpha)]$, respectively, the quantity

$$d_\lambda(A,T(A)) = \left(\int_0^1 \lambda(\alpha)[A_L(\alpha)-B_L(\alpha)]^2 d\alpha + \int_0^1 \lambda(\alpha)[A_U(\alpha)-B_U(\alpha)]^2 d\alpha\right)^{1/2} \quad (9)$$

where $\lambda : [0,1] \to [0,1]$ is a bi-symmetrical (regular) weighted function is called the bi-symmetrical (regular) weighted distance between A and B based on λ.

One can, of course, propose many regular bi-symmetrical weighted functions and hence obtain different bi-symmetrical weighted distances. Further on we will consider mainly a following function

$$\lambda(\alpha) = \begin{cases} 1-2\alpha & \text{if} \quad \alpha \in [0,\frac{1}{2}], \\ 2\alpha-1 & \text{if} \quad \alpha \in [\frac{1}{2},1], \end{cases} \quad (10)$$

which is in some sense a bi-symmetrical counterpart of the increasing weighted function such as applied in [18].

4 Trapezoidal approximations based on bi-symmetrical weighted distances

Let us go back to the trapezoidal approximation operators $T : \mathbb{F}(\mathbb{R}) \to \mathbb{F}^T(\mathbb{R})$ which produce a trapezoidal fuzzy number $T(A)$ that is the closest to given original fuzzy number

ISBN: 978-989-95079-6-8

A among all trapezoidal fuzzy numbers having identical expected interval as the original one, i.e. satisfying (7). However, now we will look for the operators which minimize the bi-symmetrical weighted distance based on bi-symmetrical function (10).

It is easily seen that the α-cut of $T(A)$ is equal to $[t_1 + (t_2 - t_1)\,\alpha, t_4 - (t_4 - t_3)\alpha]$. Since a trapezoidal fuzzy number is completely described by four real numbers that are borders of its support and core, our goal reduces to finding such real numbers $t_1 \leq t_2 \leq t_3 \leq t_4$ that characterize $T(A) = T(t_1, t_2, t_3, t_4)$. Substituting it into (9) and (7) our problem might be expressed as follows: find t_1, t_2, t_3, t_4 which minimize

$$d(A, T(A)) = \tag{11}$$
$$= \left(\int_0^1 \lambda(\alpha) \left[A_L(\alpha) - (t_1 + (t_2 - t_1)\,\alpha) \right]^2 d\alpha \right.$$
$$\left. + \int_0^1 \lambda(\alpha) \left[A_U(\alpha) - (t_4 - (t_4 - t_3)\alpha) \right]^2 d\alpha \right)^{1/2}$$

with respect to conditions

$$\frac{t_1 + t_2}{2} = \int_0^1 A_L(\alpha) d\alpha, \tag{12}$$
$$\frac{t_3 + t_4}{2} = \int_0^1 A_U(\alpha) d\alpha \tag{13}$$
$$t_1 \leq t_2 \leq t_3 \leq t_4. \tag{14}$$

Before showing final results let us introduce some notation. Firstly let us notice that by Def. 1 the centroid of a bi-symmetrical weighted function λ is $\frac{1}{2}$. Therefore, the dispersion of the bi-symmetrical weighted function λ is given by

$$\eta = \int_0^1 (\alpha - \frac{1}{2})^2 \lambda(\alpha)\; d\alpha. \tag{15}$$

Easy computation shows that the dispersion of our bi-symmetrical weighted function (10) is $\frac{1}{16}$.

We also introduce another two parameters characterizing the dispersion of the left side and of the right side of given fuzzy number A with respect to considered bi-symmetrical weighted function λ.

Definition 3
The left (lower) spread of a fuzzy number A with respect to considered bi-symmetrical weighted function λ is a number $LSP_\lambda(A)$ given by

$$LSP_\lambda(A) = \frac{1}{2\eta} \int_0^1 (\alpha - \frac{1}{2}) \lambda(\alpha) A_L(\alpha)\; d\alpha, \tag{16}$$

while the right (upper) spread of a fuzzy number A with respect to considered bi-symmetrical weighted function λ is a following number $USP_\lambda(A)$

$$USP_\lambda(A) = \frac{1}{2\eta} \int_0^1 (\frac{1}{2} - \alpha) \lambda(\alpha) A_U(\alpha)\; d\alpha. \tag{17}$$

It can be shown that $LSP_\lambda(A) \geq 0$ and $USP_\lambda(A) \geq 0$.

Let us also denote the total spread of given fuzzy number A with respect to considered bi-symmetrical weighted function λ by $TSP_\lambda(A)$, i.e.

$$TSP_\lambda(A) = LSP_\lambda(A) + USP_\lambda(A), \tag{18}$$

while the difference between the right and the left spread of a fuzzy number will be denoted by $\Delta SP_\lambda(A)$, i.e.

$$\Delta SP_\lambda(A) = USP_\lambda(A) - LSP_\lambda(A). \tag{19}$$

It is easily seen that as $TSP_\lambda(A)$ is always nonnegative, $\Delta SP_\lambda(A)$ might be positive or negative as A is more asymmetrical to the right or to the left.

We can now formulate our main result.

Theorem 4
The nearest trapezoidal approximation operator preserving expected interval with respect to distance (11) for bi-symmetrical weighted function λ given by (10) is such operator $T : \mathbb{F}(\mathbb{R}) \to \mathbb{F}^T(\mathbb{R})$, which assigns a following trapezoidal fuzzy number $T(A) = T(t_1, t_2, t_3, t_4)$ for any fuzzy number A with α-cuts $[A_L(\alpha), A_U(\alpha)]$

(a) if $w(A) \geq TSP_\lambda(A)$ then

$$t_1 = EI_L(A) - LSP_\lambda(A), \tag{20}$$
$$t_2 = EI_L(A) + LSP_\lambda(A), \tag{21}$$
$$t_3 = EI_U(A) - USP_\lambda(A), \tag{22}$$
$$t_4 = EI_U(A) + USP_\lambda(A); \tag{23}$$

(b) if $|\Delta SP_\lambda(A)| \leq w(A) < TSP_\lambda(A)$ then

$$t_1 = EI_L(A) - \frac{1}{2} w(A) + \frac{1}{2} \Delta SP_\lambda(A), \tag{24}$$
$$t_2 = t_3 = EV(A) - \frac{1}{2} \Delta SP_\lambda(A), \tag{25}$$
$$t_4 = EI_U(A) + \frac{1}{2} w(A) + \frac{1}{2} \Delta SP_\lambda(A); \tag{26}$$

(c) if $w(A) < \Delta SP_\lambda(A)$ then

$$t_1 = t_2 = t_3 = EI_L(A), \tag{27}$$
$$t_4 = 2EI_U(A) - EI_L(A); \tag{28}$$

(d) if $w(A) < -\Delta SP_\lambda(A)$ then

$$t_1 = 2EI_L(A) - EI_U(A), \tag{29}$$
$$t_2 = t_3 = t_4 = EI_U(A). \tag{30}$$

The prove this result we apply the Karush-Kuhn-Tucker theorem for the local minimizer under given constraints.

Actually, we have received four different operators providing the nearest trapezoidal fuzzy number that preserves the expected value of the original fuzzy number, where T_1 leads to trapezoidal fuzzy number, T_2 stands for the operator that leads to triangular fuzzy number with two sides, while T_3 and T_4 produce triangular fuzzy numbers with the right side only or with the left side only, respectively.

Roughly speaking, we approximate a fuzzy number A by the trapezoidal approximation operator T_1 provided the total dispersion of given fuzzy number with respect to considered bi-symmetrical weighted function, measured by the sum of the lower and upper spread, is large enough. Otherwise, we will approximate A by a triangular number. It means that for less dispersed fuzzy numbers the solution is always a triangular fuzzy number.

ISBN: 978-989-95079-6-8

However, here we have also three possible situations: to approximate a fuzzy number A we apply operator T_2 provided the asymmetry of A is not too big (i.e.there is no big difference between the lower and upper spread). If A reveals high right asymmetry (i.e. the right spread is significantly larger than the lower spread) it would be approximated by a triangular fuzzy number with the right side only, produced by operator T_3. Otherwise, a fuzzy number with high left asymmetry would be approximated by a triangular fuzzy number with the left side only, produced by operator T_4.

It is worth noticing that we have obtained four possible solutions like in the problem with non-weighted distance considered in [3] or [9]. Moreover, operators T_3 and T_4 are identical as in [3] or [9]. It means that for very asymmetrical fuzzy numbers its nearest trapezoidal approximation preserving the expected interval remains independent whether we use weighted or non-weighted distance.

Our operators possess many desired properties, like invariance to translations and scale, monotonicity, continuity, etc. Some of those properties are guaranteed by the expected interval invariance. For the detailed list of requirements related to trapezoidal approximations we refer the reader to [11].

One may ask what happen if we use a different bi-symmetrical weighted function. Thus now, contrary to previous continuous weighted function (10) let us consider a non-continuous one

$$\lambda_1(\alpha) = \begin{cases} 1 & \text{if} \quad \alpha \in [0, \frac{1}{4}] \cup [\frac{3}{4}, 1], \\ 0 & \text{if} \quad \alpha \in (\frac{1}{4}, \frac{3}{4}), \end{cases} \tag{31}$$

which appreciates only elements with high or low degree of membership and does not take into account the other. Thus λ_1 corresponds to Pedrycz's viewpoint expressed in his shadowed sets (see [14]) where we consider only these points which rather belong to a set under study or those that rather do not belong to it. The other elements with intermediate membership degree form the so-called shadow. Easy computation shows that the dispersion of our bi-symmetrical weighted function (31) is $\frac{7}{96}$. It is interesting and worth stressing that the final solution both for λ_1 is identical with that obtained before, i.e. a following theorem holds.

Theorem 5
The nearest trapezoidal approximation operator preserving expected interval with respect to distance (11) for bi-symmetrical weighted function λ given by (31) is such operator $T : \mathbb{F}(\mathbb{R}) \to \mathbb{F}^T(\mathbb{R})$, which assigns a following trapezoidal fuzzy number $T(A) = T(t_1, t_2, t_3, t_4)$ for any fuzzy number A with α-cuts $[A_L(\alpha), A_U(\alpha)]$, has the same form as for for bi-symmetrical weighted function (10), i.e. given in Theorem 4.

5 Conclusion

In the present contribution we have considered the problem of trapezoidal approximation of fuzzy numbers for bi-symmetrical weighted functions. It was shown that the choice of adequate approximation operator depends mainly on the global spread of a fuzzy number and possible asymmetry between the spread of the left-hand and right-hand part of the original fuzzy number. It is interesting that identical operators for two different bi-symmetrical weighted functions were obtained. Thus one may ask whether the solution remains valid for other weighted function too. This would be the topic of our further research.

References

[1] S. Abbasbandy, B. Asady. The nearest approximation of a fuzzy quantity in parametric form. *Applied Mathematics and Computation*, 172:624-632, 2006.

[2] S. Abbasbandy, M. Amirfakhrian. The nearest trapezoidal form of a generalized LR fuzzy number. *International Journal of Approximate Reasoning*, 43:166-178, 2006.

[3] A. Ban. Approximation of fuzzy numbers by trapezoidal fuzzy numbers preserving the expected interval. *Fuzzy Sets and Systems* 159:1327-1344, 2008.

[4] S. Chanas. On the interval approximation of a fuzzy number. *Fuzzy Sets and Systems*, 122:353–356, 2001.

[5] M. Delgado, M.A. Vila, W. Voxman. On a canonical representation of a fuzzy number. *Fuzzy Sets and Systems*, 93:125–135, 1998.

[6] D. Dubois, H. Prade. Operations on fuzzy numbers. *Int. J. Syst. Sci.*, 9:613–626, 1978.

[7] D. Dubois, H. Prade. The mean value of a fuzzy number. *Fuzzy Sets and Systems*, 24:279–300, 1987.

[8] P. Grzegorzewski. Metrics and orders in space of fuzzy numbers. *Fuzzy Sets and Systems*, 97:83–94, 1998.

[9] P. Grzegorzewski. Trapezoidal approximations of fuzzy numbers preserving the expected interval - algorithms and properties. *Fuzzy Sets and Systems*, 159:1354-1364, 2008.

[10] P. Grzegorzewski. New algorithms for trapezoidal approximation of fuzzy numbers preserving the expected interval. In: *Proceedings of the Twelfth International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU'08*, L. Magdalena, M. Ojeda-Aciego, J.L. Verdegay (Eds.), Spain, Torremolinos (Málaga), June 22-27, pp. 117-123, 2008.

[11] P. Grzegorzewski, E. Mrówka. Trapezoidal approximations of fuzzy numbers. *Fuzzy Sets and Systems*, 153:115–135, 2005.

[12] P. Grzegorzewski, E. Mrówka. Trapezoidal approximations of fuzzy numbers - revisited. *Fuzzy Sets and Systems*, 158:757–768, 2007.

[13] S. Heilpern. The expected value of a fuzzy number. *Fuzzy Sets and Systems*, 47:81–86, 1992.

[14] W. Pedrycz. Shadowed sets: representing and processing fuzzy sets. *IEEE Trans. Syst. Man. Cybern. B*, 2:103–109, 1998.

[15] C.T. Yeh. A note on trapezoidal approximations of fuzzy numbers. *Fuzzy Sets and Systems*, 158:747–754, 2007.

[16] C.T. Yeh. Trapezoidal and triangular approximations preserving the expected interval. *Fuzzy Sets and Systems*, 159:1345-1353, 2008.

[17] C.T. Yeh. On improving trapezoidal and triangular approximations of fuzzy numbers. *International Journal of Approximate Reasoning*, 48:297-313, 2008.

[18] W. Zeng, H. Li. Weighted triangular approximation of fuzzy numbers. *International Journal of Approximate Reasoning*, 46:137-150, 2007.

ISBN: 978-989-95079-6-8

Nonlinear-shaped Approximation of Fuzzy Numbers

Przemyslaw Grzegorzewski [1,2], Luciano Stefanini [3]

1. Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland.
2. Faculty of Mathematics and Information Sciences, Warsaw University of Technology, Warsaw, Poland.
3. Faculty of Economics, University of Urbino "Carlo Bo", Urbino, Italy.
Email: pgrzeg@ibspan.waw.pl, lucste@uniurb.it

Abstract— *In this paper, we present a general framework for the nearest approximation of fuzzy numbers simultaneously valid for the LR or the LU representations. We suggest two families of flexible monotonic functions characterized by four or five parameters that allow multiple approximation criteria and satisfy various requirements such as preservation of value, ambiguity, support, core, level set average, etc.*

Keywords— Trapezoidal Fuzzy Numbers, Approximation of Fuzzy Numbers, LR fuzzy numbers, Expected Interval, Shape Preserving Approximation

1 Introduction

The approximation of fuzzy numbers by linear or other simple shapes (e.g. polynomial, Gaussian, logistic) is an important problem in many fields and applications where extended or massive fuzzy computations are required. The recent literature has addressed both the linear case (see [3], [9], [10], [11], [12], [13], [18] and the references therein) and the reconstruction by nonlinear functions (e.g. [1], [2], [4], [15]).

Clearly, there are at least three key points to define:

(a) the basic general model to represent fuzzy numbers such as the LR (see [6], [7]) or the LU (see [8], [16], [17]) representations,

(b) the families of parameterized monotonic functions to be used which are sufficiently flexible to cover a large set of possible curves of membership functions,

(c) the criteria for the approximation, depending usually on the application at hand, such as least squares or other distance minimization, support or core or expected interval preservation, nearest ambiguity or value approximation among others (see [12]).

2 Basic representations and tools

In this paper, we present a general framework for the nearest approximation of fuzzy numbers valid simultaneously for the LR or the LU representations. To reach this goal we'll use special families of flexible monotonic curves described by sufficiently large number of parameters to allow multiple approximation criteria and satisfy various requirements.

Let us consider the following two families of nonlinear monotonic functions to use as shape generators ([16]).

1. the (2,2)-rational standardized monotonic spline:

$$p_{R2}(t;\beta_0,\beta_1) = \frac{t^2+\beta_0 t(1-t)}{1+(\beta_0+\beta_1-2)t(1-t)}, \qquad (1)$$

where $\beta_0,\beta_1 \geq 0$ and $t \in [0,1]$;

2. the mixed cubic-exponential spline

$$p_{MS}(t;\beta_0,\beta_1) = \frac{1}{a}\left[t^2(3-2t)+\beta_0 -\beta_0(1-t)^a+\beta_1 t^a\right], \qquad (2)$$

where $a = 1+\beta_0+\beta_1$, $\beta_0,\beta_1 \geq 0$ and $t \in [0,1]$.

It is easy to verify that for arbitrary nonnegative values of parameters β_0,β_1 the two families of functions are monotonic for all $t \in [0,1]$ and satisfy following conditions (the derivatives p' are made with respect to the first variable t):

$$p(0;\beta_0,\beta_1)=0, \qquad p(1;\beta_0,\beta_1)=1,$$
$$p\prime(0;\beta_0,\beta_1)=\beta_0, \qquad p\prime(1;\beta_0,\beta_1)=\beta_1.$$

Remark 1 All the proposed shape functions are monotonic over $[0,1]$. Note that it is not true for standard splines or other polynomials of degree greater than two.

Remark 2 Some special cases are of interest, i.e.

- if $\beta_0=\beta_1=1$ both p_{R2} and p_{MS} are linear;
- if $\beta_0+\beta_1=2$ both p_{R2} and p_{MS} are quadratic;
- if $\beta_0+\beta_1$ is integer (or zero), then p_{MS} is a polynomial function (increasing for all $t \in [0,1]$).

The properties given by following lemma are important.

Lemma 1 *For both p_{R2} and p_{MS} we have*

(i) $0 < \int_0^1 p(t;\beta_0,\beta_1)dt < 1 \quad \forall \beta_0,\beta_1 \geq 0$

(ii) $\lim_{\beta_1 \longrightarrow +\infty} \int_0^1 p(t;0,\beta_1)dt = 0$

(iii) $\lim_{\beta_0 \longrightarrow +\infty} \int_0^1 p(t;\beta_0,0)dt = 1.$

By Lemma 1 both p_{R2} and p_{MS} are able to "cover" the whole box $[0,1]\times[0,1]$ giving an infinite number of shape functions.

Some particular cases of these functions are of interest to meet specific requirements. For example, $p_{R2}(\frac{1}{2};\beta_0,\beta_1)=\frac{1}{2}$ if and only if $\beta_0=\beta_1=\beta$ and then $\int_0^1 p_{R2}(t;\beta,\beta)dt=\frac{1}{2}$ for all $\beta \geq 0$.

ISBN: 978-989-95079-6-8

A family of curves obtained from $p_{R2}(t;\beta_0,\beta_1)$ for different β is given in Fig. 1.

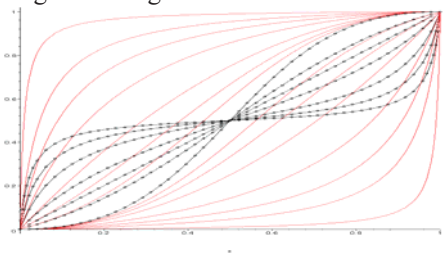


Figure 1. $p_{R2}(t;\beta_0,\beta_1)$ for different β_0,β_1; pointed lines are with $\beta_0=\beta_1$.

We can also use other single-parameter flexible forms and their combinations, like $p_{R2,1}(t,a)=p_{R2}(t;a,0)$, i.e.

$$p_{R2,1}(t,a)=\frac{t^2-at^2+at}{1+at-at^2-2t+2t^2}, \tag{3}$$

where $t\in[0,1]$, $a\geq 0$ or $p_{R2,2}(t,b)=p_{R2}(t;0,b)$, i.e.

$$p_{R2,2}(t,b)=\frac{t^2}{1+bt-bt^2-2t+2t^2}, \tag{4}$$

where $t\in[0,1]$, $b\geq 0$. Possible shapes obtained for different values of parameters a and b are shown in Fig. 2 and Fig. 3, respectively.

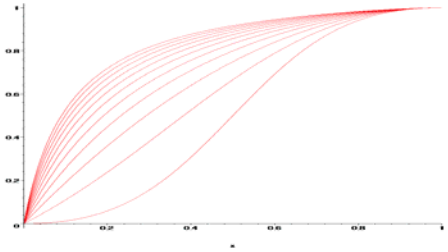


Figure 2. $p_{R2,1}(t;a)$ for different $a\geq 0$

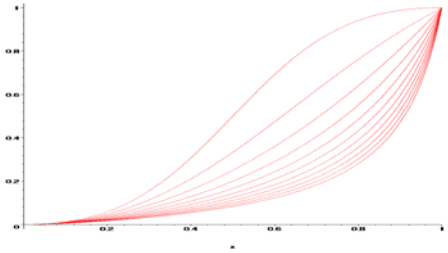


Figure 3. $p_{R2,2}(t;b)$ for different $b\geq 0$

Note that $p_{R2,1}(t;a)+p_{R2,2}(1-t;a)=1$ for all $t\in[0,1]$.

We can also combine the two using (here $\lambda\in[0,1]$)

$$P_{R2,\lambda}(t;a,b)=(1-\lambda)p_{R2,1}(t;a)+\lambda p_{R2,2}(t;b) \tag{5}$$

and have three parameters available.

Analogous constructions can be obtained by using p_{MS}, like $p_{MS,1}(t,a)=p_{MS}(t;a,0)$, i.e.

$$p_{MS,1}(t,a)=\frac{1}{1+a}[t^2(3-2t)+a-a(1-t)^{1+a}], \tag{6}$$

where $t\in[0,1]$, $a\geq 0$ or $p_{MS,2}(t,b)=p_{MS}(t;0,b)$, i.e.

$$p_{MS,2}(t,b)=\frac{1}{1+b}[t^2(3-2t)+bt^{1+b}], \tag{7}$$

where $t\in[0,1]$, $b\geq 0$, and combine the two with $\lambda\in[0,1]$:

$$P_{MS,\lambda}(t;a,b)=(1-\lambda)p_{MS,1}(t;a)+\lambda p_{MS,2}(t;b). \tag{8}$$

Examples of curves $P_{MS,0.5}(t;a,b)$ are in Fig. 4.

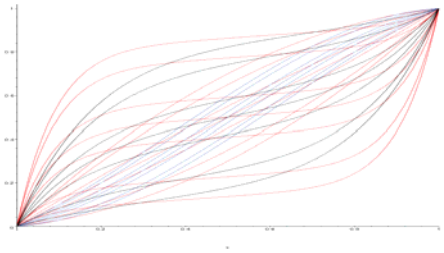


Figure 4. $P_{MS,0.5}(t;a,b)$ for different $a,b\geq 0$

The parametric functions $p_{R2}(.;\beta_0,\beta_1)$ and $p_{MS}(.;\beta_0,\beta_1)$ (or others with similar properties) can be used to construct a fuzzy number u in the LR form, by representing its membership function as

$$\mu_u(x)=\begin{cases} p(\frac{x-a}{b-a};\beta_{0,L},\beta_{1,L}) & \text{if } x\in[a,b]\\ 1 & \text{if } x\in[b,c]\\ p(\frac{d-x}{d-c};\beta_{0,R},\beta_{1,R}) & \text{if } x\in[c,d]\\ 0 & \text{otherwise}\end{cases} \tag{9}$$

or in the LU form, by representing its alpha-cuts as the compact intervals $[u_\alpha^-,u_\alpha^+]$, for $\alpha\in[0,1]$, with

$$\begin{aligned} u_\alpha^- &= a+(b-a)p(\alpha;\beta_0^-,\beta_1^-),\\ u_\alpha^+ &= d+(c-d)p(\alpha;\beta_0^+,\beta_1^+).\end{aligned} \tag{10}$$

Here, obviously, $a\leq b\leq c\leq d$ and $[a,d]$ is the support, while $[b,c]$ is the core of given fuzzy number u. If needed, we will denote a,b,c and d also by u_0^-,u_1^-,u_1^+ and u_0^+, respectively.

As it is possible to go from the LR to the LU representations by inverting the model functions, further on we will use the LU form only; however analogous formulations are possible for the LR form as well. The obvious relations between the derivatives of $\mu_u(x)$ and the derivatives of u_α^-, u_α^+ gives (provided they are not null)

$$\beta_{0,L}=\frac{1}{\beta_0^-},\beta_{1,L}=\frac{1}{\beta_1^-},\beta_{0,R}=\frac{1}{\beta_0^+},\beta_{1,R}=\frac{1}{\beta_1^+}. \tag{11}$$

Let us consider the expected interval of such fuzzy number u. It is given by $[E^-u,E^+u]$, where

$$\begin{aligned} E^-u &= u_0^-+(u_1^--u_0^-)\int_0^1 p(\alpha;\beta_0^-,\beta_1^-)d\alpha,\\ E^+u &= u_0^++(u_1^+-u_0^+)\int_0^1 p(\alpha;\beta_0^+,\beta_1^+)d\alpha\end{aligned} \tag{12}$$

and hence we have four parameters $\beta_0^-,\beta_1^-,\beta_0^+,\beta_1^+\geq 0$ free for our further purposes (or six parameters if $P_{R2,\lambda^-}(t;a^-,b^-)$ and $P_{R2,\lambda^+}(t;a^+,b^+)$ or $P_{MS,\lambda^-}(t;a^-,b^-)$ and $P_{MS,\lambda^+}(t;a^+,b^+)$ are used).

For example, if $\beta_0^-=\beta_1^-=\beta^-$ and $\beta_0^+=\beta_1^+=\beta^+$, we get $\int_0^1 p_{R2}(t;\beta^\pm,\beta^\pm)dt=\frac{1}{2}$ and the expected interval is

$$[E^-u,E^+u]=[\frac{u_0^-+u_1^-}{2},\frac{u_0^++u_1^+}{2}],$$

ISBN: 978-989-95079-6-8

for all nonnegative β^-, β^+. If a fuzzy number A with alpha cuts $[A_\alpha^-, A_\alpha^+]$ is to be approximated, we can constrain $u_0^- \leq u_1^- \leq u_1^+ \leq u_1^-$ such that

$$\begin{aligned} u_0^- + u_1^- &= 2\int_0^1 A_\alpha^- d\alpha \\ u_0^+ + u_1^+ &= 2\int_0^1 A_\alpha^+ d\alpha, \end{aligned}$$

while the two parameters $\beta^-, \beta^+ \geq 0$ still can be used for additional requirements. Thus, using our approach we obviously loose the benefits of linear shapes but the form of $p(t;\beta,\beta)$ is simple and it is easy to invert analytically by solving a quadratic equation instead of linear. It also reduces to linear if $\beta = 1$.

3 Nonlinear shape approximations

In this section, we consider the approximation problem for a fuzzy number A having membership function $\mu_A(x)$ and compact α-cuts $A_\alpha = [A_\alpha^-, A_\alpha^+]$, $\alpha \in [0,1]$.

If we like to preserve or approximate the middle set (i.e. the alpha cut for $\alpha = 0.5$) we have

$$p_{R2}(\frac{1}{2};\beta_0,\beta_1) = \frac{1+\beta_0}{\beta_0+\beta_1+2}$$

and we obtain the conditions

$$\begin{aligned} u_0^- + (u_1^- - u_0^-)\frac{1+\beta_0^-}{\beta_0^-+\beta_1^-+2} &= A_{0.5}^- \\ u_0^+ + (u_1^+ - u_0^+)\frac{1+\beta_0^+}{\beta_0^++\beta_1^++2} &= A_{0.5}^+ \end{aligned}$$

giving (if the $u_{0/1}^\pm$ are known) the following equations/constraints for the parameter β:

$$\begin{aligned} (u_1^- - A_{0.5}^-)\beta_0^- + (u_0^- - A_{0.5}^-)\beta_1^- &= 2A_{0.5}^- - u_0^- - u_1^- \\ (u_1^+ - A_{0.5}^+)\beta_0^+ + (u_0^+ - A_{0.5}^+)\beta_1^+ &= 2A_{0.5}^+ - u_0^+ - u_1^+. \end{aligned}$$

In the special case of $\beta_0 + \beta_1 = 2$ we obtain a parabolic shape function

$$p_{R2}(t;\beta_0,\beta_1) = Q(t;\beta) = t^2 + \beta t(1-t)$$

where $\beta_0 = \beta, \beta_1 = 2-\beta$ with $\beta \in [0,2]$ and

$$\int_0^1 Q(\alpha;\beta)d\alpha = \frac{1}{3} + \frac{1}{6}\beta$$

is in the range $[\frac{1}{3}, \frac{2}{3}]$. Then the equations for expected interval preservation are as follows

$$\begin{aligned} u_0^- + (u_1^- - u_0^-)(\frac{1}{3} + \frac{1}{6}\beta^-) &= \int_0^1 A_\alpha^- d\alpha, \\ u_0^+ + (u_1^+ - u_0^+)(\frac{1}{3} + \frac{1}{6}\beta^+) &= \int_0^1 A_\alpha^+ d\alpha. \end{aligned}$$

We can also consider the value and the ambiguity of fuzzy numbers or other quantities related to ranking or defuzzification. If $s : [0,1] \longrightarrow [0,1]$ is a regular reducing function (i.e. $\int_0^1 s(\alpha)d\alpha = \frac{1}{2}$) the value of the fuzzy number u relative to s is defined by (see [5])

$$Val_s(u) = \int_0^1 s(\alpha)\left(u_\alpha^- + u_\alpha^+\right)d\alpha, \tag{13}$$

while the ambiguity of u relative to s is

$$Amb_s(u) = \int_0^1 s(\alpha)\left(u_\alpha^+ - u_\alpha^-\right)d\alpha. \tag{14}$$

For our parametric forms, we obtain

$$\begin{aligned} Val_s(u) &= \frac{u_0^- + u_0^+}{2} + \\ &+(u_1^- - u_0^-)\int_0^1 s(\alpha)p(\alpha;\beta_0^-,\beta_1^-)d\alpha \\ &+(u_1^+ - u_0^+)\int_0^1 s(\alpha)p(\alpha;\beta_0^+,\beta_1^+)d\alpha \end{aligned}$$

and

$$\begin{aligned} Amb_s(u) &= \frac{u_0^+ - u_0^-}{2} + \\ &+(u_0^- - u_1^-)\int_0^1 s(\alpha)p(\alpha;\beta_0^-,\beta_1^-)d\alpha \\ &+(u_1^+ - u_0^+)\int_0^1 s(\alpha)p(\alpha;\beta_0^+,\beta_1^+)d\alpha. \end{aligned}$$

The integrals above can be computed by numerical approximations (e.g. trapezoidal or Simpson formulas), but for specific cases (e.g. $s(\alpha) = 1$ or $s(\alpha) = \alpha$) we can proceed analytically.

With reference to equations (10), define the following integrals, depending on the slope parameters:

$$I(\beta_0,\beta_1) = \int_0^1 p(\alpha;\beta_0,\beta_1)d\alpha \tag{15}$$

$$J(\beta_0,\beta_1) = \int_0^1 \alpha p(\alpha;\beta_0,\beta_1)d\alpha. \tag{16}$$

For the family of parametric functions $p_{MS}(\alpha;\beta_0,\beta_1)$ we have

$$\begin{aligned} I_{MS}(\beta_0,\beta_1) &= \frac{2+(3+2\beta_0)(\beta_0+\beta_1)}{2(1+\beta_0+\beta_1)(2+\beta_0+\beta_1)} \\ J_{MS}(\beta_0,\beta_1) &= \frac{42+(84+40\beta_0)(\beta_0+\beta_1)}{20(1+\beta_0+\beta_1)(2+\beta_0+\beta_1)(3+\beta_0+\beta_1)} \end{aligned}$$

They are nonlinear functions of β_0, β_1 but become linear if, for example, we know or we assume a given value $m =$

ISBN: 978-989-95079-6-8

$\beta_0 + \beta_1 \geq 0$. Note that fixing the value of m (we suggest m to be integer) is equivalent to fixing the "degree" of p_{MS} as a function of α. If m is integer then p_{MS} in a polynomial function of degree one (if $\beta_0 = \beta_1 = 1$), two (if $\beta_0+\beta_1 = 2$), three (if $\beta_0 = \beta_1 = 0$) or $m+1$ (if $m = \beta_0 + \beta_1 \geq 3$). In these cases we obtain

$$\begin{aligned} I_{MS}(\beta_0, m-\beta_0) &= \frac{2+(3+2\beta_0)m}{2(1+m)(2+m)} \\ J_{MS}(\beta_0, m-\beta_0) &= \frac{42+(84+40\beta_0)m}{20(1+m)(2+m)(3+m)} \end{aligned}$$

with constraint $0 \leq \beta_0 \leq m$.

For the commonly used regular reducing functions $s(\alpha) = \frac{1}{2}$ and $s(\alpha) = \alpha$, the expressions $I_{MS}(\beta_0, \beta_1)$ and $J_{MS}(\beta_0, \beta_1)$ allow to compute $Val_s(u)$ and $Amb_s(u)$ in closed form as functions of the parameters β_0^-, β_1^- and β_0^+, β_1^+ or, for fixed m^- and m^+, of β_0^-, β_0^+.

The ordering of fuzzy numbers can be approached in many ways (see [8] or [9] for recent literature and results). In most cases, the fuzzy numbers u are transformed into real numbers by a real-valued function $D(u)$. By the use of the LU-fuzzy representation, the ranking functions can be computed, either numerically or, if possible, analytically, in terms of the parameters $u_i^{\pm}$ and $\beta_i^{\pm}$ that define u.

The first example is the so called *level set average* (see [6]), also known as the expected value of a fuzzy number, and defined by

$$u_{DP}^* = \frac{1}{2}\int_0^1 \left(u_\alpha^- + u_\alpha^+\right) d\alpha. \tag{17}$$

Using (15) we obtain

$$\begin{aligned} u_{DP}^* &= \frac{1}{2}[u_0^- + (u_1^- - u_0^-)I_{MS}(\beta_0^-, \beta_1^-) + \\ &\quad + u_0^+ + (u_1^+ - u_0^+)I_{MS}(\beta_0^+, \beta_1^+)]. \end{aligned}$$

It is worth noting that the above integral is closely related to the *nearest interval approximation* of a fuzzy number (see [10]), expressed equivalently either in terms of the lower-upper functions or in terms of the membership function:

$$\begin{aligned} C(u) &= \left[\int_0^1 u_\alpha^- d\alpha\,,\ \int_0^1 u_\alpha^+ d\alpha\right] \qquad (18) \\ &= \left[u_1^- - \int_{u_0^-}^{u_1^-} \mu(x)dx\,,\ u_1^+ + \int_{u_1^+}^{u_0^+} \mu(x)dx\right]. \end{aligned}$$

Using our parametric family p_{MS} and with reference to (10) we get

$$\begin{aligned} C(u) &= \frac{1}{2}[u_0^- + (u_1^- - u_0^-)I_{MS}(\beta_0^-, \beta_1^-), \\ &\quad u_0^+ + (u_1^+ - u_0^+)I_{MS}(\beta_0^+, \beta_1^+)]. \end{aligned}$$

Other examples are the *interval valued possibilistic mean* $M(u) = [M^-(u), M^+(u)]$ and the *level-weighted average* u_{GW}^* given by

$$M^-(u) = 2\int_0^1 \alpha u_\alpha^- d\alpha, \quad M^+(u) = 2\int_0^1 \alpha u_\alpha^+ d\alpha$$

and

$$u_{GW}^* = \int_0^1 \alpha\left(u_\alpha^- + u_\alpha^+\right) d\alpha = \frac{M^-(u) + M^+(u)}{2}. \tag{19}$$

The calculations for p_{MS} family lead to

$$\begin{aligned} M^-(u) &= u_0^- + 2(u_1^- - u_0^-)J_{MS}(\beta_0^-, \beta_1^-) \\ M^+(u) &= u_0^+ - 2(u_0^+ - u_1^+)J_{MS}(\beta_0^+, \beta_1^+). \end{aligned}$$

4 Some particular approximations

We show in this section how the illustrated setting is useful to solve two specific types of approximations of a given fuzzy number A for which we know the support $[a, d]$ and the core $[b, c]$ (section 4.1) or we fix the shapes, by fixing the nonnegative parameters β_0^-, β_1^- and β_0^+, β_1^+ (section 4.2).

4.1 Approximations with known support and core

In the first case we are looking for the "best curve" under the condition that the support and the core of the approximating fuzzy number u are exactly the same of A. In other words, $u_0^- = a$, $u_0^+ = d$, $u_1^- = b$ and $u_1^+ = c$.

Now we have to estimate the shape-parameters β_0^-, β_1^- and β_0^+, β_1^+ such that a distance measure $Dist(A, u)$ is minimized, subject to the nonnegativity constraints on β_0^-, β_1^- and β_0^+, β_1^+. Here u results to be a function of the $\underline{\beta} = (\beta_0^-, \beta_1^-, \beta_0^+, \beta_1^+)$. So our problem is

$$\begin{aligned} &Min \quad Dist(A, u(\underline{\beta})) \\ &s.t. \\ &\underline{\beta} \geq 0. \end{aligned}$$

The distance $Dist(A, u(\underline{\beta}))$ can be calculated if we have other information on A. For example, if the membership function of A is known at other points, we can approximate the distance by the least squares functional.

Suppose that $\mu_A(x_j) = \mu_j$ for given $x_j \in (a, b)$, $j = 1, 2, ..., j_L$ and for given $x_j \in (c, d)$, where $j = j_L + 1, j_L + 2, ..., j_L + j_R$ (i.e. j_L values correspond to the left arm of the fuzzy number while j_R values to the right arm). So we should minimize

$$\begin{aligned} &Dist(A, u(\underline{\beta})) = \qquad (20) \\ &= \sum_{j=1}^{j_L} \left\{x_j - \left[a + (b-a)p(\mu_j; \beta_0^-, \beta_1^-)\right]\right\}^2 \\ &+ \sum_{j=j_L+1}^{j_L+j_R} \left\{x_j - \left[d + (c-d)p(\mu_j; \beta_0^+, \beta_1^+)\right]\right\}^2. \\ &s.t. \\ &\beta_0^-, \beta_1^-, \beta_0^+, \beta_1^+ \geq 0. \end{aligned}$$

Therefore, we have obtained a nonlinear least squares problem with four variables and nonnegativity constraints which can be solved by any numerical procedure.

The minimization (20) can be split into two independent problems, one to determine β_0^-, β_1^- and the other for β_0^+, β_1^+.

A special case applies if we have available a single additional observation for the left side ($j_L = 1$), say $\mu_A(x^-) =$

ISBN: 978-989-95079-6-8

μ^- with $x^- \in (a,b)$, $\mu^- \in (0,1)$ and for the right ($j_R = 1$), say $\mu_A(x^+) = \mu^+$ with $x^+ \in (c,d)$, $\mu^+ \in (0,1)$.

An interpolating solution satisfies equations

$$\begin{aligned} a + (b-a)p(\mu^-; \beta_0^-, \beta_1^-) &= x^- \\ d + (c-d)p(\mu^+; \beta_0^+, \beta_1^+) &= x^+. \end{aligned}$$

If we use $p_{R2}(t; \beta_0, \beta_1)$ for $p(t; \beta_0, \beta_1)$, we obtain two equations and four nonnegative variables

$$(b-x^-)\beta_0^- + (a-x^-)\beta_1^- = \gamma^- \quad (21)$$
$$(c-x^+)\beta_0^+ + (d-x^+)\beta_1^+ = \gamma^+, \quad (22)$$

where

$$\begin{aligned} \gamma^- &= \frac{(a-x^-)(-2\mu^{-^2} + 2\mu^- - 1) - (b-a)\mu^{-^2}}{\mu^-(1-\mu^-)} \\ \gamma^+ &= \frac{(d-x^+)(-2\mu^{+^2} + 2\mu^+ - 1) - (c-d)\mu^{+^2}}{\mu^+(1-\mu^+)}. \end{aligned}$$

Equations (21)-(22) represent a line in the plane (β_0, β_1) having an infinite number of nonnegative solutions (note that $b - x^- > 0$, $a - x^- < 0$ and $c - x^+ < 0$, $d - x^+ > 0$). So we have many possible choices. We suggest tree interesting solutions:

1. The unique solution having least norm of (β_0, β_1) (i.e. with minimal $\beta_0^2 + \beta_1^2$), is obtained at the intersections of lines (21)-(22) with the axes. It has the following closed form:

$$\begin{aligned} &\text{if } \gamma^- = 0 \text{ then } \beta_0^- = 0, \beta_1^- = 0; \quad (23) \\ &\text{if } \gamma^- > 0 \text{ then } \beta_0^- = \frac{\gamma^-}{b-x^-}, \beta_1^- = 0; \\ &\text{if } \gamma^- < 0 \text{ then } \beta_0^- = 0, \beta_1^- = \frac{\gamma^-}{a-x^-}. \end{aligned}$$

and

$$\begin{aligned} &\text{if } \gamma^+ = 0 \text{ then } \beta_0^+ = 0, \beta_1^+ = 0; \quad (24) \\ &\text{if } \gamma^+ > 0 \text{ then } \beta_0^+ = 0, \beta_1^+ = \frac{\gamma^+}{d-x^+}; \\ &\text{if } \gamma^+ < 0 \text{ then } \beta_0^+ = \frac{\gamma^+}{c-x^+}, \beta_1^+ = 0. \end{aligned}$$

Solution (23)-(24) is illustrated in Fig. 5 with the following data: $a = 0$, $b = 2.5$, $c = 3$, $d = 5$ and the two membership functions for $\mu_A(0.5) = 0.5$, $\mu_A(3.5) = 0.9$ and $\mu_A(0.6) = 0.5$, $\mu_A(3.5) = 0.8$.

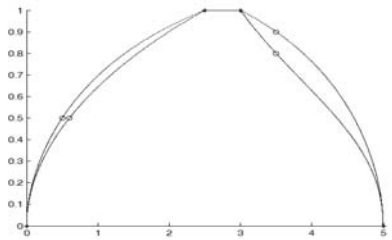


Figure 5: Solutions (23)-(24) for two examples data.

2. Recall that functions $p(t; \beta_0, \beta_1)$ represent a linear shape if and only if $\beta_0 = \beta_1 = 1$; so it is reasonable to "measure" its nonlinearity by the distance from (β_0, β_1) to $(1,1)$ and to minimize the distance $(\beta_0 - 1)^2 + (\beta_1 - 1)^2$. In this way, linear data will reproduce linear shapes. The unique solution for this criterion is obtained by the following procedure:

$$\begin{aligned} &\text{if } \widehat{\beta}_0^- \geq 0 \text{ and } \widehat{\beta}_1^- \geq 0 \text{ then } \beta_0^- = \widehat{\beta}_0^-,\ \beta_1^- = \widehat{\beta}_1^-; \quad (25) \\ &\text{if } \widehat{\beta}_0^- < 0 \text{ then } \beta_0^- = 0,\ \beta_1^- = \frac{\gamma^-}{a-x^-}; \\ &\text{if } \widehat{\beta}_1^- < 0 \text{ then } \beta_0^- = \frac{\gamma^-}{b-x^-}, \beta_1^- = 0. \end{aligned}$$

where

$$\begin{aligned} \widehat{\beta}_0^- &= \frac{\gamma^-(b-x^-) + (a-x^-)^2 - (b-x^-)(a-x^-)}{(b-x^-)^2 + (a-x^-)^2} \\ \widehat{\beta}_1^- &= 1 + \frac{(a-x^-)(\widehat{\beta}_0^- - 1)}{(b-x^-)}; \end{aligned}$$

for β_0^+, β_1^+ the procedure is analogous

$$\begin{aligned} &\text{if } \widehat{\beta}_0^+ \geq 0 \text{ and } \widehat{\beta}_1^+ \geq 0 \text{ then } \beta_0^+ = \widehat{\beta}_0^+,\ \beta_1^+ = \widehat{\beta}_1^+; \quad (26) \\ &\text{if } \widehat{\beta}_0^+ < 0 \text{ then } \beta_0^+ = 0,\ \beta_1^+ = \frac{\gamma^+}{d-x^+}; \\ &\text{if } \widehat{\beta}_1^+ < 0 \text{ then } \beta_0^+ = \frac{\gamma^+}{c-x^+}, \beta_1^+ = 0. \end{aligned}$$

where

$$\begin{aligned} \widehat{\beta}_0^+ &= \frac{\gamma^+(c-x^+) + (d-x^+)^2 - (d-x^+)(c-x^+)}{(c-x^+)^2 + (d-x^+)^2} \\ \widehat{\beta}_1^+ &= 1 + \frac{(d-x^+)(\widehat{\beta}_0^+ - 1)}{(c-x^+)}. \end{aligned}$$

Solution (25)-(26) is illustrated in Fig. 6 with the same data as before: $a = 0$, $b = 2.5$, $c = 3$, $d = 5$, $\mu_A(0.5) = 0.5$, $\mu_A(3.5) = 0.9$ and $\mu_A(0.6) = 0.5$, $\mu_A(3.5) = 0.8$.

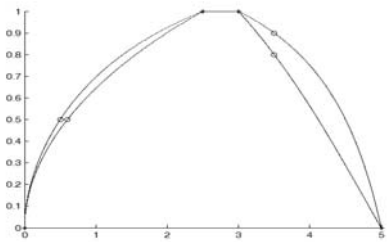


Figure 6: Solutions (25)-(26) for two examples data.

3. In cases 1. and 2. we made no assumptions on the parameters β_0^-, β_1^-, β_0^+, β_1^+, but we may know or desire to have them partially fixed. Lets consider the fitting equations in terms of the LR representation (9); if, for example, we require a differentiable membership function, this is equivalent to have $\beta_{1,L} = 0$ and $\beta_{1,R} = 0$ and we obtain the following solution

$$\beta_{0,L} = \begin{cases} \frac{\omega^-}{t^-(1-t^-)(1-\mu^-)} & \text{if } \omega^- > 0 \\ 0 & \text{if } \omega^- \leq 0, \end{cases} \quad (27)$$

where $t^- = \frac{x^- - a}{b-a}$ and $\omega^- = (1-2t^-)\mu^- - (1-2\mu^-)(t^-)^2$,

$$\beta_{0,R} = \begin{cases} \frac{\omega^+}{t^+(1-t^+)(1-\mu^+)} & \text{if } \omega^+ > 0 \\ 0 & \text{if } \omega^+ \leq 0, \end{cases} \quad (28)$$

ISBN: 978-989-95079-6-8

where $t^+ = \frac{d-x^+}{d-c}$ and $\omega^+ = (1-2t^+)\mu^+ - (1-2\mu^+)(t^+)^2$. Solution (27)-(28) is illustrated in Fig. 7 with the same data as for Fig. 5, 6. Clearly, in this case, we obtain membership functions such that $\mu(\frac{a+b}{2}) \geq \frac{1}{2}$ and $\mu(\frac{c+d}{2}) \geq \frac{1}{2}$ and interpolation is possible only for specific data.

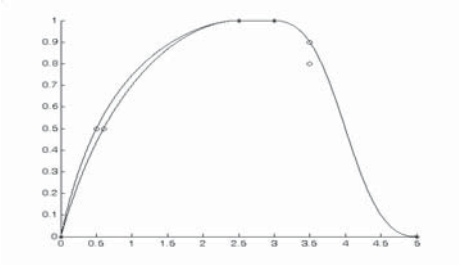


Figure 7: Solutions (27)-(28) for two examples data.

4.2 Approximations with fixed shapes

In the second case we assume that the general shape of u is fixed (by fixing the parameters $\underline{\beta}$) and we like to find the best approximation of the core and support, i.e. to find u_0^-, u_1^-, u_1^+, u_0^+ such that $d(A, u(\underline{\beta})) \rightarrow \min$.

In this case, the resulting optimization problem is easier than for Problem 1. Now, $u = u(\underline{\eta})$ is a function of $\underline{\eta} = (u_0^-, u_1^-, u_0^+, u_1^+)$ and we require $u_0^- \leq u_1^- \leq u_1^+ \leq u_0^+$.

Suppose also here that we know the membership of A at $j_L + j_R$ points, $\mu_A(x_j) = \mu_j$ for given $x_j \in [a, b]$, $j = 1, 2, ..., j_L$ and for given $x_j \in [c, d]$, $j = j_L + 1, j_L + 2, ..., j_L + j_R$.

Define $p_j = p(\mu_j; \beta_0^-, \beta_1^-)$ for $j = 1, 2, ..., j_L$ and $p_j = p(\mu_j; \beta_0^+, \beta_1^+)$ for $j = j_L + 1, j_L + 2, ..., j_L + j_R$.

Now we have to minimize

$$
\begin{aligned}
& Dist(A, u(\underline{\eta})) = \qquad (29)\\
= & \sum_{j=1}^{j_L} \left\{x_j - \left[u_0^- + (u_1^- - u_0^-)p_j\right]\right\}^2 \\
& + \sum_{j=j_L+1}^{j_L+j_R} \left\{x_j - \left[u_0^+ + (u_1^+ - u_0^+)p_j\right]\right\}^2 \\
& s.t. \\
& u_0^- - u_1^- \leq 0 \\
& u_1^- - u_1^+ \leq 0 \\
& u_1^+ - u_0^+ \leq 0.
\end{aligned}
$$

We obtain a linear least squares problem with four variables and three linear constraints and can be solved using standard well known procedures (see [14]).

5 Conclusions

In the present contribution we have suggested a general approach to fuzzy number approximation based on two families of parameterized functions. Our method enables a very rich set of approximating curves which are both easy to handle and useful for practical purposes.

References

[1] S. Abbasbandy, B. Asady, *The nearest approximation of a fuzzy quantity in parametric form*, Applied Mathematics and Computation, 172, 2006, 624-632.

[2] S. Abbasbandy, M. Amirfakhrian, *The nearest trapezoidal form of a generalized LR fuzzy number*, International Journal of Approximate Reasoning, 43, 2006, 166-178.

[3] Ban A., *Approximation of fuzzy numbers by trapezoidal fuzzy numbers preserving the expected interval*, Fuzzy Sets and Systems 159, 2008, 1327-1344.

[4] Y. Chalco-Cano, H. Roman-Flores, F. Gomide, *A new type of approximation for fuzzy intervals*, Fuzzy Sets and Systems, 159, 2008, 1376-1383.

[5] M. Delgado, M.A. Vila, W. Voxman, *A fuzziness measure for fuzzy number: applications*, Fuzzy Sets and Systems 94 (1998), 205–216.

[6] D. Dubois, H. Prade, *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, New York, 1980.

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

[8] S. Heilpern, *Representation and application of fuzzy numbers*, *Fuzzy Sets and Systems*, 91, 1997, 259-278.

[9] P. Grzegorzewski, *Metrics and orders in space of fuzzy numbers*, Fuzzy Sets and Systems, 97, 1998, 83-94.

[10] P. Grzegorzewski, *Nearest interval approximation of a fuzzy number*, Fuzzy Sets and Systems, 130, 2002, 321-330.

[11] P. Grzegorzewski, *Trapezoidal approximations of fuzzy numbers preserving the expected value: algorithms and properties*, Fuzzy Sets and Systems, 159, 2008, 1354-1364.

[12] P. Grzegorzewski, E. Mrówka, *Trapezoidal approximations of fuzzy numbers*, Fuzzy Sets and Systems, 153, 2005, 115-135.

[13] P. Grzegorzewski, E. Mrówka, *Trapezoidal approximations of fuzzy numbers - revisited*, Fuzzy Sets and Systems 158 (2007), 757–768.

[14] C. L. Lawson, C. R. J. Hanson, *Solving least squares problems*, Prentice-Hall, 1974.

[15] E. N. Nasibov, S. Peker, *On the nearest parametric approximation of a fuzzy number*, Fuzzy Sets and Systems, 159, 2008, 1365-1375.

[16] L. Stefanini, L. Sorini, M. L. Guerra, *Parametric representation of fuzzy numbers and application to fuzzy calculus*, Fuzzy Sets and Systems, 157, 2006, 2423 – 2455.

[17] L. Stefanini, L. Sorini, M.L. Guerra, *Fuzzy Numbers and Fuzzy Arithmetic*, in W. Pedrycz et al. (Eds) "*Handbook of Granular Computing*" , Chapter 12, J. Wiley & Sons, 2008.

[18] C.T. Yeh, *Trapezoidal and triangular approximations preserving the expected interval*, Fuzzy Sets and Systems 159, 2008, 1345-1353.

ISBN: 978-989-95079-6-8

Characterizing Fuzzy Modal Semantics by fuzzy multimodal systems with crisp accessibility relations

Félix Bou[1], Francesc Esteva[1], Lluís Godo[1], Ricardo Rodríguez[2]

1. Institut d'Investigació en Intel·ligència Artificial, IIIA - CSIC,
Campus UAB, 08193 Bellaterra, Spain
2. Dpto. de Computación, Universidad Buenos Aires,
1428 Buenos Aires, Argentina
Email: {fbou,esteva,godo}@iiia.csic.es, ricardo@dc.uba.ar

***Abstract*—** *In [1] the authors considered finitely-valued modal logics with Kripke style semantics where both propositions and the accessibility relation are valued over a finite residuated lattice. Unfortunately, the necessity operator does not satisfy in general the normality axiom (K). In this paper we focus on the case of finite chains, and we consider a different approach based on introducing a multimodal logic where the previous necessity operator is replaced with a family, parametrized by truth values different from zero, of necessity operators each one semantically defined using the crisp accessibility relation given by the corresponding cut of the finitely-valued original accessibility relation. This multimodal logic is somehow more appealing than the original modal one because axiom (K) holds for each necessity operator. In this paper we axiomatize this multimodal logic and we prove that, in the case the starting residuated lattice is a finite BL chain, the modal and the multimodal languages have the same expressive power iff this algebra is an MV chain.*

***Keywords*—** many-valued modal logic, fuzzy modal logic, Łukasiewicz modal logic, fuzzy logic.

1 Introduction

Fuzzy modal logics is a subfield of mathematical fuzzy logic with growing interest. The interested reader is referred to [1] and the references therein for recent developments. Indeed, in [1] the same authors proposed a mathematical definition of what a many-valued (uni)modal logic is. According to this definition *many-valued modal logics* are sets[1] of modal formulas, denoted by $\Lambda(\mathsf{K}, \mathbf{A})$, arising naturally from the semantics given by a complete residuated lattice $\mathbf{A}$, whose support A corresponds to the intended set of truth values, and a class K of Kripke frames valuated over the set A.

In [1], due to simplicity reasons, the authors only considered *modal formulas* obtained enriching the propositional language of residuated lattices with a necessity operator $\Box$ (i.e., without a primitive possibility operator $\Diamond$). These modal formulas may include or not canonical constants to talk about the truth values. By adding *canonical constants* we mean to add one constant $\overline{a}$ for every element a in the residuated lattice $\mathbf{A}$ in such a way that each one of these constants is semantically interpreted by its canonical interpretation (i.e., the very element a). In the case that we allow canonical constants in the modal language we use the notation $\Lambda(\mathsf{K}, \mathbf{A^c})$ to stress their presence. In the present paper we will always assume that the language has canonical constants.

An assumption that we consider throughout the present paper is that $\mathbf{A}$ is a chain (to keep us inside the fuzzy realm) and finite, i.e., $\mathbf{A}$ is a finite MTL chain. We remind the reader that finiteness is crucial in the results obtained in [1], but linearity is not required there. In the next paragraphs we point out some of the results in [1] for the particular case of finite chains.

One of the main results in [1] is the presentation of a complete calculus for the many-valued modal logic $\Lambda(\mathsf{K}, \mathbf{A^c})$ when K is the class Fr of all frames valued over $\mathbf{A}$, assuming a complete calculus is already known for the non-modal logic without canonical constants (denoted by $\mathbf{\Lambda(A)}$). This axiomatization is shown in Table 1. It is worth pointing out that in the case that $\mathbf{A}$ is a finite BL chain, it is known [1, Propositions 2.5 and 2.7] that there are axiomatizations of $\mathbf{\Lambda(A)}$ based on only one rule, the Modus Ponens rule; but there are examples of finite MTL chains where it is strictly necessary to add more rules besides Modus Ponens.

Among the difficulties to find such an axiomatization is that while the meet distributivity axiom

$$(\Box\varphi \wedge \Box\psi) \leftrightarrow \Box(\varphi \wedge \psi) \qquad (\mathsf{MD})$$

is valid (as in the classical modal case), in general this is not the case in the many-valued modal setting for the normality axiom

$$\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi) \qquad (\mathsf{K})$$

The normality axiom is indeed valid in $\Lambda(\mathsf{Fr}, \mathbf{A^c})$ iff the residuated lattice is a Heyting algebra (i.e., the interpretation of the strong conjunction coincides with the meet).

The situation is much more difficult in case that there are no canonical constants in the language, and as far as the authors are aware the only known axiomatizations for many-valued modal logics of the form $\Lambda(\mathsf{Fr}, \mathbf{A})$, where $\mathbf{A}$ is not a Heyting algebra, are the ones given in [1] for the case that $\mathbf{A}$ is a finite MV chain. The main drawback of the axiomatization there given is that it is rather artificial, and hence it is not absolutely clear what are the basic principles of these many-valued modal logics.

On the other hand, if one considers K as the class CFr of *crisp Kripke frames* (i.e., those Kripke frames valuating the accessibility relation over the set[2] $\{0, 1\}$) then the problem

[1] Besides these sets in the same paper it is also considered the consequence relations $\mathbf{\Lambda}(l, \mathsf{K}, \mathbf{A})$ and $\mathbf{\Lambda}(g, \mathsf{K}, \mathbf{A})$ corresponding to the local and global many-valued modal consequence relations.

[2] It is worth pointing out that since $\mathbf{A}$ is a chain it holds that the set $\{0, 1\}$ coincides with the set $\{a \in A : 1 = a \vee \neg a\}$ of Boolean elements of $\mathbf{A}$.

ISBN: 978-989-95079-6-8

Table 1: Axiomatization of the set $\Lambda(\mathsf{Fr}, \mathbf{A^c})$ when $\mathbf{A}$ is a finite chain

- the set of axioms is the smallest set closed under substitutions containing
  - the axiomatic basis for $\mathbf{\Lambda(A)}$,
  - the witnessing axiom $\bigvee_{a\in A}(\varphi \leftrightarrow \overline{a})$
  - the bookkeeping axioms $(\overline{a_1} * \overline{a_2}) \leftrightarrow \overline{a_1 * a_2}$ (for every $a_1, a_2 \in A$ and every $* \in \{\wedge, \vee, \odot, \rightarrow\}$),
  - $\Box 1$, $(\Box\varphi \wedge \Box\psi) \rightarrow \Box(\varphi \wedge \psi)$ and $\Box(\overline{a} \rightarrow \varphi) \leftrightarrow (\overline{a} \rightarrow \Box\varphi)$ (for every $a \in A$),
- the rules of the basis for $\mathbf{\Lambda(A)}$, the rule $\overline{k} \vee \varphi \vdash \varphi$ (where k is the coatom of $\mathbf{A}$) and the Monotonicity rule $\varphi \rightarrow \psi \vdash \Box\varphi \rightarrow \Box\psi$.

resembles much more the one in the classical modal setting since now the normality axiom is valid. It is known [1] that $\Lambda(\mathsf{CFr}, \mathbf{A^c})$ is the set of modal formulas derivable in the calculus given in Table 2. This table is, roughly speaking, saying that we only need to add to $\Lambda(\mathsf{Fr}, \mathbf{A^c})$ the normality axiom plus the axiom $\Box(\overline{k}\vee\varphi) \rightarrow (\overline{k}\vee\Box\varphi)$ (where k is the coatom of $\mathbf{A}$) in order to capture the logic of crisp frames. We point out it is known that in general it is not enough to add the normality axiom[3] (cf. Remark 3.2); and we notice that the last axiom is a particular case of the formulas $\Box(\overline{a} \vee \varphi) \rightarrow (\overline{a} \vee \Box\varphi)$ with $a \in A$, all these formulas being valid in $\Lambda(\mathsf{CFr}, \mathbf{A^c})$. Notice that in Table 2 we have replaced the Monotonicity rule by the Necessity rule, but this is not important due to the presence of the normality axiom.

The aim of this paper is to apply to the realm of many-valued modal logics the well known [2] correspondence between

- a fuzzy binary relation over A (i.e., a function $R : W \times W \longrightarrow A$), and
- an $A \setminus \{0\}$-indexed and decreasing family of crisp binary relations (i.e., a family $\{R_a : a \in A, a \neq 0\}$ such that if $a \leqslant b$ then $R_b \subseteq R_a \subseteq W \times W$).

This correspondence is given by the following identities

$$R_a = \{(w_1, w_2) \in W \times W : R(w_1, w_2) \geqslant a\} \qquad (1)$$

$$R(w_1, w_2) = \sup\{a \in A : (w_1, w_2) \in R_a\} \qquad (2)$$

The fact that $\mathbf{A}$ is a finite chain (in particular a complete lattice) is crucial in order to see that the previous identities induce a bijective correspondence.

In other words, the previous correspondence transforms a Kripke frame into a family of crisp Kripke frames. This transformation suggests to use a multimodal language (with one modality $\Box_a$ for every element $a \in A \setminus \{0\}$) to describe properties of Kripke frames. The advantage of this method is that, since the modalities $\Box_a$'s are induced by crisp Kripke frames, the normality axioms

$$\Box_a(\varphi \rightarrow \psi) \rightarrow (\Box_a\varphi \rightarrow \Box_a\psi)$$

are valid.

[3] In [1] it is proved that if we only add the normality axiom to $\Lambda(\mathsf{Fr}, \mathbf{A^c})$ then we get an axiomatization for the modal logic given by the class of idempotent frames (i.e., those Kripke frames valuating the accessibility relation over the set $\{a \in A : a = a \odot a\}$).

The aim of the present paper is to pursue this research line. To this purpose in Section 2 these multimodal logics are introduced, in Section 3 a complete calculus is given for them, and in Section 4 we compare the expressive power of the crisp multimodal approach with the unimodal one from [1]. Finally, in Section 5 the authors analyze what it is known about these results when there are no canonical constants in the language.

2 Defining the multimodal systems

Throughout this paper we assume that $\mathbf{A}$ is a fixed *finite residuated lattice* whose underlying order is a *chain*. We remind the reader that a residuated lattice is an algebra $\mathbf{A}$ such that:

- $\langle A, \wedge, \vee, 0, 1\rangle$ is a bounded lattice with a linear associated order $\leqslant$,
- $\langle A, \odot, 1\rangle$ is a commutative monoid with the unit 1,
- $x \odot z \leqslant y \Leftrightarrow z \leqslant x \rightarrow y$ (the law of *residuation*).

In the literature these algebras are well known under different names: residuated lattices, integral, commutative residuated monoids, $\mathsf{FL_{ew}}$-algebras, etc. [3, 4, 5]. We stress that finite MTL-algebras [6], finite BL-algebras [7] and finite MV-algebras [8] are particular cases satisfying our assumption.

We stress that the algebraic language of $\mathbf{A}$ is given by $\langle\wedge, \vee, \odot, \rightarrow, 1, 0\rangle$ (with arities $\langle 2, 2, 2, 2, 0, 0\rangle$). The *multimodal language* is the one obtained by enriching the previous one with canonical constants and a unary operator $\Box_a$ for every $a \in A \setminus \{0\}$. We remind the reader that in this paper (see p. 1) with the term *modal language* we denote the language obtained enriching $\langle\wedge, \vee, \odot, \rightarrow, 1, 0\rangle$ with canonical constants and one unary operator $\Box$. In case we are interested in the expansion having simultaneously, besides canonical constants, the operators $\Box_a$'s and $\Box$ we will use the expression *full modal language*. Of course we adopt the analogous convention to talk about *multimodal formulas*, *modal formulas* and *full modal formulas*.

A *Kripke frame* is a pair $\mathfrak{F} = \langle W, R\rangle$ where W is a non empty set (whose elements are called *worlds*) and R is a binary relation valued in A (i.e., $R : W \times W \longrightarrow A$) called *accessibility relation*. $\mathfrak{F}$ is said to be *crisp* in case that the range of R is included in $\{0, 1\}$. The classes of Kripke frames and crisp Kripke frames will be denoted, respectively, by Fr and CFr.

A *Kripke model* is a 3-tuple $\mathfrak{M} = \langle W, R, V\rangle$ where $\langle W, R\rangle$ is a Kripke frame and V is a map, called *valuation*, assigning

ISBN: 978-989-95079-6-8

Table 2: Axiomatization of the set $\Lambda(\mathsf{CFr}, \mathbf{A^c})$ when $\mathbf{A}$ is a finite chain

- the set of axioms is the smallest set closed under substitutions containing
  - the axiomatic basis for $\mathbf{\Lambda(A)}$,
  - the witnessing axiom $\bigvee_{a\in A}(\varphi \leftrightarrow \overline{a})$
  - the bookkeeping axioms $(\overline{a_1} * \overline{a_2}) \leftrightarrow \overline{a_1 * a_2}$ (for every $a_1, a_2 \in A$ and every $* \in \{\wedge, \vee, \odot, \rightarrow\}$),
  - $\Box 1$, $(\Box\varphi \wedge \Box\psi) \rightarrow \Box(\varphi \wedge \psi)$ and $\Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi)$
  - $\Box(\overline{a} \rightarrow \varphi) \leftrightarrow (\overline{a} \rightarrow \Box\varphi)$ (for every $a \in A$) and $\Box(\overline{k} \vee \varphi) \rightarrow (\overline{k} \vee \Box\varphi)$ (where k is the coatom of $\mathbf{A}$),
- the rules of the basis for $\mathbf{\Lambda(A)}$, the rule $\overline{k} \vee \varphi \vdash \varphi$ (where k is the coatom of $\mathbf{A}$) and the Necessity rule $\varphi \vdash \Box\varphi$.

to each propositional variable and each world in W an element of A (i.e., $V : \mathsf{Var} \times W \longrightarrow A$ where Var is the set of propositional variables). In such a case we say that $\mathfrak{M}$ is based on the Kripke frame $\langle W, R\rangle$.

If $\mathfrak{M} = \langle W, R, V\rangle$ is a Kripke model, then the map V can be uniquely extended to a map V' assigning to each full modal formula (in particular also multimodal and modal formulas) and each world in W an element of A (i.e., $V' : Fm \times W \longrightarrow A$) satisfying that:[4]

- V' is an algebraic homomorphism, in its first component, for the connectives $\wedge, \vee, \odot, \rightarrow, 1$ and 0,
- $V'(\overline{a}, w) = a$ for every $a \in A$.
- $V'(\Box\varphi, w) = \bigwedge\{R(w, w') \rightarrow V'(\varphi, w') : w' \in W\}$.
- $V'(\Box_a\varphi, w) = \bigwedge\{R_a(w, w') \rightarrow V'(\varphi, w') : w' \in W\} = \bigwedge\{V'(\varphi, w') : w' \in W, R(w, w') \geqslant a\}$.

Although V and V' are different mappings there is no problem, since one is an extension of the other, to use the same notation V for both. As usual we say that two formulas φ and ψ are *equivalent* (in symbols $\varphi \equiv \psi$) iff for every Kripke model $\mathfrak{M}$ it holds that $V(\varphi) = V(\psi)$.

Definition 2.1. The *local (many-valued) full modal logic* $\mathbf{F\Lambda}(l, \mathsf{Fr}, \mathbf{A^c})$ is the consequence relation obtained by defining, for all sets $\Gamma \cup \{\varphi\}$ of full modal formulas,

- $\Gamma \vdash_{\mathbf{F\Lambda}(l,\mathsf{Fr},\mathbf{A^c})} \varphi$, iff
- For every Kripke model $\langle W, R, V\rangle$ and $w \in W$, if $V(\gamma, w) = 1$ for every $\gamma \in \Gamma$, then $V(\varphi, w) = 1$.

And the *global (many-valued) full modal logic* $\mathbf{F\Lambda}(g, \mathsf{Fr}, \mathbf{A^c})$ is the one given by defining

- $\Gamma \vdash_{\mathbf{F\Lambda}(g,\mathsf{Fr},\mathbf{A^c})} \varphi$, iff
- For every Kripke model $\langle W, R, V\rangle$, if $V(\gamma, w) = 1$ for every $\gamma \in \Gamma$ and every $w \in W$, then $V(\varphi, w) = 1$ for every $w \in W$.

In the case that we restrict our attention to multimodal formulas or modal ones we will analogously use the notations $\mathbf{M\Lambda}(l, \mathsf{Fr}, \mathbf{A^c})$, $\mathbf{M\Lambda}(g, \mathsf{Fr}, \mathbf{A^c})$, $\mathbf{\Lambda}(l, \mathsf{Fr}, \mathbf{A^c})$ and $\mathbf{\Lambda}(g, \mathsf{Fr}, \mathbf{A^c})$. A formula is *valid* iff it is a theorem of the local (or the global) consequence relation. We will write $\mathsf{M\Lambda}(\mathsf{Fr}, \mathbf{A^c})$, $\mathsf{F\Lambda}(\mathsf{Fr}, \mathbf{A^c})$ and $\Lambda(\mathsf{Fr}, \mathbf{A^c})$ to denote the set of valid formulas in the corresponding language.

From this definition it is obvious that all these local and global consequence relations are conservative expansions of the non-modal consequence relations with and without canonical constants, denoted respectively by $\mathbf{\Lambda(A^c)}$ and $\mathbf{\Lambda(A)}$.

3 Completeness of the multimodal systems

In this section we give an sketch of the proof of the following completeness theorem.

Theorem 3.1 (Completeness)**.** *Let* $\mathbf{A}$ *be a finite* MTL *chain.*

1. *The consequence relation* $\mathbf{M\Lambda}(g, \mathsf{Fr}, \mathbf{A^c})$ *is the consequence relation axiomatized by the axioms and rules given in Table 3.*
2. *The set* $\mathsf{M\Lambda}(\mathsf{Fr}, \mathbf{A^c})$ *is the set of modal formulas derivable in the calculus given in Table 3.*
3. *The consequence relation* $\mathbf{M\Lambda}(l, \mathsf{Fr}, \mathbf{A^c})$ *is axiomatized by (i)* $\mathsf{M\Lambda}(\mathsf{Fr}, \mathbf{A^c})$ *as the set of axioms, and (ii) the rules of the basis for* $\mathbf{\Lambda(A)}$ *together with the rule* $\overline{k} \vee \varphi \vdash \varphi$ *(where* k *is the coatom of* $\mathbf{A}$*).*

We notice that the behaviour of the $\Box_b$ operators given in this axiomatization (Table 3) is the same one that was given for $\Box$ in the case of crisp Kripke frames (cf. Table 2), plus the addition of some axioms telling that the $\Box_b$'s are somehow nested.

Remark 3.2. Before explaining the proof it is worth pointing out that if we delete the axioms $\Box_b(\overline{k} \vee \varphi) \rightarrow (\overline{k} \vee \Box_b\varphi)$ from Table 3 then we get an incomplete system. This fact can be checked using an standard matrix argument. Let us consider the Gödel algebra $\mathbf{G_3}$ with three elements $\{0, 0.5, 1\}$, and expand it with canonical constants (i.e., $\overline{0.5} = 0.5$) and with the connectives $\Box_{0.5}$ and $\Box_1$ being interpreted by the unary function $f : \mathbf{G_3} \longrightarrow \mathbf{G_3}$ defined by $f(x) := 0.5 \rightarrow x$. Then, this algebra enriched with the set $\{1\}$ of designated elements is a matrix that is a model of all axioms and rules given in Table 3 except for the ones of the form $\Box_b(\overline{k} \vee \varphi) \rightarrow (\overline{k} \vee \Box_b\varphi)$. It is not a model of these last axioms because for example $\Box_1(\overline{0.5} \vee 0) \rightarrow (\overline{0.5} \vee \Box_1 0) = 0.5$.

[4] The infimun of the empty set is taken, as usual, equal to 1.

ISBN: 978-989-95079-6-8

Table 3: Axiomatization of the set $\mathsf{M}\Lambda(\mathsf{Fr}, \mathbf{A^c})$ when $\mathbf{A}$ is a finite chain

- the set of axioms is the smallest set closed under substitutions containing
  - the axiomatic basis for $\mathbf{\Lambda(A)}$,
  - the witnessing axiom $\bigvee_{a\in A}(\varphi \leftrightarrow \overline{a})$
  - the bookkeeping axioms $(\overline{a_1} * \overline{a_2}) \leftrightarrow \overline{a_1 * a_2}$ (for every $a_1, a_2 \in A$ and every $* \in \{\wedge, \vee, \odot, \rightarrow\}$),
  - $\Box_b 1$, $(\Box_b\varphi \wedge \Box_b\psi) \rightarrow \Box_b(\varphi \wedge \psi)$ and $\Box_b(\overline{a} \rightarrow \varphi) \leftrightarrow (\overline{a} \rightarrow \Box_b\varphi)$ (for every $a \in A$ and $b \in A \setminus \{0\}$),
  - $\Box_b(\varphi \rightarrow \psi) \rightarrow (\Box_b\varphi \rightarrow \Box_b\psi)$ and $\Box_b(\overline{k} \vee \varphi) \rightarrow (\overline{k} \vee \Box_b\varphi)$ (for every $b \in A \setminus \{0\}$) (where k is the coatom of $\mathbf{A}$),
  - $\Box_{b_1}\varphi \rightarrow \Box_{b_2}\varphi$ (for every $b_1, b_2 \in A \setminus \{0\}$ such that $b_1 \leqslant b_2$),
- the rules of the basis for $\mathbf{\Lambda(A)}$, the rule $\overline{k} \vee \varphi \vdash \varphi$ (where k is the coatom of $\mathbf{A}$) and the Necessity rules $\varphi \vdash \Box_b\varphi$ (for every $b \in A \setminus \{0\}$).

Next we give some hints on the proof of Theorem 3.1. To this purpose for the rest of this section we will use the symbol L to denote the set of modal formulas derivable from the calculus given in Table 3, and the symbol $\vdash_{\mathsf{L}}$ to denote the consequence relation axiomatized by L as set of axioms and the rules of the basis for $\mathbf{\Lambda(A)}$ together with the rule $\overline{k} \vee \varphi \vdash \varphi$ (where k is the coatom of $\mathbf{A}$). The completeness theorem can be proved using two steps: (i) a reduction of the multimodal completeness problem to the already known strong completeness of the non-modal logic $\mathbf{\Lambda(A^c)}$, and (ii) a Truth Lemma based on a canonical Kripke model construction. Next we briefly sketch the proofs of each one of these steps in the completeness proof.

The first step is based on the fact that $\vdash_{\mathsf{L}}$ is strongly complete by definition with respect to $\mathbf{\Lambda(A)}$, and hence by [1, Corollary 2.16] we know that it is also strongly complete with respect to $\mathbf{\Lambda(A^c)}$. We point out that the rule $\overline{k} \vee \varphi \vdash \varphi$ plays a remarkable role in the proof of [1, Corollary 2.16]. Therefore, we get the following trivial consequence.

Lemma 3.3 (Non-Modal Reduction)**.** *Let $\Gamma \cup \{\varphi\}$ be a set of multimodal formulas. Then*

- *$\Gamma \vdash_{\mathsf{L}} \varphi$, iff*
- *$\Gamma \vdash_{\mathbf{\Lambda(A^c)}} \varphi$, i.e., for every homomorphism h from the algebra of multimodal formulas[5] into the algebra $\mathbf{A^c}$, if $h[\Gamma] \subseteq \{1\}$ then $h(\varphi) = 1$.*

On the other hand, the second step consists on a canonical Kripke model construction. The definition of this construction is the following one.

Definition 3.4. The *multi canonical Kripke model* $\mathfrak{M}_{\mathtt{mcan}}$ is the Kripke model $\langle W_{mcan}, R_{mcan}, V_{mcan}\rangle$ where

- the set W_{mcan} is the set of non-modal homomorphisms $v : \mathbf{Fm} \longrightarrow \mathbf{A^c}$ (we point out that the algebra $\mathbf{Fm}$ is the one given by multimodal formulas) such that $v[\mathsf{L}] = \{1\}$.
- the accessibility relation R_{mcan} is defined by[6]

$$R_{mcan}(v_1, v_2) := \bigvee\{b \in A\setminus\{0\} : \forall\varphi(v_1(\Box_b\varphi) \leqslant v_2(\varphi))\}.$$

- the valuation map is defined by $V_{mcan}(p, v) := v(p)$ for every variable p.

It is obvious that for every $b \in A \setminus \{0\}$, it holds that $b \leqslant R_{mcan}(v_1, v_2)$ iff $v_1(\Box_b\varphi) \leqslant v_2(\varphi)$ for every multimodal formula φ.

Lemma 3.5 (Truth Lemma)**.** *The multi canonical Kripke model $\mathfrak{M}_{\mathtt{mcan}}$ satisfies that $V_{mcan}(\varphi, v) = v(\varphi)$ for every multimodal formula φ and world v.*

Proof. The proof is done by induction on the multimodal formula. The only non trivial case is when this formula starts with a necessity operator $\Box_b$. By the inductive hypothesis it is clear that it is enough to prove that

$$\bigwedge\{v'(\varphi) : v' \in W_{mcan}, b \leqslant R_{mcan}(v, v')\} = v(\Box_b\varphi)$$

where v is a world. And indeed the only non trivial inequality is $\leqslant$. Hence, let us consider $a := \bigwedge\{v'(\varphi) : v' \in W_{mcan}, b \leqslant R_{mcan}(v, v')\}$ and try to prove that $a \leqslant v(\Box_b\varphi)$.

First of all we claim that

$$\mathsf{L} \cup \{\overline{d} \rightarrow \psi : \psi \in Fm, d = v(\Box_b\psi) \in A\} \vdash_{\mathbf{\Lambda(A^c)}} \overline{a} \rightarrow \varphi.$$

Why? Let us consider a homomorphism h from the algebra of multimodal formulas into the algebra $\mathbf{A^c}$ such that $h[\mathsf{L}\cup\{\overline{d} \rightarrow \psi : \psi \in Fm, d = v(\Box_b\psi) \in A\}] = \{1\}$. We have to prove that $h(\overline{a} \rightarrow \varphi) = 1$, i.e., $a \leqslant h(\varphi)$. The assumptions on h imply that $h \in W_{mcan}$, and that $v(\Box_b\psi) \leqslant h(\psi)$ for every ψ. Hence, $b \leqslant R_{mcan}(v, h)$. Using the definition of a we get that $a \leqslant h(\varphi)$. This finishes the proof of this first claim.

Hence, using that $\mathbf{\Lambda(A^c)}$ is a finitary logic (because A is finite) we get from the previous claim that there is some $m \in \omega$, some multimodal formulas $\psi_1, \ldots, \psi_m$ and some elements $d_1, \ldots, d_m \in A$ such that $d_i = v(\Box_b\psi_i)$ for every $i \in \{1, \ldots, m\}$ and

$$\mathsf{L} \cup \{\overline{d_1} \rightarrow \psi_1, \ldots, \overline{d_m} \rightarrow \psi_m\} \vdash_{\mathbf{\Lambda(A^c)}} \overline{a} \rightarrow \varphi.$$

From here it is clear[7] that $\mathsf{L} \vdash_{\mathbf{\Lambda(A^c)}} ((\overline{d_1} \rightarrow \psi_1)\wedge\ldots\wedge(\overline{d_m} \rightarrow \psi_m)) \rightarrow (\overline{k} \vee (\overline{a} \rightarrow \varphi))$. Using Lemma 3.3 we get from the

[5] When we look at the multimodal formulas as non-modal ones we are thinking that $\{\Box_a\varphi : a \in A \setminus \{0\}, \varphi$ multimodal formula$\}$ are the variables of this non-modal language.

[6] The supremum of the empty set is taken, as usual, equal to 0.

[7] The trick used here is a particular case of the following more general statement:

$$\Gamma, \gamma \vdash_{\mathbf{\Lambda(A^c)}} \varphi \quad \text{iff} \quad \Gamma \vdash_{\mathbf{\Lambda(A^c)}} \gamma \rightarrow (\overline{k} \vee \varphi).$$

This statement is a consequence of the fact that k is the coatom of $\mathbf{A}$.

ISBN: 978-989-95079-6-8

previous claim that $\mathsf{L} \vdash_{\mathsf{L}} ((\overline{d_1} \to \psi_1) \wedge \ldots \wedge (\overline{d_m} \to \psi_m)) \to (\overline{k} \vee (\overline{a} \to \varphi))$. Thus, $((\overline{d_1} \to \psi_1) \wedge \ldots \wedge (\overline{d_m} \to \psi_m)) \to (\overline{k} \vee (\overline{a} \to \varphi)) \in \mathsf{L}$. Therefore, $\Box_b((\overline{d_1} \to \psi_1) \wedge \ldots \wedge (\overline{d_m} \to \psi_m)) \to \Box_b(\overline{k} \vee (\overline{a} \to \varphi)) \in \mathsf{L}$ by the Monotonicity Rule. Thus, $((\overline{d_1} \to \Box_b\psi_1) \wedge \ldots \wedge (\overline{d_m} \to \Box_b\psi_m)) \to (\overline{k} \vee (\overline{a} \to \Box_b\varphi)) \in \mathsf{L}$ using some of the axioms in Table 3. Finally, using that $v(\overline{d_1} \to \Box_b\psi_1) = \ldots = v(\overline{d_m} \to \Box_b\psi_m) = 1$ and that $k \neq 1$, we obtain that $v(\overline{a} \to \Box_b\varphi) = 1$, i.e., $a \leqslant v(\Box_b\varphi)$. This finishes the proof. □

Remark 3.6. In the calculus given in Table 3 we have included for each one of the $\Box_b$ modalities the meet distributivity axiom, the normality axiom, the Monotonicity rule and the Necessity rule. Checking the details of the proof of Lemma 3.5 the reader can realize that among the previous four axioms/rules it is enough to include the meet distributivity axiom and the Monotonicity rule in order to get completeness. And the same is also true if we only include the normality axiom and the Necessity rule, but this time the argument is slightly different (just replace $\wedge$ with $\odot$ in the proof of Lemma 3.5).

4 Comparing Expressive Power

The aim of this section is to compare the expressive power of the multimodal language with the one of the modal language. To this purpose we introduce the following concepts.

Definition 4.1. Given two pointed[8] Kripke models $\langle \mathfrak{M}, w\rangle$ and $\langle \mathfrak{M}', w'\rangle$, we will say that they are *modally equivalent*, in symbols $\langle \mathfrak{M}, w\rangle \equiv \langle \mathfrak{M}', w'\rangle$, in the case that $V(\varphi, w) = V'(\varphi, w')$ for every modal formula φ. Analogously we will talk about *multimodally equivalent* and *full modally equivalent*, in symbols $\equiv_{\mathsf{M}}$ and $\equiv_{\mathsf{F}}$, in the case we focus, respectively, on multimodal formulas or full modal formulas.

The first remark comparing expressive powers is that in every finite residuated lattice $\mathbf{A}$ (here it is not needed the chain assumption) it holds that

$$\Box\varphi \;\equiv\; \bigwedge\{\overline{a} \to \Box_a\varphi : a \in A \setminus \{0\}\} \tag{3}$$

Therefore, the modality $\Box$ is explicitly definable using the modalities $\Box_a$'s (and these last ones have the advantage that satisfy the normality axiom). In other words, the expressive power of the modal language is smaller than the one of the multimodal one. Thus, $\Lambda(\mathsf{Fr}, \mathbf{A^c})$ can be seen as a fragment of $\mathsf{M}\Lambda(\mathsf{Fr}, \mathbf{A^c})$. It is obvious that if two pointed Kripke models are multimodally equivalent then they are also modally equivalent.

What about the converse direction in the statements from last paragraph? That is, (i) is it possible to explicitly define the modalities $\Box_a$'s in the modal language?, and (ii) are modally equivalent pointed Kripke models also multimodally equivalent? In the rest of the section we will discuss these two questions. It is worth pointing out that as far as the authors are aware there is no general result in the many-valued modal setting relating these last two questions: an study of Beth definability in this setting has not been undertaken (cf. [9, p. 277]). As a matter of fact it is necessary to distinguish between having an empty set Var of propositional variables or not.

[8] By a *pointed Kripke model* we mean a Kripke model together with a distinguished point.

Proposition 4.2 (Case $\mathsf{Var} = \emptyset$)**.** *Let* $\mathbf{A}$ *be a finite* MTL *chain. Then, if two pointed Kripke models are modally equivalent then they are also multimodally equivalent.*

Proof. First of all we point out that it is enough to prove that if $\langle \mathfrak{M}, w\rangle \equiv \langle \mathfrak{M}', w'\rangle$, $w_0 \in W$ and $R(w, w_0) \geqslant a \neq 0$, then there is some $w_0' \in W'$ such that $R'(w', w_0') \geqslant a$ and $\langle \mathfrak{M}, w_0\rangle \equiv \langle \mathfrak{M}', w_0'\rangle$. The proof finishes by realizing that the previous statement is true because for every $a \in A \setminus \{0\}$, it holds that $V(\Box\overline{\mathsf{pred}(a)}, w) \neq 1$ iff there is some w_0 such that $R(w, w_0) \geqslant a$. The notation $\mathsf{pred}(a)$ refers to the predecessor of element a. □

Therefore, it is obvious that if $\mathsf{Var} = \emptyset$ then two pointed Kripke models are modally equivalent iff they are multimodally equivalent. It is still open whether the modalities $\Box_a$'s are explicitly definable or not when there are no propositional variables.

In the case that there is some propositional variable then the situation is quite different as next proposition shows. We will see later in Theorem 4.5 that the assumption in this proposition is also a necessary condition for the case of finite BL chains.

Proposition 4.3 (Case $\mathsf{Var} \neq \emptyset$)**.** *Let* $\mathbf{A}$ *be the ordinal sum* $\mathbf{A_1} \oplus \mathbf{A_2}$ *of two finite* MTL *chains such that* $\mathbf{A_1}$ *and* $\mathbf{A_2}$ *are non trivial (i.e.,* $\min\{|A_1|, |A_2|\} \geqslant 2$*). Then, there are two pointed Kripke models that are modally equivalent but not multimodally equivalent.*

Proof. Let us define $a \in A$ as the minimum element of $\mathbf{A_2}$ (i.e., a is the idempotent element separating both components). Since $\mathbf{A_i}$'s are non trivial it is obvious that $a \notin \{0^{\mathbf{A}}, 1^{\mathbf{A}}\}$, and also that for every $b \in A$, $a \odot b = a \wedge b$. Next we consider the two Kripke models given[9] in Fig. 1. It is obvious that for every modal formula φ it holds that $V(\varphi, w_1) = V'(\varphi, w_1')$. On the other hand, by induction on the length of formulas the reader can easily prove[10] that for every modal formula φ (we remind that canonical constants are allowed), it holds both that

- $a \leqslant V'(\varphi, w_1')$ iff $a \leqslant V'(\varphi, w_2')$,
- if $a \not\leqslant V'(\varphi, w_1')$ then $V'(\varphi, w_1') = V'(\varphi, w_2')$.

These last two properties guarantees that for every modal formula φ, it holds that $V'(\varphi, w_1') \leqslant a \to V'(\varphi, w_2')$; and so $V'(\Box\varphi, w_0') = V'(\varphi, w_1')$. Once we know this last fact, it becomes easy to show by an straightforward induction that for every modal formula φ, it holds that $V(\varphi, w_0) = V'(\varphi, w_0')$. Thus, $\langle \mathfrak{M}, w_0\rangle$ and $\langle \mathfrak{M}', w_0'\rangle$ are modally equivalent. Finally, using that $V(\Box_a p, w_0) = 1 \neq a = V'(\Box_a p, w_0')$ we get that $\langle \mathfrak{M}, w_0\rangle$ and $\langle \mathfrak{M}', w_0'\rangle$ are not multimodally equivalent. □

Lemma 4.4. *Let* $\mathbf{A}$ *be the finite* MV *chain of cardinal* n*. Then, for every* $a \in A \setminus \{0\}$ *it holds that*[11]

$$\Box_a\varphi \;\equiv\; \bigwedge\left\{\left(\overline{a} \to \neg\Box\neg((\varphi \leftrightarrow \overline{b})^{n-1})\right)^{n-1} \to \overline{b} : b \in A\right\}$$

[9] The convention here adopted about presenting Kripke models as diagrams is the same one stated in [1, Convention 3.6].

[10] In this inductive proof it plays a crucial role the definition of ordinal sum.

[11] Notice that φ^{n-1} only takes crisp values (i.e., in $\{0, 1\}$). Indeed, φ^{n-1} takes value 1 when φ takes value 1, and φ^{n-1} takes value 0 elsewhere. Hence, φ^{n-1} takes the same value than $\Delta\varphi$ where Δ is the well-known operator used in fuzzy logic (see [7]).

ISBN: 978-989-95079-6-8

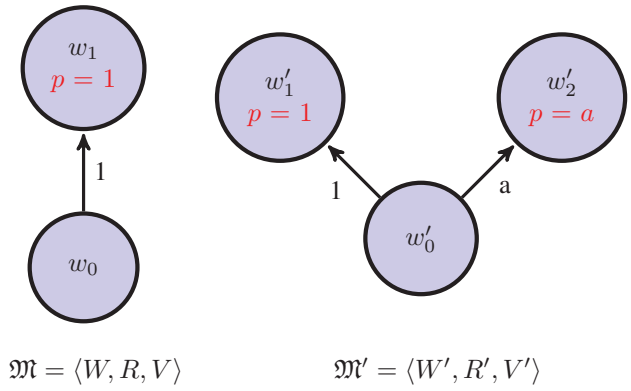


Figure 1: Two interesting Kripke models

Proof. Let us introduce the abbreviations $\Diamond\varphi := \neg\Box\neg\varphi$ and $\Delta\varphi := \varphi^{n-1}$. Then, the statement says that

$$\Box_a\varphi \;\equiv\; \bigwedge \{\Delta(\overline{a} \to \Diamond\Delta(\varphi \leftrightarrow \overline{b})) \to \overline{b} : b \in A\}$$

In this proof we will use the notation

$$f(w, \varphi, b) := \bigvee\{R(w, w') : w' \in W, V(\varphi, w') = b\}.$$

The reader can easily check the following steps.

$$V(\Diamond\Delta(\varphi \leftrightarrow \overline{b}), w) = f(w, \varphi, b)$$

$$V(\Delta(\overline{a} \to \Diamond\Delta(\varphi \leftrightarrow \overline{b})), w) = \begin{cases} 1, & \text{if } a \leqslant f(w, \varphi, b) \\ 0, & \text{if not} \end{cases}$$

$$V(\Delta(\overline{a} \to \Diamond\Delta(\varphi \leftrightarrow \overline{b})) \to \overline{b}, w) = \begin{cases} b, & \text{if } a \leqslant f(w, \varphi, b) \\ 1, & \text{if not} \end{cases}$$

Then, it easily follows that $V(\bigwedge \{\Delta(\overline{a} \to \Diamond\Delta(\varphi \leftrightarrow \overline{b})) \to \overline{b} : b \in A\}, w) = \bigwedge\{V(\varphi, w') : w' \in W, R(w, w') \geqslant a\} = V(\Box_a\varphi, w)$. This finishes the proof. □

Theorem 4.5 (Case $\mathsf{Var} \neq \emptyset$)**.** *Let* $\mathbf{A}$ *be a finite* BL *chain. Then, the following statements are equivalent.*

1. $\mathbf{A}$ *is a finite* MV *chain (i.e., the only idempotent elements of* A *are* 0 *and* 1*),*
2. *The modalities* $\Box_a$*'s are explicitly definable*[12] *in the modal language.*
3. *Two pointed Kripke models are modally equivalent iff they are multimodally equivalent,*

Proof. For the case that $|A_1| \leqslant 2$ it is trivial that these statements are equivalent. Hence, we only have to deal with the case that $|A_1| \geqslant 3$; and this case is a consequence of Proposition 4.3 and Lemma 4.4 together with the decomposition of BL chains as ordinal sums. □

5 Concluding Remarks

As it has been stressed in this paper, and also in [1], the role played by canonical constants is crucial in several of the proofs given above. For example, it is unknown to the authors whether there is a general method for converting an axiomatization of $\mathbf{\Lambda(A)}$ (and so without canonical constants) into one of $\Lambda(\mathsf{CFr}, \mathbf{A})$. One of the few statements that we know that remains true when we remove canonical constants from the language is Theorem 4.5. That is, for finite MV chains it is possible to find a modal formula (and so using $\Box$) without canonical constants that is equivalent to $\Box_a p$. The proof of this stronger version of Theorem 4.5 is completely different and is based on a combinatorial analysis of all possibilities; it is a rather involved proof and due to space limitations has not been included in this paper.

[12] That is, there is some modal formula φ such that $\varphi \equiv \Box_a p$. We remind that p refers to a propositional variable.

Acknowledgments

The collaborative work between the authors of this paper has been made possible by several research grants. The first three authors acknowlege partial support by the Spanish project MULOG2 TIN2007-68005-C04-01/04, the EU-ROCORES(LogICCC) project FP006 LoMoReVi (linked to the Spanish project FFI2008-03126-E/FILO), and the grants 2005SGR-00083 and 2005SGR-00093 of the Catalan Government. The last author has been partially financed by CyT-UBA project X484 and CONICET project PIP 112-200801-02543.

References

[1] F. Bou, F. Esteva, L. Godo, and R. Rodríguez. On the minimum many-valued modal logic over a finite residuated lattice, 2008. Submitted. Available at `http://arxiv.org/abs/0811.2107`.

[2] L. A. Zadeh. Similarity relations and fuzzy orderings. *Information Sciences*, 3:177–200, 1971.

[3] U. Höhle. Commutative, residuated l-monoids. In U. Höhle and E. P. Klement, editors, *Non-classical logics and their applications to fuzzy subsets (Linz, 1992)*, volume 32 of *Theory Decis. Lib. Ser. B Math. Statist. Methods*, pages 53–106. Kluwer Acad. Publ., Dordrecht, 1995.

[4] P. Jipsen and C. Tsinakis. A survey of residuated lattices. In J. Martinez, editor, *Ordered Algebraic Structures*, pages 19–56. Kluwer Academic Publishers, Dordrecht, 2002.

[5] H. Ono. Substructural logics and residuated lattices - an introduction. In V. F. Hendricks and J. Malinowski, editors, *50 Years of Studia Logica*, volume 21 of *Trends in Logic—Studia Logica Library*, pages 193–228. Dordrecht, 2003.

[6] F. Esteva and L. Godo. Monoidal t-norm based logic: towards a logic for left-continuous t-norms. *Fuzzy Sets and Systems*, 124:271–288, 2001.

[7] P. Hájek. *Metamathematics of fuzzy logic*, volume 4 of *Trends in Logic—Studia Logica Library*. Kluwer Academic Publishers, Dordrecht, 1998.

[8] R. Cignoli, I. M. L. D'Ottaviano, and D. Mundici. *Algebraic foundations of many-valued reasoning*, volume 7 of *Trends in Logic—Studia Logica Library*. Kluwer Academic Publishers, Dordrecht, 2000.

[9] P. Blackburn, J. F. A. K. van Benthem, and F. Wolter, editors. *Handbook of Modal Logic, Volume 3 (Studies in Logic and Practical Reasoning)*. Elsevier Science Inc., New York, NY, USA, 2006.

ISBN: 978-989-95079-6-8

Fuzzy Optimization in Vehicle Routing Problems

J. Brito, J. A. Moreno[1] J. L. Verdegay[2]

1 Grupo de Computación Inteligente, I.U.D.R.
Universidad de La Laguna, E-38271 La Laguna, Spain
2 Departamento de Ciencias de la Computación e Inteligencia Artificial
Universidad de Granada, E-18071 Granada, Spain
Email: {jbrito,jamoreno}@ull.es, verdegay@decsai.ugr.es

Abstract—*This paper focuses on analyzing practical solution approaches for Vehicle Routing Problems (VRP) with uncertain information. Several variants of the basic problem and fuzzy optimization problem formulations are described. The fuzzy VRP is obtained when some of the elements in the formulation are fuzzy. The main versions of Fuzzy VRP that have appeared in the literature are reviewed and the standard approaches for solving the corresponding models are analyzed.*

Keywords—Vehicle Routing Problem, Fuzzy Optimization, Fuzzy Vehicle Routing Problem, VRP with Time Windows.

1 Introduction

Route planning problems are among the activities that have the highest impact in logistical planning, transport and distribution because of their effects on efficiency in resource management, service levels, and client satisfaction. Route distribution planning problems, also known as Vehicle Routing Problems (VRP), have been thoroughly studied in a variety of areas, such as Operations Research, Artificial Intelligence, etc. The standard VRP was originally introduced by Dantzig and Ramser (1959), and is NP-hard, which is a complex combinatorial optimization problem [5]. Several variants of the basic problem have been put forward and strong formulations have been proposed. Most of these problems can be modeled as linear programming problems. The most common solution techniques are exact methods that guarantee finding an optimal solution if it exists. These approaches have also been applied together with numerous heuristics solution techniques developed with enough flexibility in optimization systems and can be adapted to various practical contexts.

Given that the complex, flexible, and dynamic nature of real logistical planning produces a high degree of uncertainty related to the decision making process, is not always possible to have all of the necessary information available at the onset of the problem. Consequently the most common scenario provides incomplete or imprecise information of the parameters and variables. The use of fuzzy sets to approach these situations is very appropriate to build computing systems for the solution of solving decision and optimization problems. The modeling of these problems is complicated at various levels: not only are they difficult to define accurately and need to manage uncertainty, but there is also imprecision in the available information and stated preferences, restrictions and objectives by decision makers.

A frequent occurrence in real decision-making problems, such as those found in VRP, is the lack of precision or uncertainty in the information available in them. This characteristic complicates the process concerning the definition of their objectives and parameters. Although this type of uncertainty found in the nature of the data and their settings has traditionally been handled by means of probability theory, in most of the cases they cannot be considered random phenomena and therefore probability theory cannot be applied successfully.

The aim of this paper is to analyze how the relevance and the variety of applications found in VRP can be considered under the fuzzy context point of view. The organization of the paper follows. Section 2 introduces the broad class of optimization problems known as VRP. We also survey the concept of fuzzy optimization where the objective function and constraints have fuzzy parameters, and then we apply these concepts to the simple and classical VRP. Finally, we review the corresponding fuzzy versions, which are called Fuzzy VRP.

2 Vehicle Routing Problems

The VRP are one of the most widely studied class of problems in combinatorial optimization, and the literature provides several exact and heuristic solution techniques of general applicability. VRP are problems where several vehicles that must serve points of demands and satisfy a finite set of constraints while minimizing costs, distances or times [20].

The standard VRP (usually called capacitated VRP; CVRP) calls for the determination of a set of *m* routes whose total travel length is minimized such that: (1) each customer is visited exactly once by one route, (2) each route starts and ends at a single depot, (3) the total demand of the customers served by a route does not exceed a given vehicle capacity Q, and (4) the length of each route does not exceed a preset limit L. A constant speed is typically assumed so that to minimize distances, travel times and travel costs are considered equivalent. If each vehicle *i* is assigned to a route R_i, a feasible solution for the VRP is made up of a partition from V into *m* routes $R_1, R_2, ..., R_m$ and the corresponding permutations σ_i of $R_i \cup 0$ that specify client order along the routes.

VRP allow us to consider several types of constraints, depending on the specific characteristics of the problem and decision-making process [4]. These possibilities lead to a variety of problems, beginning with the standard or basic VRP. Their descriptions follow:

ISBN: 978-989-95079-6-8

MFVRP- *Mix Fleet VRP*. This is a VRP in which the vehicles have different capacities, in other words, fleet of vehicles with a heterogeneous capacity. Each resource must consider these capacities for each route.

MDVRP- *Multi-depot VRP*. A company can have several depots or warehouses for use when supplying client demand. If clients are grouped around the depots then the problem can be viewed as a set of independent VRP problems. However, if the clients and depots are spread apart, then this scenario is known as a MDVRP.

PVRP - *Period VRP*. In the classical VRP the planning period of the route is one day. The PVRP expands the planning period to *M* days. During this period of *M* days, each client should be visited a given number of days.

SDVRP - *Split-up Delivery VRP*. This problem is a VRP in which a client is allowed to be serviced by several vehicles if the total cost decreases. Order size is important since the total size of the client orders must exceed the capacity of one vehicle.

PDVRP - *Pickup and Delivery VRP*. This variation is a VRP since it allows clients the possibility of returning certain goods that have to be delivered to other clients. Therefore, each vehicle that is fitted must consider client pickup and delivery amounts as well as a planned order.

VRPB – *VRP with Backhauls.* This problem is VRP where clients can order or return articles. Therefore the formulation requires that returned goods from clients fit in the vehicle. In addition, all deliveries must be completed before pickups begin because vehicles are back-loaded and pickups in the vehicles are considered uneconomical or unfeasible. Orders and pickups must be known ahead of time.

VRPTW- *VRP with Time Windows*. This variation is a VRP with the additional constraint that associates a time window to each client which identifies the only interval of time when the client is willing to receive goods or services. If a vehicle arrives to the client early the vehicle has to wait. If it arrives during the interval of the time window, the vehicle makes the delivery at that moment. Finally, if it arrives late the client is not serviced.

OVRP - *Open VRP*. This problem is a special class of VRP within all open delivery routes, since vehicles are not required to return to the depot.

Finally, other versions of VRP do exist. Some versions are mixes or hybrids of the previously mentioned typologies, such as the MDVRPTW or OPDVRP. Other problem types introduce specific characteristics which require the corresponding modifications in the model. For example, one version of a particularly interesting (and increasingly complex) problem is a dynamic version of the route problem with or without time windows (DVRP, Dynamic VRP). In this version the problem characteristics vary as time passes.

3 Fuzzy optimization problems

An *optimization problem* can be described as the search for the value of specific *decision variables* so that identified *objective functions* attain their optimum values. The value of the variables is subject to stated *constraints*. In these problems the objective functions are defined on a set of solutions that we will denote by *X*. The objective function is not subject to any condition or property nor is the definition of the set *X*. Typically the number of elements of *X* is very high, essentially eliminating the possibility of a complete evaluation of all its solutions while determining the optimal solution.

Optimization problems in their most general form involve finding an optimal solution according to stated criteria. In practice, however, many situations lack the exact information that is needed in the problem, including its constraints, or in other cases, where it is unreasonable to access such specific constraints or clearly defined objective functions. In these situations it is advantageous to model and solve the problem using soft computing and fuzzy techniques.

Among all the optimization problems, the models that have received the most attention and have offered the most useful applications in different areas are Linear Programming (LP) models, which is the single objective linear case with linear constraints. The classic problem of LP is to find the maximum or minimum values of a linear function subject to constraints that are represented by linear equations or inequalities. The most general formulation of the LP problem is:

$$\begin{aligned} &\max && z = cx \\ &\text{subject to} && Ax \le b \\ & && x \ge 0 \end{aligned} \qquad (1)$$

The vector $x = (x_1, x_2, ..., x_n) \in \mathbb{R}^n$ represents the decision variables. The objective function is denoted by *z*, the numbers c_j are coefficients and the vector $c = (c_1, c_2, ..., c_n) \in \mathbb{R}^n$ is known as the cost vector. The matrix $A = [a_{ij}] \in \mathbb{R}^{n \times m}$ is called the constraint or technological matrix and the vector $b = (b_1, b_2, ..., b_m) \in \mathbb{R}^m$ represents the independent terms or right-hand-side of the constraints.

In many real situations not all the constraints and objective functions can be valued in a precise way. In these situations we are dealing with the general problem form of Fuzzy Linear Programming (FLP). FLP is characterized as follows: a_{ij}, b_j and c_i can be expressed as *fuzzy numbers*, x_i as variables whose states are fuzzy numbers, addition and multiplication operates with fuzzy numbers, and the inequalities are among fuzzy numbers.

Different FLP models can be considered according to the elements that contain imprecise information that is taken as a basis for the classification proposed in [22], [2]. Short descriptions of these models follow.

Models with feasible fuzzy set (fuzzy constraints)

This is the case where constraints can be satisfied, and consequently the feasible region can be defined as a fuzzy set; it should be defined by means of a membership function $\mu : \mathbb{R}^n \to [0,1]$. In such a situation, for each constraint, a desirable quantity *b* is considered, but the possibility that it is greater is accepted until a maximum *b+t* (*t* is referred to as a violation *tolerance level*). This model is represented by:

ISBN: 978-989-95079-6-8

$$\begin{aligned} &\max && z = cx \\ &\text{subject to} && Ax \le_f b \\ & && x \ge 0 \end{aligned} \quad (2)$$

where the symbol $\le_f$ indicates the imprecision of the constraints and where each fuzzy constraint $a_i x \le_f b_i$ is specified by a membership function in the form:

$$\mu_i(a_i x) = \begin{cases} 1 & \text{if } a_i x \le b_i \\ f_i(a_i x) & \text{if } b_i \le a_i x \le b_i + t_i \\ 0 & \text{if } b_i + t_i \le a_i x \end{cases} \quad (3)$$

which means that, for each constraint *i*, given the level of tolerance t_i, to each point (*n*-dimensional vector) *x is* associated a number $\mu_i(x) \in [0,1]$ known as the degree of fulfillment (or verification) of the constraint *i*. The functions f_i are assumed to be continuous and monotonous non-decreasing functions. In particular, Verdegay [21], using the representation theorem for fuzzy sets, proves that the solutions for the case of linear functions f_i can be obtained from the auxiliary model:

$$\begin{aligned} &\max && z = cx \\ &\text{subject to} && Ax \le b + t(1-\alpha) \\ & && x \ge 0, \alpha \in [0,1] \end{aligned} \quad (4)$$

where $t = (t_1, t_2, ..., t_m)$.

Models with fuzzy goals

A optimization problem with fuzzy goals allows the objective function value to be slightly below the minimum goal for a maximization problem, and similarly for a minimization problem. The corresponding linear model is expressed in the following way:

$$\begin{aligned} &\widetilde{\max} && z = cx \\ &\text{subject to} && Ax \le b \\ & && x \ge 0 \end{aligned} \quad (5)$$

If t_0 is the maximum quantity that the objective function should be inferior to the minimum goal c_0, then each vector *x* is associated to a number $\mu_0(x)$, which represents the degree that the decision maker considers to be an achieved goal. It is defined according to the following function:

$$\mu_0(x) = \begin{cases} 1 & \text{if } cx > c_0 \\ f_0(cx) & \text{if } c_0 - t_0 \le cx \le c_0 \\ 0 & \text{if } cx < c_0 - t_0 \end{cases} \quad (6)$$

where f_0 is a continuous, monotonous non-decreasing function. An operative model that provides satisfactory solutions can be found in [27].

Models with fuzzy costs as objective function coefficients

These models are those whose costs are not fully known (with imprecision). Therefore, they are represented by an *m*-dimensional fuzzy vector $c^f = (c_1^f, c_2^f, ..., c_n^f)$, and the following model:

$$\begin{aligned} &\max && z = c^f x \\ &\text{subject to} && Ax \le b \\ & && x \ge 0 \end{aligned} \quad (7)$$

Evidently, *z* is also a fuzzy number, but *x* can be a vector of fuzzy or non-fuzzy numbers, and each fuzzy cost is described by its corresponding membership function $\mu_i(x)$. Each coefficient c_i^f of the objective function is a plane fuzzy number of the L-R type with modal interval $[\underline{c}_j, \overline{c}_j]$ and membership functions g_j and h_j (which can be linear, parabolic, etc.). Delgado et al. [7] prove that the solution can be obtained with the multi-objective auxiliary model:

$$\begin{aligned} &\max && z = \left[c^1 x, ..., c^{2^n} x \right] \\ &\text{subject to} && Ax \le b \\ & && x \le 0, \alpha \in [0,1], c_j^k \in \left\{ g_j^{-1}(1-\alpha), h_j^{-1}(1-\alpha) \right\} \\ & && k = 1, ..., 2^n, \; j = 1, ..., n. \end{aligned} \quad (8)$$

Models with fuzzy coefficients in the technological matrix

Consider a problem of this type:

$$\begin{aligned} &\max && z = cx \\ &\text{subject to} && A^f x \le_f b^f \\ & && x \ge 0 \end{aligned} \quad (9)$$

where the values of the technological matrix and the coefficients are fuzzy numbers. Fuzzy constraints can also be included. Delgado et al. [6] also include imprecision in the constraints. They propose considering fuzzy solutions that are solved with the application of an ordered function *g* for the constraints. This new formulation is expressed by the auxiliary problem:

$$\begin{aligned} &\max && z = cx \\ &\text{subject to} && a_i^f x \le_g b_i^f + t_i^f (1-\alpha), \; i = 1, ...m \\ & && x \ge 0, \alpha \in [0,1] \end{aligned} \quad (10)$$

where the symbol $\le_g$ stands for a comparison relation between fuzzy numbers.

4 Fuzzy optimization Vehicle Routing Problems

Although there are different stochastic approaches to modeling and solving the VRP in the literature, this is not the case with the proposed approaches from fuzzy set theory. Furthermore, the literature offers very little in terms of modeling VRP proposals, both from the standpoint of the solutions modeled as Fuzzy Mathematical Programming, and Fuzzy VRP (FVRP). However, as we will discuss later, several proposals have been introduced in recent years, such as the VRPTW and Dynamic VRP.

Specifically, if we look at the FVRP models in the literature, the majority only assume vagueness for some of the following elements that are described in the model: a) Fuzzy demands (to be collected): customer demand is a imprecise variable and 2) Fuzzy times: service time and travel time can be imprecise variables.

The first problem deals with the demand vector of each client's ordered goods. Planning the ordered quantity by the clients is difficult to establish with sufficient notice and precise form, therefore we do not have access to a specific quantity. In other words, the information about vehicle demand at some nodes is often not precise enough. Thus,

ISBN: 978-989-95079-6-8

there is often uncertainty regarding the amount of demand at some nodes. This FVRP with fuzzy customer demand was first instructed by Teodorovic and Kikuchi [18]. In this paper, they treated the travel time and the transportation costs between two nodes in a network as fuzzy numbers. They modified the Clarke and Wright algorithm where travel times in a network are treated as fuzzy numbers. Later, Teodorović and Pavković [19] solve a VRP when demand at the nodes is uncertain and is represented by a triangular fuzzy number. The model is based on the heuristic "sweeping" algorithm, which uses fuzzy approximate reasoning procedures to decide whether or not to include a node in the route. It first uses the approximate reasoning algorithm to calculate the preference index. Once the membership function of the preference index has been determined, defuzzification must take place. In recent years, these same authors have proposed solutions to this problem [15], [16], where actual demand value is known only after the visit to the node. Their solution combines Bee and Ant systems with rules based on fuzzy logic. A new paper [14] has recently appeared which considers the VRP with uncertain demand at node. It uses the approximate reasoning algorithm to determine the preference strength to send the vehicle to next node, and the improved sweeping algorithm with vehicle coordinated strategy to determine a set of vehicle routes that minimizes costs.

The second problem, (fuzzy times in both service and travel) is characterized by other pieces of information that are increasingly imprecise, given the daily circumstances found in routing networks and traffic. In these cases service time, time windows and travel time are expressed as fuzzy numbers.

The traditional deterministic VRP is expanded to the situation so that the VRP has fuzzy travel time features. In [11] a simple description of the VRP with fuzzy traveling time, that is, a mathematical model for the problem, is built. It puts forward the concept of level effect function which can quantify the location of fuzzy number intensively and globally, and sets up the uncertain degree of measurement. In this paper the solution to the problem is based on a genetic algorithm.

In [17] the travel time based on the fuzzy mathematical model of the VRP takes the time window as a fuzzy variable. Information entropy and the path chosen by the use of random disturbance control strategy to the Ant algorithm is used to the vehicle routing problem with fuzzy travel time.

Reference [9] focuses on modeling and solution of the dynamic VRP with time-dependent and fuzzy travel time. A model of this problem is set up based on fuzzy service times of the customer, its demand and its time windows, which is regarded and ordering as a triangular fuzzy number. A hybrid genetic algorithm, which is seasoned with the model and combined with the ant colony algorithm, is presented.

In [10] the authors present a model of the real world vehicle routing and dispatching problem. The time-dependent and fuzzy travel speeds are introduced into the model. A dispatching period is divided into some time slices and each time slice is designated a triangular fuzzy speed. The method of comparing two triangular fuzzy numbers is applied to check whether or not customers' time windows are satisfied. A hybrid intelligent approach combining a genetic algorithm and an ant colony algorithm is proposed for solving the dispatching model.

Results in both types of models with uncertainty in demand or in time have been published based on fuzzy variables, fuzzy random variables, stochastic programming and Chance Constrained Programming (CCP). These concepts were introduced by Charnes and Cooper and later by Liu, and generally are applied with heuristics to find solutions to the VRP [26], [24].

In [13] the author considers the VRP with time window while assuming that the travel times cannot be precisely known, but can be regarded as fuzzy variables. Since the travel times are fuzzy variables, every customer will be visited at a fuzzy time. Credibility is introduced as a measure of confidence in the constraints so that it ensures that all customers are visited within their time windows with a confidence level, then following Chance Constraint Programming and also a hybrid intelligent algorithm by integrating fuzzy simulation and GA to solve the VRP.

To the best of our knowledge, there is little evidence in the literature on the properties of mathematical programming with random fuzzy coefficients and VRP with random fuzzy demands. The only related study appears to be a paper [8] which proposed a method to solve a class of model with random fuzzy coefficients in both the objective functions and constraint functions and applied it in the CVRP with random fuzzy demands. Based on the concept of random fuzzy variables introduced in [3], the objective is to provide workable formulations and exact algorithms for a class of uncertainty.

Reference [12] deals with a variation of uncertain VRP where the customers' demands are random fuzzy variables and the travel times between customers are random variables. The travel times between customers follow given probability distributions. The authors develop a programming model with random fuzzy and random variables for VRP which consider capacity and arrival time constraints and present a stochastic programming formulation which includes probabilistic constraints and apply a pure genetic algorithm (GA).

In [23] and [25] the authors considered a fuzzy multi-objective modeling approach for capacitated VRP with fuzzy random parameters. The first paper is based on the mixed integer linear programming model, and suggested an interactive heuristics approach using triangular fuzzy number and crisp equivalent formulae to address the demand solution. The authors in the second paper use travel time and customer demand and treat them as fuzzy random variables and chance-constrained programming is presented and converted to a crisp equivalent model under some assumptions. The authors present a hybrid multi-objective particle swarm optimization that incorporates specific heuristics to solve such a problem.

All of the above VRP can be formalized as problems of combinatorial optimization. These problems in their most general form, have integer LP formulations, based on the proposal given by Bodin *et al*. [1], and can be described, assumed that the depot is the node 0, N is the number of

ISBN: 978-989-95079-6-8

customers to be served by K vehicles and the decision variables are $x_{ij}^k \in \{0,1\}$, $i,j = 1, 2, \ldots, N$, $k = 1, 2, \ldots, K$, where $x_{ij}^k = 1$ if vehicle k travels from customer i to j and 0 otherwise.

In particular, objectives and constraints may be formulated as follows.

Objectives

a. If c_{ij}^k is the *cost of travelling* from customer i to customer j by vehicle k. The total travel cost is an objective function to be minimized and is expressed by:

$$\min \sum_{k=1}^{K} \sum_{i=0}^{N} \sum_{j=0}^{N} c_{ij}^k x_{ij}^k \tag{11}$$

In general, this cost can be expressed as total time or distance traveled by the vehicles.

b. If t_{ij}^k denotes the time needed to go from customer i to customer j and δ_i^k is the required time by vehicle k to unload the demand to the customer i. The objective function is to minimize the total travel time.

$$\min \left(\sum_{k=1}^{K} \sum_{i=0}^{N} \sum_{j=0}^{N} t_{ij}^k x_{ij}^k + \sum_{k=1}^{K} \sum_{i=0}^{N} \sum_{j=0}^{N} \delta_i^k x_{ij}^k \right) \tag{12}$$

c. Other possible objective function is to maximize the number of customers served by the vehicles.

$$\max \sum_{k=1}^{K} \sum_{i=0}^{N} \sum_{j=0}^{N} x_{ij}^k \tag{13}$$

Constraints

Constraints (14) and (15) ensure that each customer is served exactly once.

$$\sum_{k=1}^{K} \sum_{i=0}^{N} x_{ij}^k = 1, \quad j = 1,2,...,N \tag{14}$$

$$\sum_{k=1}^{K} \sum_{j=0}^{N} x_{ij}^k = 1, \quad i = 1,2,...,N \tag{15}$$

Constraint (16) ensures route continuity.

$$\sum_{i=0}^{N} x_{it}^k - \sum_{j=0}^{N} x_{tj}^k = 0, \quad t = 1,2,...,N; k = 1,2,...,K \tag{16}$$

Constraint (17) shows that the total length of each route k has a limit D_k, with d_{ij}^k the distance from i to j.

$$\sum_{k=1}^{K} \sum_{i=0}^{N} d_{ij}^k x_{ij}^k \le D_k, \quad k = 1,2,...,K \tag{17}$$

Constraint (18) ensures that the total demand of any route must not exceed the capacity Q_k of the vehicle k, with the demand of customer i, q_i.

$$\sum_{j=0}^{N} q_j \left(\sum_{i=0}^{N} x_{ij}^k \right) \le Q_k, \quad k = 1,2,...,K \tag{18}$$

Constraints (19) and (20) ensure that each vehicle is used no more than once.

$$\sum_{j=0}^{N} x_{0j}^k \le 1, \quad k = 1,2,...,K \tag{19}$$

$$\sum_{i=0}^{N} x_{i0}^k \le 1, \quad k = 1,2,...,K \tag{20}$$

This formalization assumes that the decision-maker has access to specific information on the components that define the problem; that is, on objective functions and constraints. However in real world problems it is more common that the information one has is actually imprecise or incomplete.

Real world situations reveal the difficulties of analyzing different types of models and problems that are affected by incomplete, imprecise or vague information or constraints. These problems, however, can be modeled in terms of fuzzy sets, leading to the field of fuzzy optimization. The solutions to these problems are fuzzy solutions. Discussions concerning solutions do not focus on their feasibility, nor if they are optimal solutions or not. We, in turn, have chosen to discuss the degree of feasibility and optimality of the solution.

Thus, if we suppose that any of the parameters: cost c_{ij}^k, distance d_{ij}^k, times t_{ij}^k and δ_i^k, and demand q_i can be fuzzy, the traditional model becomes a Fuzzy VRP. Intuitively, when any of these quantities are fuzzy numbers, the objective functions become fuzzy as well. If these parameters are approximately known, they can be represented by the fuzzy numbers $\tilde{c}_{ij}^k$, $\tilde{d}_{ij}^k$, $\tilde{t}_{ij}^k$, $\tilde{\delta}_i^k$ and $\tilde{q}_i$ respectively, with their corresponding membership functions. Then, for instance, the objective function can be expressed as:

$$\widetilde{\min} \sum_{k=1}^{K} \sum_{i=0}^{N} \sum_{j=0}^{N} \tilde{c}_{ij}^k x_{ij}^k \tag{21}$$

Similar changes occur with other objective functions. In the same way, constraints (17) would be expressed by:

$$\sum_{k=1}^{K} \sum_{i=0}^{N} \tilde{d}_{ij}^k x_{ij}^k \preceq \tilde{D}_k, \quad k = 1,2,...,K \tag{22}$$

with similar changes to the other constraints. However, constraints (14) to (16) are the same as those found in the crisp model.

Note that quantities Q_k and D_k also must be considered as fuzzy parameters. In addition, the summation symbol Σ in the objective functions and constraints refers to an addition of fuzzy numbers and $\preceq$ a fuzzy relation between fuzzy numbers. The meaning of min operator $\widetilde{\min}$ is also ambiguous, because it depends on the fuzzy numbers ranking index to be used. Hence there is a need to seek appropriate procedures for its solution.

Following the pattern in general Fuzzy Optimization described in a previous section, four different types of problems can be considered. Two of these problems include imprecision/uncertainty in the objective function(s), such as the case with fuzzy goals and the case with fuzzy costs. The remaining two problems consider fuzzy comparison in the constraints and in the coefficients of the technological matrix. In addition a fifth problem, the general fuzzy problem could be studied in which all of the parameters will be subject to fuzzy considerations.

ISBN: 978-989-95079-6-8

In practice, the search for optimal solutions to FVRP can be done with the following approaches. The simplest approach applies procedures for the fuzzification and defuzzification of variables. It transforms the imprecise information in fuzzy parameters and uses procedures that integrate fuzzy arithmetic to obtain fuzzy solutions. The fuzzy solution is then transformed into a crisp one using some known formulation. This approach may also be used for the introduction of sophisticated fuzzy rules in the decision-making processes to improve their quality. Linguistic variables could be used to facilitate the incorporation of "intelligent" procedures as automatic reasoning, adaptive control or automatic learning.

Another common approach consists in applying the Theory of Possibility and Chance Constraint Programming. It treats the fuzzy models, becoming the crisp equivalent, and by adding Chance Constraints changes the model. It assumes that the constraints will hold with at least a possibility α, and the chance is represented by the possibility that the constraints are satisfied.

The previous approaches combined with some metaheuristics provide robust optimization methods to obtain efficient solutions. Although in each planning problem we can establish a set of constraints, many of which can define different planning problems, this set which is defined with parameters and imprecise variables can also be considered the same if the constraints are not necessarily strictly or exactly established. These parameters or imprecise variables could include the times that cannot exceed a specified bound, or the time windows required to visit clients. In other words, it is possible to establish problems with constraints where not all of the constraints need to be satisfied and with the same degree of precision. In this formulization constraints can be called hard, which represent those constraints that must be satisfied exactly, or soft, which are not subject to the same demanding criteria.

5. Conclusions

Fuzzy Vehicle Routing Problems have been analyzed. Strong formulations are available but most algorithmic development focuses on a limited number of prototype problems. We analyze the use of sufficiently flexible and comprehensive fuzzy approaches to address the various imprecision required in practice.

Acknowledgements

The two first authors have been supported by projects TIN2008-06873-C04-01 (70% FEDER) from the Spanish Government and PI042005/044 from the Canary Islands Government. J.L. Verdegay has been supported by Project TIC-02970-JA.

References

[1] B. Bodin et al., 1983. Routing and scheduling of vehicles and crews: The state of the art. *Computers and Operations Research* 10, pp. 63–211. MathSciNet.

[2] J.M. Cadenas, J.L. Verdegay, 1999 *Optimization Models with Imprecise Data*. Ed. Universidad de Murcia. (in spanish).

[3] A. Charnes, W.W. Cooper, 1959. Chance-Constrained Programming. *Management Science*, 6:73-79.

[4] J.F. Cordeau et al., 2007. Vehicle Routing. *Handbook in OR & MS*, Vol. 14. Chapter 6. C. Barnhart and G. Laporte (Eds.), Elsevier.

[5] G.B. Dantzig, R.H. Ramser, 1959. The truck dispatching problem. *Management Science*, 6:80-91.

[6] M. Delgado, J.L. Verdegay and M.A. Vila., 1987. Imprecise costs in mathematical programming problems. *Fuzzy Sets and Systems*, 29, 21-29.

[7] M. Delgado, J.L. Verdegay and M.A. Vila, 1989. A general model for fuzzy linear programming. *Control and Cybernet*, 16 (2), 113–121.

[8] Y. He, J. Xu, 2005 A class of random fuzzy programming model and its application to vehicle routing problem. *World Journal of Modelling and Simulation* Vol. 1. N. 1, pp. 3-11

[9] L. Hong, M. Xu, 2008. A model of MDVRPTW with fuzzy travel time and time-dependent and its solution. In proceeding, *Fifth International Conference on Fuzzy Systems and Knowledge Discovery*. IEEE.

[10] L. Hong, M. Xu,, 2008. Real Vehicle Routing and Dispatching with Dynamic Fuzzy Travel Times. In proceeding, *Second International Conference on Genetic and Evolutionary Computing*. IEEE.

[11] J. Jia, et al., 2008. Genetic algorithm for fuzzy logistics distribution vehicle routing problem. In *proceeding International Conference on Service Operations and Logistics, and Informatics*. IEEE.

[12] B. Liu, 2001. Fuzzy random Chance-Constrained Programming. IEEE *Transactions on Fuzzy systems* Vol. 9, N. 5.

[13] B. Liu, 2002.*Theory and practice of uncertain programming*. Physica-berlag.

[14] C.-S. Liu, M-Y. Lai, 2009. The vehicle routing problem with uncertain demand at nodes. *Transportation Research Part E: Logistics and Transportation Review.* Available on line, 21/01/2009. doi:10.1016/j.tre.2008.11.002

[15] P. Lucić, D. Teodorović, 2003. Vehicle routing problem with uncertain demand at nodes: The Bee System and fuzzy logic Approach. *In Fuzzy Sets Based Heuristics for Optimization*, 67-82 (Editor J.L. Verdegay), Springer – Verlag.

[16] P. Lucić, D. Teodorović, 2007. The fuzzy ant system for the vehicle routing problem when demand at nodes is uncertain. *Journal on Artificial Intelligence Tools (IJAIT)*.Vol, 16 Issue: 5, 751-770.

[17] L. Tang, et al. 2007. Ant Colony Algorithm Based on Information Entropy Theory to Fuzzy Vehicle Routing Problem. *In proceedings ISKE*, series: Advances in Intelligent Systems Research.

[18] D. Teodorovic, S. Kikuchi, 1991. Application of fuzzy sets theory to the saving based vehicle routing algorithm. *Civil Engineering Systems* 8, pp. 87–93.

[19] D. Teodorovíc, G. Pavkovíc, 1996. The fuzzy set theory approach to the vehicle routing problem when demand at nodes is uncertain. *Fuzzy Sets and Systems* 82, 307-317. Elsevier.

[20] P. Toth, D.Vigo, 2002. The vehicle routing problem. *Monographs on Discrete Mathematics and Applications*, Vol. 9. SIAM

[21] J.L. Verdegay, 1982. *Fuzzy mathematical programming* In: Gupta MM, Sanchez E (eds) Fuzzy Information and Decision Processes.

[22] J.L. Verdegay, 1995. Fuzzy optimization: models, methods and perspectives. *6thIFSA-95 World Congress*. p. 39-71.

[23] B. Werners, M. Drave, 2003. Capacitated vehicle routing problem with fuzzy demand. *In Fuzzy Sets Based Heuristics for Optimization*, 317-334 (Editor J.L. Verdegay), Springer – Verlag.

[24] J. Xu, Y. He, 2004. A class of random fuzzy linear programming model and its application to vehicle routing problem. *Tech. report*, Sichuan University.

[25] Y. Yang , J. Xu, 2008. A class of multi-objective vehicle routing optimal model under fuzzy random environment and its application. *World Journal of Modelling and Simulation* Vol. 4 (2008) No. 2, pp. 112-119.

[26] Y. Zheng, and B. Liu, 2004. Fuzzy vehicle routing model with credibility measure and its hybrid intelligent algorithm, *Tech. report*, Tsinghua University.

[27] HJ. Zimmermann, 1976. Description and optimization of fuzzy system. *International Journal of general System*, 2, 209-216.

ISBN: 978-989-95079-6-8

Weighted decisions in a Fuzzy Random Forest

P.P. Bonissone[1] J.M. Cadenas[2] M.C. Garrido[2] R.A. Díaz-Valladares[3] R. Martínez[2]

1. GE Global Research. One Research Circle.
Niskayuna, NY 12309. USA
2. Dept. Ingeniería de la Información y las Comunicaciones. Universidad de Murcia.
Murcia, Spain
3. Dept. Ciencias Computacionales. Universidad de Montemorelos.
Montemorelos, México
Email: bonissone@crd.ge.com,{jcadenas,carmengarrido}@um.es,rdiaz@um.edu.mx,raquel.m.e@alu.um.es

Abstract— *A multi-classifier system - obtained by combining several individual classifiers - usually exhibits a better performance (precision) than any of the original classifiers. In this work we use a multi-classifier based on a forest of randomly generated fuzzy decision trees (Fuzzy Random Forest), and we propose a new method to combine their decisions to obtain the final decision of the forest. The proposed combination is a weighted method based on the concept of local fusion and on the data set Out Of Bag (OOB) error.*

Keywords— combination methods, fuzzy trees, local fusion, multi-classifier, random forest

1 Introduction

We have witnessed a variety of methods and algorithms addressing the classification issue. In the last few years, we have also seen an increase of multi-classifiers based approaches, which have shown to deliver better results than individual classifiers [10].

The multi-classifier Fuzzy Random Forest [2], doesn't aim to obtain the best system multi-classifier. Rather, we will focus on how to start from a multi-classifier system with performance comparable to the best classifiers and extend it to handle and manipulate imperfect information (linguistic labels, missing values, noise in nominal attributes, etc.) maintaining a proper performance. To build the multi-classifier, we follow the methodology of random forest and to incorporate the processing of imperfect data, we construct the random forest using fuzzy trees as base classifiers. Therefore, we try to use the robustness of a tree ensemble, the power of the randomness to increase the diversity of the trees in the forest, and the flexibility of fuzzy logic and fuzzy sets for imperfect data managing.

Usually, the majority vote weighted method has been the most common combination method used to combine the outputs of different classifiers. This combination method is optimal in the case of two classes and classifiers with independent outputs. In addition, although the classifiers are designed independently, it will be unlikely that they produce independent outputs [9]. In [2] we have presented various combination methods to obtain the final decision of the proposed multi-classifier.

In this work, we propose new inference implementations to obtain the decision of the multi-classifier Fuzzy Random Forest and we compare them with two of the best performing implementations in Fuzzy Random Forest: the proposed implementations are weighted implementations based on local fusion concept [3] and we compare them with weighted implementations based in the OOB cases.

In the section 2, we review the major elements that constitute a multi-classifier and some attempts for the incorporation of fuzzy logic in some techniques. In section 3, we provide a brief description of the Fuzzy Random Forest multi-classifier, showing the learning and inference phases. In section 4, we present the inference implementations that are the subject of this work. We present some comparative results of the analyzed implementations in section 5. Finally, we present our conclusions in section 6.

2 Multi-classifiers and Fuzzy Logic

When individual classifiers are combined appropriately, we usually obtain a better performance in terms of classification precision and/or speed to find a better solution. Multi-classifiers are the result of combining several individual classifiers. Multi-classifiers differ among themselves by their diverse characteristics: the number and the type of the individual classifiers; the characteristics of the subsets used by every classifiers of the set; the consideration of the decisions; and the size and the nature of the training sets for the classifiers [9].

Segrera [10] divides the methods for building multi-classifiers in two groups: ensemble and hybrid methods. The first type, such as Bagging and Boosting, induces models that merge classifiers with the same learning algorithm, while introducing modifications in the training data set. The second, type such as Stacking, creates new hybrid learning techniques from different base learning algorithms.

An ensemble uses the predictions of multiple base-classifiers, typically through majority vote or averaged prediction, to produce a final ensemble-based decision. The ensemble-based predictions typically have lower generalization error rates than the ones obtained by a single model. The difference depends on the type of base-classifiers used, ensemble size, and the diversity or correlation between classifiers [1]. Ahn [1] indicates that, over the last few years, three ensemble-voting approaches have received attention by researchers: boosting, bagging and random subspaces.

ISBN: 978-989-95079-6-8

2.1 Random Forest: A Multi-Classifier based on Decision Trees

Breiman [4] further defines a random forest as a classifiers composed by decision trees where every tree h_t has been generated from the set of data training and a vector θ_t of random numbers identically distributed and independent from the vectors θ_1, θ_2,..,θ_{t-1} used to generate the classifiers $h_1, h_2, .., h_{t-1}$. Each tree provides his unitary vote for the majority class given the entry. Examples of random forest are: randomization, Forest-RI and Forest-RC, double-bagging.

Hamza [6] concludes that: Random Forests are significantly better than Bagging, Boosting and a single tree; their error rate is smaller than the best one obtained by other methods; and they are more robust to noise than the other methods. Consequently, random forest is a very good classification method with the following characteristics: it's easy to use; it does not require models, or parameters to select except for the number of predictors to choose at random at each node.

2.2 Fuzzy Logic and Decision Trees

Decision tree techniques have proved to be interpretable, efficient and capable of treating with applications of great scale. However, they are highly unstable when small disturbances are introduced in data learning. For this reason, the fuzzy logic has been incorporated in the decision tree construction techniques.

Fuzzy logic offers an improvement in the disadvantages previously commented on the decision trees due to the elasticity of the fuzzy set's formalism. In [7, 8] we can find approaches in which fuzzy sets and their underlying approximate reasoning capabilities have been successfully combined with decision trees. This combination has preserved the advantages of both components: uncertainty management with the comprehensibility of linguistic variables, and popularity and easy application of decision trees. The resulting trees show an increased immunity to noise, an extended applicability to uncertain or vague contexts, and a support for the comprehensibility of the tree structure, which remains the principal representation of the resultant knowledge.

In the literature, we can find several proposals for building trees of fuzzy information starting with algorithms already known for building traditional trees. Fuzzy CART [7] was one of the first examples of this approach, being based on the CART algorithm. However, most authors have preferred to use the ID3 algorithm to construct trees for recursive partition of the data set of agreement to the values of the selected attribute. To use the ID3 algorithm in the construction of fuzzy trees, we need to develop attribute value space partitioning methods, branching attribute selection method, branching test method to determine the degree to which data follow the branches of a node, and leaf node labeling methods to determine classes [8].

Fuzzy decision trees are constructed in a top-down manner by recursive partitioning the training set into subsets. Some particular features of fuzzy tree learning are: the member-ship degree of examples, the selection of test attributes, the fuzzy tests (to determine the membership degree of the value of an attribute to a fuzzy set), and the stop criteria (besides the classic criteria when the measure of the information is under a specific threshold) [8].

3 Fuzzy Random Forest

Following Breiman's methodology [4], FRF is a multi-classifier that is a forest of randomly-generated fuzzy decision trees. In this section we provide a brief description of the learning and inference phases in FRF.

3.1 Fuzzy Random Forest Learning

To built a Fuzzy Random Forest (FRF) we use the algorithm 1.

FRFLearning (M,Fuzzy Random Forest)
begin
1. Take a random sample of N observations from the data set with replacement of the complete set of M observations.
2. Apply algorithm 2 (construct a fuzzy tree) to the subset of examples.
3. Repeat steps 1 and 2 up to building all fuzzy trees.
end.

Algorithm 1. FRF learning

Each tree in the forest will be a fuzzy tree generated following the guidelines of [8], adapting it where is necessary. So, the following algorithm (Algorithm 2) has been designed so that the trees could be constructed without considering all the attributes to split the nodes. We select the set of attributes as a random subset of the total set of available attributes. Then, we perform a new random selection for each split. As in the original case, some attributes (inclusive the best) might not be considered for each split, but a attribute excluded in one split might be used by other splits in the same tree.

FuzzyDecisionTree(Examples,Fuzzy Tree)
begin
1. Start with examples set of entry, having the weights of the examples (in root node) equal to 1.
2. At any node N still to be expanded, compute the number of examples of each class. The examples are distributed in part or in whole by branches. The distributed amount of each example to a branch is obtained as the product of its current weight and the membership degree to the node.
3. Compute the standard information content.
4. At each node search the set of remaining attributes to split the node.
4.1.Select with any criteria, the candidate attributes set to split the node.
4.2. Compute the standard information content to each child node obtained from from each candidate attribute.
4.3.Select the candidate attribute such that information gain is maximal.
5. Divide N in sub-nodes according to possible outputs of the attribute selected in the previous step.
6. Repeat steps 2-5 to stop criteria is satisfied in all nodes.
end.

Algorithm 2. Fuzzy decision tree learning

With Algorithm 2, we integrate the concept of fuzzy tree within the design philosophy of Breiman's random forest. In this way, we augment the capacity of diversification of random forests with the capacity of approximate reasoning of fuzzy logic.

When we built a Fuzzy Random Forest using the previous algorithm, about $1/3$ of the cases are excluded from each tree

ISBN: 978-989-95079-6-8

in the forest. These cases are called the "out of bag" (OOB); each tree will have a different set of OOB cases. The OOB cases are not used to build the tree and constitute an independent test sample for the tree. We use the set of OOB cases to obtain one of the weighted combination methods proposed in this work.

3.2 Fuzzy Random Forest Inference

In this section we will describe how the inference is carried out using Fuzzy Random Forest. First, we introduce the notation that we will use in the following sections. Then, we define two general strategies to obtain the decision of the forest for a target example. Concrete instances of these strategies are defined in the next section where we present different inference implementations in the Fuzzy Random Forest.
Notations

- T is the trees' number of the forest. We will use the index t to refer to a particular tree in the forest.
- N_t is the number of reached leaf nodes by an example, in the tree t. A characteristic inherent in fuzzy trees is that the classification of an example can derive in two or more leaves due to the overlapping of the fuzzy sets. We will use the index n to refer to a particular leaf in a tree of the forest.
- I is the number of classes that we consider. We will use the index i to refer to a particular class.
- e is an example from the dataset used to built and to infer with the Fuzzy Random Forest.
- $\chi_{t,n}(e)$ is the grade which the example e active the leaf n from t tree.
- $\overline{\omega}_{t,n}$ is a vector with I elements indicating the weights of the I possible classes in the leaf n of tree t, $\overline{\omega}_{t,n}$=$(\omega_{t,n,1}, \omega_{t,n,2}, ..., \omega_{t,n,I})$, where $\omega_{t,n,i} = \frac{E_i}{\sum_{j=1}^{I} E_j}$ and E_j is the sum of the weights of examples with class j in the leaf.

To make the inference with the Fuzzy Random Forest, we will define the fuzzy classifier module. The fuzzy classifier module operate on fuzzy trees of the forest with two possible strategies:

Strategy 1: Combining the information from the different reached leaves in each tree to obtain the decision of each individual tree and then apply the same one or another combination method to generate the global decision of the forest (Fig. 1).

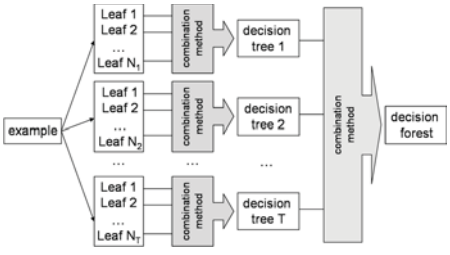


Figure 1: Inference with strategy 1

Strategy 2: Combining the information from all reached leaves from all trees to generate the global decision of the forest (Fig. 2).

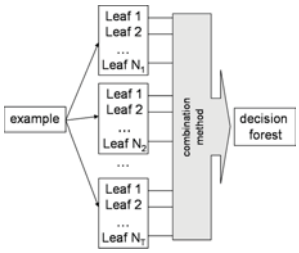


Figure 2: Inference with strategy 2

Moreover, in strategies 1 and 2 we use two functions: $K(\cdot,\cdot,\cdot)$ and $Faggre$. With $K(\cdot,\cdot,\cdot)$ we obtain information about the classes that provides a leaf reached n in a tree. This function depends of $\chi_{t,n}(e)$ and vector $\overline{\omega}_{t,n}$ and returns the weight assigned by the node n to class i. Concrete cases of this function will be defined in the next section. And, $Faggre$ is defined as a frequently used multi-classifiers combination method [9], e.g., majority vote, minimum, maximum, average, product or a weighting combination method defined in the next section.

In Algorithm 3 we implement strategy 1. In this algorithm first, $Faggre$ is used to obtain $A[t][i]$ (the weight of each tree for each class). In this case, $Faggre$ aggregates the information provided by the reached leaves in a tree. Later, the values obtained in each tree t, will be aggregating by means of the function $Faggre$ to obtain the vector $\overline{F}$ that contains the weight proposed by the Fuzzy Random Forest for the different classes. This algorithm takes a target example e and the random forest, and generates the vector $\overline{F}$ as output.

```
Begin
  TreeClasification.
  ForestClasification.
End

TreeClasification (in : e, in : Random Forest, out : A)
Begin
  For each Tree t
    For each Class i
      A[t][i] = Faggre(K(1, χ_{t,1}(e), ω_{t,1}), ...,
        K(1, χ_{t,N_t}(e), ω_{t,N_t}), ..., K(I, χ_{t,1}(e), ω_{t,1}),
        , ..., K(I, χ_{t,N_t}(e), ω_{t,N_t}))
    End For each Class
  End For each Tree
End

ForestClasification (in : A, out : F̄)
Begin
  For each Class i
    F[i] = Faggre(A[1][i], ..., A[T][i])
  End For each Class
End
```

Algorithm 3. FRF Inference (Strategy 1)

To implement strategy 2, the previous algorithm is simplified so that it does not add the information for each tree, but

ISBN: 978-989-95079-6-8

uses directly the information of all leaves reached by the example e in the different trees of the forest. The algorithm 4 implement the strategy 2 and uses as target values the example e to classify and the random forest and provides the vector $\overline{F}$ that contains the weight proposed by fuzzy random forest for the different classes.

ForestClasification ($in : e, in : Random\ Forest$,$out : \overline{F}$)
Begin
  For each Class i
    $F[i] = Faggre(K(i, \chi_{1,1}(e), \omega_{1,1}), ...,$
      $K(i, \chi_{1,N_1}(e), \omega_{1,N_1}), ..., K(i, \chi_{T,1}(e), \omega_{T,1}),$
      $..., K(i, \chi_{T,N_T}(e), \omega_{T,N_T}))$
  End For each Class
End

Algorithm 4. FRF Inference (Strategy 2)

4 Inference Implementations

In the previous section we have shown the general scheme of inference that we use to obtain the forest's final decision. In [2] specific inference instances to FRF are described to both strategies. These specific instances use the combination methods commented in previous sections to implement $Faggre$ function. To implement concrete instances of the strategies we need to define particular examples to the function K previously defined.

Given a reached leaf n in tree t, the particular case of $K(i, \chi_{t,n}(e), \omega_{t,n})$ that we use provides the weight $\chi_{t,n}(e)$ if i is the majority class in this node and 0 for all other classes. This particular case is denoted K1.

In this work we present new inference implementations to FRF based on local fusion concept. Also, we compare the results of these new implementations with the results obtained to the weighted implementations based on the OOB cases. Next we describe these implementations:

• **Weighted Implementations using the OOB cases-** Within this group we define the following implementations:

- *Majority Vote Weighted by Leaf and by Tree applied to strategy 1* (MWLT1): Uses a weight for each tree obtained by testing each individual tree with the OOB data file commented in the previous section. Lets be $\overline{p} = (p_1, p_2, ..., p_T)$ the vector with the weights assigned to each tree. Each p_i is obtained as:

$$\frac{N_success_OOB_i}{size_OOB_i}$$

  where $N_success_OOB_i$ is the number of examples inferred correctly from the OOB file used for testing the i th tree and $size_OOB_i$ is the total number of examples in this file.

  Once the vector $\overline{p}$ is obtained, we used it in the final decision of forest, so that the forest votes for the class c if:

$$c = \max_{i=1}^{I} \sum_{t=1}^{T} p_t \cdot A[t][i]$$

  where $A[t][i]$ is obtained using the majority vote combination method but applied to the values of reached leaves in each tree. Each reached leaf provides values to each class using the $K1$ function:

$$A[t][i] = \begin{cases} 1 & \text{if } i = \max_{j=1}^{I} \sum_{n=1}^{N_t} K1(j, \chi_{t,n}(e), \omega_{t,n}) \\ 0 & \text{other case} \end{cases}$$

- *Majority Vote Weighted by Leaf and by Tree to strategy 2* (MWLT2): Every reached leaf of the forest is considered to vote. In this case, given the example e, the forest votes for the class c if

$$c = \max_{i=1}^{I} \sum_{t=1}^{T} p_t \sum_{n=1}^{N_t} K1(i, \chi_{t,n}(e), \omega_{t,n})$$

• **Weighted Implementations based in local fusion-** Within this group we define the following implementations:

- *Local fusion applied to strategy 1* (FUS1):

  To apply this combination method, first, during the learning of the forest, by each tree generated we obtain an additional tree, that we call error tree.

  The procedure to construct the error tree associated to t-th tree in the forest is the following:

  With the training data set of the t-th tree in the forest (tds tree$_t$ in figure 3) we make a test of the tree. So, we consider the training data set as the test data set. With the results of this test we build a new dataset (tds error_tree$_t$ in the figure 3) with the same data but replacing the attribute class for the binary attribute error. The attribute error indicates whether that example has been classified correctly or not by the t-th tree in the forest (for example, that binary attribute can take the value 0 if the example has been correctly classified by the tree or 1 if it has been incorrectly classified and therefore it is a mistake made by the tree). With this new dataset is constructed a tree to learn the attribute error. This is the error tree.

  In the figure 3, tds tree$_t$ is the training dataset of the t-th tree in the forest and contains examples represented by vectors where:

  - $e_{j,t}$ is the j-th example in the training dataset of the t-th tree in the forest.
  - $class_{j,t}$ is the value of the attribute class to the j-th example. This attribute is the classification aim for the forest.

  tds error_tree$_t$ is the training dataset of the error tree associated to the t-th tree in the forest. Contains vectors represented as:

  - $e_{j,t}$ is the j-th example in the training dataset of the t-th tree in the forest.
  - $error_{j,t}$ is the attribute that acts as class in this dataset. It's a binary attribute with value:
    * 1 if $e_{j,t}$ is incorrectly classified by the t-th tree in the forest.
    * 0 if $e_{j,t}$ is correctly classified by the t-th tree in the forest.

ISBN: 978-989-95079-6-8

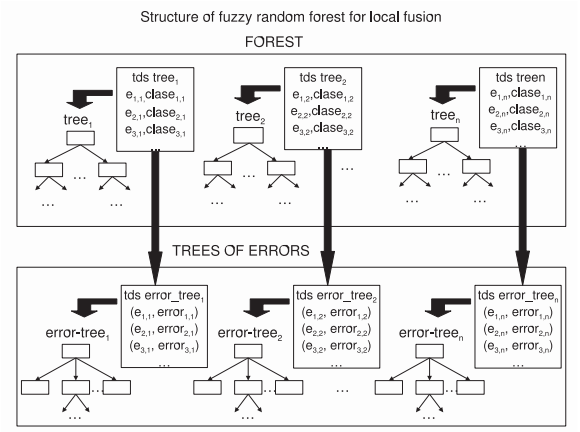


Figure 3: Inference with local fusion

Once the forest and the additional error trees are built, for each example e to classify and each tree of the forest we will obtain a vector $\overline{p}_e = (p_{e,1}, p_{e,2}, ..., p_{e,T})$ with the weights assigned to each tree to this example (local weights). Each $p_{e,t}$ is obtained as:

$$p_{e,t} = \sum_{n=1}^{N_{error_t}} \chi_{error_t,n}\omega_{error_t,n,0}, \quad \forall t = 1, .., T$$

where with the index $error_t$ we refer to the error tree of t-th tree in the forest, N_{error_t} is the number of leaves reached by the example e in the tree of errors $error_t$, $\chi_{error_t,n}$ is the weight of the leaf n of the tree of errors $error_t$ reached by the example and $\omega_{error_t,n,0}$ is the weight of the class 0 (value of binary attribute error=0) in the leaf n of the tree of errors.

The key idea that we want to capture with this implementation is that use of a local weight or a local fusion mechanism [3]. For a given probe (new example), first we evaluate the tree's performance with those examples (in the training set) which are similar to the probe we want to classify. Then, we associate a weight to the decision of that tree on the basis of its performance in the neighborhood of the probe.

Finally, the decision of the forest for the probe e is obtained by weighing the decisions of each tree with the computed weight. Therefore, the forest votes for the class c is:

$$c = \max_{i=1}^{I} \sum_{t=1}^{T} p_{e,t} \cdot A[t][i]$$

where $A[t][i]$ is obtained using the majority vote combination method but applied to the values of reached leaves in each tree. Each reached leaf provides values to each class using the $K1$ function:

$$A[t][i] = \begin{cases} 1 & \text{if } i = \max_{j=1}^{I} \sum_{n=1}^{N_t} K1(j, \chi_{t,n}(e), \omega_{t,n}) \\ 0 & \text{other case} \end{cases}$$

- *Local fusion applied to strategy 2* (FUS2):

Every reached leaf of the forest is considered to vote. In this case, given the example e, the forest votes for the class c if

$$c = \max_{i=1}^{I} \sum_{t=1}^{T} p_{e,t} \sum_{n=1}^{N_t} K1(i, \chi_{t,n}(e), \omega_{t,n})$$

5 Experiments and results

In this section we will make the FRF inference with different databases of the UCI repository and with different percentages of missing values and linguistic labels (to introduce the missing values and linguistic labels we have used the NIP 1.5 tool [5]). The percentage is divided equally between missing and labels values. In table 1 show the results of four inference implementations that we are considering in this work for each database without imperfection, with 5%, 15% and 30% of missing and linguistic labels. For each of them show the average and the standard deviation of 5 x 10 cross validation. We also show the results of the combined outputs of the trees by simple majority vote applied to strategy 1 (SM1) where each tree votes for the majority class of the reached leaves and the forest votes for the majority class among the trees and simple majority vote applied to strategy 2 (SM2) where each reached leaf in the several trees of the forest votes for its majority class and the forest votes for the majority class among the reached leaves.

Analyzing data from Table 1 we can say that the four weighted inference implementations have have similar behavior but the best inference implementation is FUS2, followed by MWLT2, FUS1 and MLWT1. It seems that the weighted inference implementations behave better with the strategy 2 that the strategy 1 and within each one of them behave better those who make use of the weighted with local fusion. We may also note that although generally not weighted strategies (SM1 and SM2) have very good performance with databases without imperfection, weighted strategies have improved the not weighted when we add imperfection in the databases.

6 Summary

In this paper we have defined implementations to combine the outputs of classifiers that composing the multi-classifier FRF. These implementations are based on the combination methods used frequently in the literature but weighting the decision of each classifier by local fusion. The results of these combination implementation is compared with other weighted implementation based on the OOB dataset that obtain a good performance in FRF and implementations based on simple majority vote.

We have presented experimental results obtained by applying these implementations to various databases. The results show that all weighted implementations obtain a similar behavior but FUS2 is the implementation that gets better results. Also the results show that the general strategy 2 provide better results with these implementations than the strategy 1. Furthermore, the weighted strategies improves the non weighted strategies when we add imperfection in the databases.

Acknowledgements

The authors thank the "Ministerio de Educación y Ciencia" of Spain and the European Fund of Regional Development

ISBN: 978-989-95079-6-8

Table 1: Testing accuracies of FRF for different levels of labels and missing

Dataset	Combination methods	original (0% missing and labels)	5% missing and labels	15% missing and labels	30% missing and labels
Ionosphere	FUS2	92.66 (2.15)	92.2 (2.16)	92.08 (1.96)	87.07 (3.29)
	FUS1	92.6 (2.11)	92.26 (2.17)	87.01 (3.36)	80.63 (2.96)
	MWLT2	92.66 (2.15)	92.2 (2.17)	91.51 (2.07)	85.82 (3.09)
	MWLT1	92.66 (2.15)	92.43 (2.17)	84.22 (3.01)	80.0 (2.77)
	SM2	93.0 (1.99)	92.14 (2.13)	85.93 (2.73)	79.83 (2.78)
	SM1	92.83 (2.05)	92.03 (2.15)	84.62 (2.82)	79.37 (2.81)
Wisconsin Breast Cancer	FUS2	96.37 (1.05)	95.94 (1.17)	96.02 (1.0)	94.79 (1.17)
	FUS1	96.34 (1.07)	95.62 (1.3)	95.54 (1.07)	94.22 (1.32)
	MWLT2	96.37 (1.05)	96.02 (1.15)	95.77 (1.04)	94.93 (1.11)
	MWLT1	96.34 (1.07)	95.65 (1.32)	95.34 (1.08)	94.05 (1.31)
	SM2	96.6 (1.03)	95.62 (1.32)	95.31 (1.07)	94.05 (1.31)
	SM1	96.6 (1.05)	95.62 (1.32)	95.31 (1.07)	94.05 (1.31)
Wine	FUS2	95.27 (2.35)	93.59 (2.62)	93.0 (2.75)	87.84 (3.85)
	FUS1	95.05 (2.49)	93.02 (2.81)	91.42 (3.14)	85.15 (4.56)
	MWLT2	95.27 (2.35)	93.47 (2.77)	92.78 (2.85)	87.61 (3.77)
	MWLT1	94.93 (2.57)	92.68 (2.91)	88.95 (3.27)	83.81 (4.43)
	SM2	95.27 (2.35)	92.01 (3.08)	88.39 (3.26)	83.58 (4.2)
	SM1	95.16 (2.45)	91.9 (3.12)	88.16 (3.36)	82.9 (4.5)
Iris Plants	FUS2	94.93 (3.13)	96.4 (2.5)	94.0 (3.2)	85.2 (5.13)
	FUS1	94.93 (3.13)	96.4 (2.5)	93.33 (3.2)	82.67 (5.67)
	MWLT2	94.93 (3.13)	96.4 (2.5)	94.0 (2.95)	85.33 (5.03)
	MWLT1	95.07 (3.16)	96.4 (2.5)	93.33 (3.2)	82.67 (5.71)
	SM2	94.93 (3.08)	96.4 (2.5)	93.33 (3.2)	82.4 (5.7)
	SM1	94.93 (3.08)	96.4 (2.5)	93.33 (3.2)	82.4 (5.7)
Pima Indian Diabetes	FUS2	76.59 (2.31)	73.36 (2.86)	72.11 (2.3)	72.42 (2.66)
	FUS1	76.51 (2.32)	73.02 (2.84)	71.72 (2.26)	72.61 (2.51)
	MWLT2	76.46 (2.34)	73.31 (2.9)	71.95 (2.31)	72.29 (2.83)
	MWLT1	76.66 (2.31)	72.89 (2.81)	71.8 (2.26)	72.16 (2.63)
	SM2	77.26 (2.23)	73.1 (2.58)	72.5 (2.38)	72.48 (2.76)
	SM1	77.0 (2.23)	73.54 (2.69)	72.4 (2.33)	72.5 (2.84)

for the support given to this work under the project TIN2008-06872-C04-03 and the "Fundación Séneca, Agencia de Ciencia y Tecnología de la Región de Murcia (Plan Regional de Ciencia y Tecnología 2007/2010)" which give support under the "Programa de Ayudas a Grupos de Excelencia de la Región de Murcia".

References

[1] H. Ahn, H. Moon, M.J. Fazzari, N. Lim, J.J. Chen and R.L. Kodell. Classification by ensembles from random partitions of high dimensional data. *Computational Statistics and Data Analysis*, 51:6166–6179, 2007.

[2] P.P. Bonissone, J.M. Cadenas, M.C. Garrido and R.A. Díaz. Combination methods in a fuzzy random forest. *Proceedings IEEE International Conference on Systems, Man and Cybernetics (SMC2008)* 1794–1799, 2008.

[3] P.P. Bonissone, F. Xue, and R. Subbur. Fast meta-models for local fusion of multiple predictive models. *Applied Soft Computing Journal*, doi:10.1016/j.asoc.2008.03.06, 2008.

[4] L. Breiman. Random forests. *Machine Learning* 45(1): 5–32, 2001.

[5] J.M. Cadenas, J.V. Carrillo, M.C. Garrido, R. Martínez and E. Muñoz. A tool to handle imperfect information in data sets *URL: http://heurimind.inf.um.es*

[6] M. Hamza and D. Larocque. An empirical comparison of ensemble methods based on classification trees. *Statistical Computati. & Simulation* 75(8):629-643, 2005.

[7] J. Jang. Structure determination in fuzzy modeling: A fuzzy CART approach. *Proceedings IEEE Conference on Fuzzy Systems*, 480–485, 1994.

[8] C.Z. Janikow. Fuzzy decision trees: issues and methods. *Transaction on Systems, Man and Cybernetics, Part B* 28(1): 1–15, 1998.

[9] L.I. Kuncheva. Fuzzy vs non-fuzzy in combining classifiers designed by boosting. *IEEE Trans. on Fuzzy Systems* 11(6): 729–741, 2003.

[10] S. Segrera, M. Moreno. Multiclassifiers: applications, methods and architectures. *Proc. of International Workshop on Practical Applications of Agents and Multiagents Systems, IWPAAMS05*, 263–271, 2005.

ISBN: 978-989-95079-6-8

Evolutionary learning of Quantified Fuzzy Rules for hierarchical grouping of laser sensor data in intelligent control

M. Mucientes, I. Rodríguez-Fdez, A. Bugarín

Intelligent Systems Group
Department of Electronics and Computer Science, University of Santiago de Compostela
Email: manuel.mucientes@usc.es, ismael.rodriguez1@rai.usc.es, alberto.bugarin.diz@usc.es

Abstract— *In complex systems it often occurs that relevant information about the system state and behavior is provided by groups of low-level variables rather than single variables. This grouping into high-level variables introduces a hierachy in the knowledge that can only be captured by means of rules involving propositions with a representation capability that is more complex than usual ones. In this paper we describe a genetic programming based approach for automated learning of Quantified Fuzzy Rules that are capable to deal with such representation capability. An application of this approach for hierarchical grouping of the distance measures provided by the laser sensors of a mobile robot (for the wall-following behaviour) is presented. Experimentation results show the control action is acceptable although no prior knowledge on the variables definition and structure was introduced in the controller.*

Keywords— Evolutionary algorithms, Genetic Programming, Fuzzy Quantification, Intelligent control.

1 Introduction

Modeling of systems using rules overcomes the interpretability issue and is a more adequate approach for scenarios when the model has to cope with noisy data or uncertainty. In this case fuzzy rules are a good choice.

Evolutionary algorithms have some characteristics that make them specially adequate for learning fuzzy rules, such as the flexibility in the representation of solutions through chromosomes, that may range from codifying a complete knowledge base, a single rule, some parameters of the fuzzy sets, ... The well-known combination of fuzzy logic with evolutionary algorithms (genetic fuzzy systems [1]) allows the designer to determine the most appropriate trade-off between accuracy and interpretability for a given system.

However, there are systems that can only be correctly modeled with rule bases that contain rules with different structures, and many times these structures are unknown. In order to learn such knowledge bases, the algorithm must have the ability to represent a set of structures. This ability is provided by genetic programming, where each chromosome of the population is represented as a tree of variable length. A complete flexibility in the structure of the rules is, generally, not desired and some restrictions have to be imposed, as not all the structures are valid. A compact representation of the valid structures of rules can be defined through a context-free grammar.

In systems with a huge number of (low level) input variables it is frequent that the values of the individual variables are not meaningful. In these cases relevant information about the system is only obtained by analyzing sets of low level variables, grouped into high level variables. This provides a hierarchical structure of knowledge that is meaningful for the experts, where the meaning of high level variables cannot be extracted taking neither an average value of the low level variables of the set, nor the maximum or minimum values. Usually this meaning can be directly captured using Quantified Fuzzy Propositions (QFPs) like *"most of the variables in the set take a high value"*.

In this paper we propose a genetic based approach to learn Quantified Fuzzy Rules (QFRs), defined as fuzzy rules of variable structure involving QFPs. The algorithm is based on the genetic programming approach and has been designed to solve regression problems in which the number of input variables is really high, and the model of the system needs the grouping of these variables in several sets (high level variables). In order to illustrate the utility of the algorithm, an application of use for hierarchical grouping of the distance measures provided by the laser sensors in a mobile robot is presented.

The paper is structured as follows: Sec. 2 introduces the motivation of the use of Quantified Fuzzy Rules (QFRs) with variable structure. Sec. 3 presents the QFRs model, while Sec. 4 describes the genetic programming algorithm that has been used to learn the QFRs. Finally, Sec. 5 points out some results, conclusions and future improvements of the algorithm.

2 Motivation for QFRs

Modeling the behavior of a mobile robot that performs a task, autonomously or remotely controlled by a human operator, is a challenging problem if no prior knowledge is provided. In previous papers [2] two successful approaches to learn fuzzy rules for the control of a robot in two different tasks were described. In both cases, the learned rules were conventional fuzzy rules and all the input variables were defined by a human expert. For the wall following task these variables were *right distance*, *left distance*, *orientation*, and *velocity*. For the moving object tracking task the variables were *distance to the object*, *deviation*, *difference in velocity with the object* and *difference in angle*.

The question that comes up is: is it possible to model the behavior of the robot without using any prior knowledge? This means that learning relies only on the input-output data pairs, i.e. on the sensors information and on the control orders that have been given for those situations. Fig. 1 shows how a robot equipped with a laser range finder senses the environment.

Laser range finders emit at the same time beams in different directions. When a beam hits an obstacle, it is reflected and registered by the scanner's receiver. With this information, the distance measured in the direction of each beam can be calculated. Fig. 1 shows in the shadowed area the space covered by the laser when the robot is placed in an typical office environment. The laser range finder provides the distances to the

ISBN: 978-989-95079-6-8

closest obstacle in each direction with a given angular resolution (number of degrees between two consecutive beams).

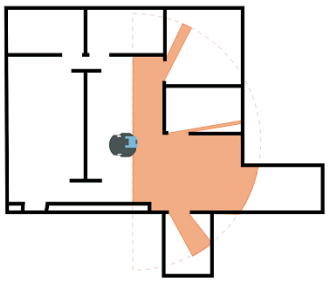


Figure 1: Sensors information for a mobile robot equipped with a laser range finder.

A robot equipped with a single laser range finder provides 361 distance measurements in each acquisition for an angular resolution of 0.5 degrees. When equipped with two laser range finders (to get information of the whole surrounding environment) it provides 722 distance input variables. As the output variables the controller must provide the linear and angular velocities. This control action cannot be decided by simply analyzing the individual distance values of each beam, since noisy measurements, gaps between objects (very frequent in cluttered environments) may occur. Usually the relevant information of the environment is modeled (or extracted) with propositions involving a more complex structure in the variables. For instance, the input variables can be arranged taking into account the importance of the information they provide to model the system.

In general, two categories can be established:

- High-level input variables: variables that provide, by themselves, information relevant and meaningful to the expert for modelling the system (e.g. the linear velocity of the robot).
- Low-level input variables: variables for which their individual values do not contribute to model the system. Their importance stems from the analysis of sets of these variables (e.g. the distance measured by a single beam in a laser range finder).

However, the values of a group of low level variables can provide valuable and meaningful information. For example, the "'frontal sector"' of a laser range finder is a high-level variable made up of a set of distances of single beams (low-level variables). Within this context Quantified Fuzzy Propositions (QFPs) such as *"some of the distances of the frontal sector are low"* are useful for representing relevant knowledge for the experts and therefore for performing intelligent control. QFPs require the definition of several elements:

- *Some*: how many distances of the frontal sector must be low?
- *frontal sector*: which beams belong to the frontal sector?
- *low*: what is the meaning of low?

This example clearly sets out the need to use propositions different from the conventional ones. In this paper QFPs (as "X is A in Q of S") are used for representing knowledge about high-level variables that are defined as the grouping of low-level variables. Conventional ("X is A") are used for other high-level variables, non related to low-level ones.

3 Quantified Fuzzy Rules (QFRs) model

An example of QFR is shown in Fig. 2, and involves both QFPs 1 and conventional ones 2 :

IF $d(h)$ *is* $HIGH$ *in most of* F^1_{sector} and (1)
...
velocity is F_{vel} (2)
THEN *linear velocity is* F_{lv} and *angular velocity is* F_{av}

Figure 2: A typical QFR to model the behavior of a mobile robot.

The general expression for QTPs in our case is:

$$d(h) \text{ is } F^i_d \text{ in } Q^i \text{ of } F^i_{sector} \quad (3)$$

where, for each i=1,...,N_d (N_d being the number of analyzed sectors of distances:

- $d(h)$ is the signal. In this particular case, it represents the distance measured by beam h.
- F^i_d is a linguistic value for variable $d(h)$.
- Q^i is a (spatial, defined in the laser beam domain) fuzzy quantifier.
- F^i_{sector} is a fuzzy set in the laser beam domain (e.g., the "frontal sector").

Evaluation of the Degree of Fulfillment (DOF) for 3 is carried out using Zadeh's quantification model for proportional quantifiers (such as "most of", "part of", ...) [3], that allows to consider non-persistence, partial persistence and total persistence situations for event $d(h)$ is F^i_d in the range of laser beams (spatial interval F^i_{sector}). This is a relevant characteristic of this model, since it allows to consider partial, single or total fulfillment of an event within the laser beams set.

Automatic learning of QFRs for this application (Fig. 2) demands an algorithm with the ability to represent rules with different structures, as the number N_d of analyzed sectors of distances, and therefore the number of QFPs per rule (Fig. 2) can change among rules. Moreover, the number of parameters involved in the definition of a proposition is higher in QFP than in conventional propositions (3 vs. 1).

4 Genetic programming algorithm

The proposed evolutionary algorithm is a genetic programming algorithm based on the genetic cooperative-competitive learning (GCCL) approach. In genetic programming, an individual is a tree of variable length. Each individual in the population can have a different structure, and the introduction of restrictions in that structure of the chromosome can be solved using, for example, a context-free grammar. As the number of beams sectors is unknown and can change among rules, the

ISBN: 978-989-95079-6-8

flexibility in the structure of the rules can be managed by the genetic programming algorithm.

In the GCCL approach [4, 5] rules evolve together but competing among them to obtain the highest fitness. In this approach it is fundamental to include a mechanism to maintain the diversity of the population (niche induction). The mechanism must warrant that there is competition among individuals of the same niche, but also has to avoid the deletion of those weak individuals that occupy a niche not covered by other individuals of the population.

We have chosen token competition [6] as the mechanism for maintaining the diversity. According to [7], this mechanism is adequate for genetic programming, as in this kind of evolutionary algorithms the structure of the individuals can be completely different and, thus, the evaluation of the similarities is hard. The advantage of token competition over other approaches, like crowding or fitness sharing, is that it is not necessary to estimate the similarities between pairs of individuals.

The learning process is based on a set of training examples. Each example e^l is represented by a tuple:

$$e^l = (d(1), \ldots, d(N_{LB}), \mathit{velocity}, \mathit{vlin}, \mathit{vang}) \quad (4)$$

where $d(h)$ is the distance measured by beam h, *velocity* is the velocity of the robot, and *vlin* and *vang* are the output variables linear and angular velocity. In token competition, each example of the training set has a token and, of all the individuals that cover this example, the token will be seized by the individual with the highest raw fitness. In this way, the individual with the highest strength in the niche will exploit it, while individuals that are weaker will reduce its strength as they cannot compete with the best individual in the niche.

4.1 Description of the context-free grammar

As we pointed out before, in genetic programming each individual is a tree of variable size. Thus, the structure of the individuals can be very different among them. In order to generate valid individuals of the population, and to produce right structures for the individuals after crossover and mutation, some restrictions have to be applied. With a context-free grammar all the valid structures of a tree (chromosome) in the population can be defined in a compact form. A context-free grammar is a quadruple (V, Σ, P, S), where V is a finite set of variables, Σ is a finite set of terminal symbols, P is a finite set of rules or productions, and S is an element of V called the start variable.

The grammar is described in Fig. 3. The first item enumerates the variables, then the terminal symbols, in third place the start variable is defined, and finally the rules for each variable are enumerated. When a variable has more than one rule, rules are separated by symbol |. Fig. 4 represents a typical chromosome generated with this context-free grammar. Terminal symbols (leafs of the tree) are represented by circles, and variables are shown as flatted circles. Terminal symbols F_{vel}, F_d, F_{lv}, F_{av}, Q correspond with the linguistic labels and the quantifier defined in Fig. 2. Moreover, F_{sector} has been codified with *initialB* and *finalB*, which represent the extreme values of a trapezium.

A description of the evolutionary algorithm is shown in Fig. 5. It is based on the GCCL approach with a population of variable size. The following sections describe each of the stages of the algorithm.

- V = { rule, antecedent, consequent, distances }
- Σ = { F_{lv}, F_{av}, F_{vel}, F_d, *initialB*, *finalB*, Q }
- S = rule
- Productions:
  - rule $\longrightarrow$ antecedent consequent
  - antecedent $\longrightarrow$ distances F_{vel} | distances
  - consequent $\longrightarrow$ F_{lv} F_{av}
  - distances $\longrightarrow$ F_d *initialB* *finalB* Q distances | F_d *initialB* *finalB* Q

Figure 3: Context-free grammar.

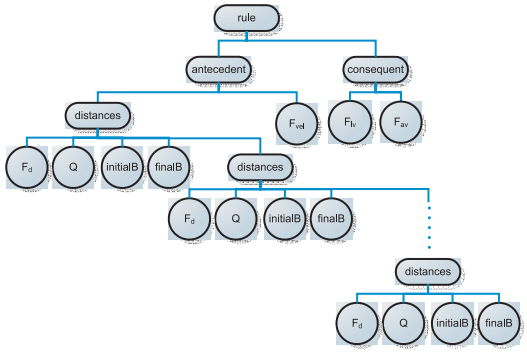


Figure 4: A chromosome representing a QFR to model the behavior of a robot.

1. Initialization
2. for iteration = 1 to *maxIterations*
   (a) Selection
   (b) Crossover and mutation
   (c) Evaluation
   (d) Resize population

Figure 5: Evolutionary algorithm.

4.2 Initialization

The first step of the algorithm is the initialization of the population: a chromosome (Fig. 4) is generated for each of the first $pop_{maxSize}$ examples in the training set. All the fuzzy sets of conventional fuzzy propositions (F_{vel}) and those of the consequent part (F_{lv} and F_{av}) are initialized with a triangular membership function centered in the corresponding example value, and with the extremes of the triangle at a distance equal to $prec_j$ from the center of the triangle. $prec_j$ represents the lowest meaningful change in the value of variable j.

The initialization of F_d, *initialB*, *finalB* and Q is quite more complex. Firstly, the number of QFPs that should be included in each rule is unknown and may be different for each example. Therefore, starting with the first beam, the average value

ISBN: 978-989-95079-6-8

of the distance is calculated incorporating consecutive beams until the variance of the distances is over a threshold. The process is repeated until all the beams have been asigned to a sector. Those sectors with a low number of beams are discarded. For each of the other sectors, *initialB* and *finalB* are directly defined using the information of the sector. F_d is generated in the same way as the fuzzy sets of the conventional propositions, but the triangular membership function is centered in the average value of the distances of the beams of the sector. Finally, Q is calculated as the percentage of beams of the sector that fulfill F_d.

4.3 Selection

The first stage of the iterative part of the algorithm is the selection of the individuals of the population. We have used a binary tournament selection. In a k-tournament selection, k individuals are randomly picked from the population with replacement, and the best of them is selected. In this case, $k = 2$ (binary tournament selection).

4.4 Evaluation

In order to estimate the fitness of an individual of the population, we first need to know the probability that an example e^i matches the output C_j:

$$P\left(C_j|e^i\right) = \exp\left(-\frac{error_j^i}{ME}\right) \tag{5}$$

where $error_j^i$ is the error due to the selection of an output different from the one codified in the example, and *ME* is a parameter that defines the meaningful error for the application. $error_j^i$ can be defined as:

$$error_j^i = \left(y_i^i - y_j^i\right)^2 \tag{6}$$

where y_i^i is the state of the system reached when the output of example e^i is applied to the state defined by e^i, and y_j^i is the state of the system reached when output C_j is applied to the state defined by e^i. As we are working with regression problems, an example can have several outputs that are different from the one codified in the example, but that make small errors, i.e., that are very similar to the desired output.

The accuracy of an individual of the population can be described as:

$$confidence = \frac{GC_{ex}}{covered_{ex}} \tag{7}$$

where GC_{ex} is the number of covered examples with $P\left(C_j|e^i\right)$ over a threshold P_{min}. The ability of generalization of a rule is calculated as:

$$support = \frac{GC_{ex}}{GC_{ex} + GU_{ex}} \tag{8}$$

where GU_{ex} is the number of uncovered examples with $P\left(C_j|e^i\right) > P_{min}$. Finally, we can define $fitness_{raw}$ as the combination of both values:

$$fitness_{raw} = \alpha_f \cdot confidence + (1 - \alpha_f) \cdot support \tag{9}$$

which represents the strength of an individual without taking into account the others. $\alpha_f \in [0,\ 1]$ is a parameter that codifies the trade-off between accuracy and generalization.

As we are using the GCCL approach, $fitness_{raw}$ will be used to decide which individual seizes each example. Thus, the fitness of an individual is defined as:

$$fitness = fitness_{raw} \cdot \frac{seized_{ex}}{covered_{ex}} \tag{10}$$

where $seized_{ex}$ is the number of examples seized by the individual, while $covered_{ex}$ is the number of examples that have been covered by it.

4.5 Crossover and mutation

Once pop_{size} individuals have been selected, each couple of them is crossed with probability p_c. The crossover operator is the parent-centric BLX (PCBLX) [8, 9]. This crossover operator is valid for genes representing real numbers. However, in the defined chromosomes (Fig. 4), there are also genes representing trapezoids. Given two trapezoids represented by tuples, $(\kappa_1\ \ldots\ \kappa_4)$ and $(\rho_1\ \ldots\ \rho_4)$ ($\kappa_j, \rho_j \in [a_j, b_j]$, $j = 1, \ldots, 4$), that are going to be crossed, the following offspring are generated:

- $(\kappa'_1\ \ldots\ \kappa'_4)$, where κ'_j is randomly selected from the interval $\left[l_j^\kappa, r_j^\kappa\right]$, with $l_j^\kappa = \max\left\{a'_j, \kappa_j - I_j\right\}$, $r_j^\kappa = \min\left\{b'_j, \kappa_j + I_j\right\}$, $I_j = |\kappa_j - \rho_j| \cdot \alpha$, $\alpha \in [0, 1]$, $a'_j = a_j$ if $j \in \{1,\ 2\}$ or $a'_j = \kappa_2$ if $j \in \{3,\ 4\}$, and $b'_j = \kappa_3$ if $j \in \{1,\ 2\}$ or $b'_j = b_j$ if $j \in \{3,\ 4\}$.
- $(\rho'_1\ \ldots\ \rho'_4)$, where ρ'_j is randomly selected from the interval $\left[l_j^\rho, r_j^\rho\right]$, with $l_j^\rho = \max\left\{a'_j, y_j - I_j\right\}$ and $r_j^\rho = \min\left\{b'_j, y_j + I_j\right\}$. Finally, $a'_j = a_j$ if $j \in \{1,\ 2\}$ or $a'_j = \rho_2$ if $j \in \{3,\ 4\}$, and $b'_j = \rho_3$ if $j \in \{1,\ 2\}$ or $b'_j = b_j$ if $j \in \{3,\ 4\}$.

When two individuals are crossed, the PCBLX operator is applied to all the genes of type terminal symbol in the chromosomes. However, as the chromosomes are trees of variable structure and/or size, for a gene of the first individual, which gene should be selected in the second of the individuals? For gene F_{vel} the choice is simple: if this gene exists in both individuals crossover is performed, otherwise not. Nevertheless, for genes of type *distances* (as there are many possible choices), the gene selected in the second individual is the one that has a higher overlap with the gene of the first individual in the definition of the sector. If there is not overlap, the gene is not crossed. For genes of type *distances*, PCBLX is applied to F_d, *initialB*, *finalB*, and Q with their corresponding counterparts in the other chromosome.

After crossover, it could happen that the nodes of type *distances* are unsorted, or that two nodes overlap (they should be merged). The algorithm to repair the chromosome has the following steps:

1. Sort all the sectors taking into account *initialB* and insert them on a list.
2. While the list is not empty.
   (a) Choose the first element of the list and check if it overlaps with other sectors.
      i. If there is overlap, then merge the sectors and sort the list.

ISBN: 978-989-95079-6-8

ii. If there is not overlap, delete the sector from the list and add it to the new tree.

When crossover is not performed, both individuals are mutated. Mutation can be implemented to generate a more general or a more specific rule. The higher the value of $seized_{ex}/covered_{ex}$, the higher the probability to generalize the rule by mutation. This occurs with rules that are very strong in their niche and that could be modified to seize examples in other niches. On the contrary, when the number of seized examples in very low, this means that the rule is weak in its niche, and in order to improve its performance some examples that are currently covered should be discarded.

In order to select the examples that are going to be covered (generalization) or uncovered (specialization), the value of $P\left(C_j|e^i\right)$ (Eq. 5) for each example is considered, where C_j is the consequent of the rule. For generalization, the best C_{ex} uncovered examples are selected while, for specialization, the worst C_{ex} covered ones are chosen. Then, from the selected set of examples, S_{ex} are randomly selected to modify the rule. The selection probability of each example is directly (generalization) or inversely (specialization) proportional to $P\left(C_j|e^i\right)$. For each selected example, each proposition of the antecedent part of the rule is analyzed to decide its mutation. For generalization, if the DOF of the proposition is null, then the proposition is mutated. The opposite occurs for specialization. The mutation of a proposition for generalization can be done with the following mutation operators:

1. Delete the proposition.
2. Merge two adjacent distance sectors. The new F_d value will be obtained applying PCBLX operator to the F_d values of the old sectors. This mutation operator can only be applied for propositions of type distances.
3. Modify the fuzzy label to cover the new example.
4. Modify the quantifier to cover the new example (only for distance propositions).

On the other hand, the mutation operators for specialization are:

1. Insert a *distances* sector (only valid for distance propositions). The insertion position is randomly selected. The values for F_d and Q are generated mutating the values of the previous or next sector. Also, *initialB* and *finalB* are randomly selected using the definitions of the contiguous sectors.
2. Divide a distances sector (only valid for distance propositions). One of the new sectors will have the same F_d and Q values, while the other sector will be generated applying non-uniform mutation to the original values.
3. Modify the fuzzy label to uncover the selected example.
4. Modify the quantifier to uncover the selected example (only for distance propositions).

4.6 Resize population

Those individuals with null fitness are removed from the population, and the best $pop_{maxSize}$ individuals are selected for the final population. If $pop_{size} < pop_{maxSize}$, and there are still uncovered examples, then new individuals are added. These individuals are chosen from the *examples population*, randomly selecting those rules that cover examples that have not been seized yet by the individuals of the population.

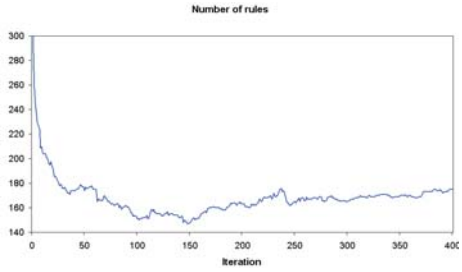
4.7 Population evaluation

The performance of the knowledge base obtained after each iteration has been measured as:

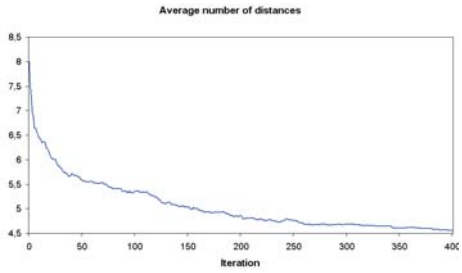
$$fitness_{pop} = \frac{n_{ex}}{\sum_l error_{pop^l}} \tag{11}$$

where n_{ex} is the total number of examples, and $error_{pop^l}$ is the error of the knowledge base when the input is example l.

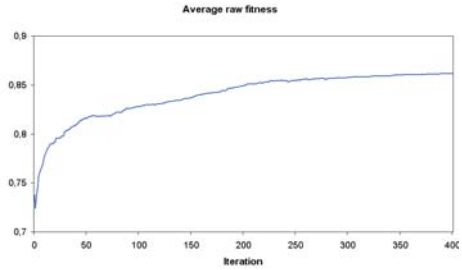
5 Results and discussion



(a) Number of rules.



(b) Average number of distances.



(c) Average $fitness_{raw}$.

Figure 6: Evolution of several variables for an execution of the algorithm.

The learning algorithm has been tested with a set of 795 examples recorded when the robot was following the wall in an environment. The values that have been used for the parameters of the evolutionary algorithm are: $maxIterations = 400$,

ISBN: 978-989-95079-6-8

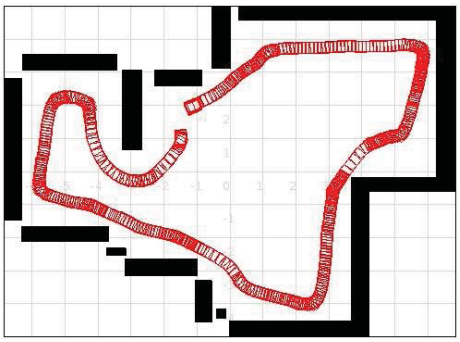


Figure 7: Path of the robot along the test environment.

$pop_{maxSize} = 300$, $p_c = 0.5$, $\alpha = 0.5$ (PCBLX crossover), and $P_{min} = 0.4$.

Fig. 6 shows the evolution of several variables for an execution of the genetic programming algorithm. These variables are the number of rules in the population after each iteration, the average (over all the individuals of the population) number of distance variables in the antecedent part of the rules, and the average $fitness_{raw}$. As can be seen from Fig. 6(a), until iteration 150 the number of rules decreases to less than 150. As the rules in the initial population where very specific (covering sometimes only one example) the algorithm finds rules with a good value of the confidence and with a higher value of the support. In the second stage, as there are rules with a high support, new rules with a higher confidence (and a lower support) seize part of the examples of the weaker rules, thus generating a higher number of rules. At the last iteration of the algorithm, the population had 175 rules. The improvement in the quality of the rules of the population is reflected by the evolution of the average raw fitness (eq. 9) of the rules (Fig. 6(c)).

Finally, the evolution in the average number of distances in the antecedent part of the rules (Fig. 6(b)) reflects the ability of the algorithm to extract valuable information about the distance sectors that are meaningful to model the system. From an initial value of more than eight distance sectors (in average), the algorithm obtains rules that, in average, use around 4.5 distance variables.

Fig. 7 shows the path of the robot along the environment that was used to test the learned controller. In this environment, the robot was following the right-hand wall. The higher the concentration of marks of the robot, the lower the linear velocity. In order to evaluate the quality of the controller we have measured four different indicators: the right distance, the linear velocity, the change in the linear velocity between two consecutive cycles —which reflects the smoothness in the control—, and the time. The average values of the indicators are calculated for each lap that the robot performs in the environment. Results presented in table 1 are the average and standard deviation values over five laps of the average values of the indicators over one lap.

Results show a distance over the expected one (0.5). However, this is in part because the distance has been measured with the information coming from only one beam (in [2, 10] the distance was measured using the input variable defined by an expert). Moreover, the velocity is low. Nevertheless, the smoothness of the control is very good.

Table 1: Results ($x \pm \sigma$) for the environment in Fig. 7

Distance (m)	0.8137 ± 0.0091
Velocity (m/s)	0.13274 ± 0.00095
Velocity change (m/s)	0.00486 ± 0.00053
Time (s)	259.26 ± 2.20

These preliminary results are promising and show an acceptable control, although no expert knowledge was introduced for the definition of the variables. Moreover, as the system has been trained with data of a unique environment, further analysis on the ability of the algorithm to generalize the extracted model of the behavior of the robot will be conducted in the future.

Acknowledgment

This work was supported by the Spanish Ministry of Science and Innovation under grant TIN2008-00040. M. Mucientes is supported by the *Ramón y Cajal* program of the Spanish Ministry of Science and Innovation.

References

[1] 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.

[2] M. Mucientes and J. Casillas. Quick design of behaviors in mobile robotics by fuzzy controllers with good interpretability. *IEEE Transactions on Fuzzy Systems*, 15(4):636–651, 2007.

[3] L.A. Zadeh. A computational approach to fuzzy quantifiers in natural languages. *Computers and Mathematics with Applications*, 9:149–184, 1983.

[4] A. Giornada and F. Neri. Search-intensive concept induction. *Evolutionary Computation*, 3(4):375–416, 1995.

[5] D.P. Greene and S.F. Smith. Competition-based induction of decision models from examples. *Machine Learning*, 3:229–257, 1993.

[6] K.S. Leung, Y. Leung, L. So, and K.F. Yam. Rule learning in expert systems using genetic algorithm: 1, concepts. In *Proceedings of the 2nd International Conference on Fuzzy Logic and Neural Networks*, pages 201–204, Iizuka (Japan), 1992.

[7] F.J. Berlanga, M.J. del Jesus, and F. Herrera. Learning compact fuzzy rule-based classification systems with genetic programming. In *Proceedings of the 4th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT)*, pages 1027–1032, Barcelona (Spain), 2005.

[8] F. Herrera, M. Lozano, and A.M. Sánchez. A taxonomy for the crossover operator for real-coded genetic algorithms: An experimental study. *International Journal of Intelligent Systems*, 18:309–338, 2003.

[9] M. Lozano, F. Herrera, N. Krasnogor, and D. Molina. Real-coded memetic algorithms with crossover hill-climbing. *Evolutionary Computation*, 12:273–302, 2004.

[10] M. Mucientes, R. Alcalá, J. Alcalá-Fdez, and J. Casillas. Learning weighted linguistic rules to control an autonomous robot. *International Journal of Intelligent Systems*, 24:226–251, 2009.

ISBN: 978-989-95079-6-8

Graded Equipollence of Fuzzy Sets

Michal Holčapek[1], Matěj Turčan[2]

1. Institute for Research and Applications of Fuzzy Modeling, University of Ostrava,
30. dubna 22, 701 03 Ostrava 1, Czech Republic
2. Faculty of Economics, VŠB-Technical University of Ostrava
Sokolská 33, 701 21 Ostrava 1, Czech Republic
Email: michal.holcapek@osu.cz, matej.turcan@vsb.cz

Abstract— *The aim of this contribution is to propose an approach to comparing fuzzy sets from the cardinality point of view. The idea is based on a transformation of one fuzzy set to another one in some degree, where one-to-one mappings between suitable sets are used. To ensure the correctness of degrees in which there exist correspondences between fuzzy sets we introduce a class of sets and consider fuzzy sets and fuzzy classes with universes of discourse from this universe. We introduce a degree in which two fuzzy sets are equipollent and show that these degrees define a similarity relation on the class of all fuzzy sets. Some simplifications of the original definition of the concept of "being equipollent" are proposed and several properties are demonstrated.*

Keywords— fuzzy sets, equipollence of fuzzy sets, cardinal theory of fuzzy sets.

1 Introduction

The classical concept of cardinality of a set is a measure of the "number of elements of the set". A formal definition without knowing anything about numbers is based on the notion of one-to-one correspondence between sets. In particular, we say that two sets x and y are *equipollent* (*equipotent, equivalent, bijective*, or have *the same cardinality* etc.) and write $|x| = |y|$, if there exists a one-to-one mapping of x onto y. This definition gives rise to equivalence relation defined on the class of all sets. This equivalence is called *equipollence* (or also *equipotence, equinumerosity* etc.) and using it the concept of cardinality is introduced. For instance, if we work with the Zermelo-Fraenkel axiomatic set theory with the axiom of choice, the cardinality (or the cardinal) of a set x is defined as the least ordinal number in the class of all sets equipollent with x.

The concept of equipollent fuzzy sets is not new in fuzzy set theory. Some classical-like approaches to the concept of equipollence of fuzzy sets are presented by D. Klaua [1] or S. Gottwald [2, 3]. In the second case, a degree in which two fuzzy sets are equipollent is defined using fuzzy uniqueness of fuzzy mappings. Equivalence classes (fuzzy sets) are *fuzzy cardinals*. Note that fuzzy mappings are here defined as special fuzzy relations.

Another approach was suggested by M. Wygralak [4], where the concept of equipollence of fuzzy sets is introduced using the cardinality of α-cuts of fuzzy sets. More precisely, two fuzzy sets A and B are equipollent, if

$$\bigvee\{\alpha \in [0,1] \mid |A_\alpha| \geq i\} = \bigvee\{\alpha \in [0,1] \mid |B_\alpha| \geq i\} \quad (1)$$

for any cardinal number i. Some modifications of (1) can be found in [5, 6, 7].

In this contribution, we propose a new approach to the equipollence of fuzzy set over a universe of sets (e.g., a universe of all finite or countable sets, or all sets, or a Grothendieck universe). Analogously to the Gottwald's approach, a graded equipollence is considered where, however, the degrees of being equipollent are obtained by a simpler way than in the mentioned approach. More precisely, our approach is based on graded one-to-one correspondences between fuzzy sets where the crisp mappings between universe are considered. Such equipollence, contrary to the Gottwald's definition, can be used in practice, e.g., for a comparison of a number of well satisfied criteria for different applicants. Note that our approach is, in a certain sense, equivalent to the Wygralak's approach, if we restrict ourselves to finite fuzzy sets.

2 Preliminaries

2.1 Structures of membership degrees of fuzzy sets

In this contribution the membership degrees of fuzzy sets will be interpreted in a complete residuated lattice.

Definition 2.1. An algebra $\mathbf{L} = \langle L, \wedge, \vee, \otimes, \rightarrow, \bot, \top\rangle$ is a *residuated lattice*, if $\langle L, \wedge, \vee, \bot, \top\rangle$ is a bounded lattice, where $\bot$ and $\top$ are the least and greatest elements of L, respectively, $\langle L, \otimes, \top\rangle$ is a commutative monoid and

$$\alpha \leq \beta \rightarrow \gamma \quad \text{if and only if} \quad \alpha \otimes \beta \leq \gamma, \qquad (2)$$

holds for arbitrary $\alpha, \beta, \gamma \in L$ ($\leq$ denotes the corresponding lattice ordering).

The operations $\otimes$ and $\rightarrow$ are called a *multiplication* and *residuum*, respectively. A residuated lattice is *complete* or *linearly ordered*, if $\langle L, \wedge, \vee, \bot, \top\rangle$ is a complete or linearly ordered lattice, respectively.

Example 2.1. Let T be a left continuous t-norm (see [8]). Then

$$\mathbf{L} = \langle [0,1], \min, \max, T, \rightarrow_T, 0, 1\rangle,$$

where $\alpha \rightarrow_T \beta = \bigvee\{\gamma \in [0,1] \mid T(\alpha, \gamma) \leq \beta\}$, is a complete linearly ordered residuated lattice. One of the most important residuated lattices on the unit interval is *Łukasiewicz algebra* that is determined by $T_{\mathrm{L}}(a,b) = \max(0, a + b - 1)$. The residuum is then given by $a \rightarrow_{T_{\mathrm{L}}} b = \min(1 - a + b, 1)$.

ISBN: 978-989-95079-6-8

Example 2.2. Let us consider the lattice $L = \{\bot, a, b, \top\}$ with two incomparable elements a and b. Put $x \otimes y = x \wedge y$ and $x \to y = \bigvee\{z \mid x \wedge z \leq y\}$ (see Table 1). Then

$$\mathbf{L} = \langle\{\bot, a, b, \top\}, \wedge, \vee, \to, \bot, \top\rangle$$

is a non-linearly ordered residuated lattice. Note that this is an example of a lattice called a *Heyting algebra*.

The *biresiduum* in $\mathbf{L}$ is a binary operation $\leftrightarrow$ on L defined by

$$\alpha \leftrightarrow \beta = (\alpha \to \beta) \wedge (\beta \to \alpha). \tag{3}$$

Example 2.3. One checks easily that the biresiduum for the Łukasiewicz algebra is given by $a \leftrightarrow_{\mathrm{L}} b = 1 - |a - b|$. The definition of biresiduum for the residuated lattice from Example 2.2 is given in Table 1.

$\to$	$\bot$	a	b	$\top$
$\bot$	$\top$	$\top$	$\top$	$\top$
a	b	$\top$	b	$\top$
b	a	a	$\top$	$\top$
$\top$	$\bot$	a	b	$\top$

$\leftrightarrow$	$\bot$	a	b	$\top$
$\bot$	$\top$	b	a	$\bot$
a	b	$\top$	$\bot$	a
b	a	$\bot$	$\top$	b
$\top$	$\bot$	a	b	$\top$

Table 1: Definitions of $\to$ and $\leftrightarrow$ in Example 2.2 and 2.3

2.2 Several notions from the set theory

In this contribution we work with Zermelo-Fraenkel axiomatic set theory with the axiom of choice (ZFC). We shall use small latin letters as $x, y, z, \ldots$ to denote common sets and capital latin letters as $A, B, R, \ldots$ to denote special sets (e.g., special mappings and relations). Although, the objects of ZFC are sets, we consider the informal notion of a class. This is useful for description of collections of sets that are defined by formulas of the language of set theory, i.e., if $\varphi(x, p_1, \ldots, p_n)$ is a formula, we call

$$\{x \mid \varphi(x, p_1, \ldots, p_n)\} \tag{4}$$

a *class*. We shall use calligraphic capital letters as $\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \ldots$ to denote classes. Every set can be consider as a class, i.e., if y is a set, then it is sufficient to consider the class $\{x \mid x \in y\}$. A class that is not a set is a *proper class*.

We shall use $\emptyset$ to denote the empty set and $\mathrm{P}(x)$ to denote the power set. Further, we shall use the standard symbols $\cap, \cup, \sqcup, \times, \setminus$ etc. to denote the intersection, union, disjoint union, product, difference of sets etc. (and analogously for classes), respectively. The common symbols $\subseteq$ and $\subset$ denote the inclusion relations of sets (and analogously for classes). If $R \subseteq x \times y$ (and analogously for $\mathcal{R} \subseteq \mathcal{X} \times \mathcal{Y}$) is a relation, then we shall use

$$\mathrm{Dom}(R) = \{t \mid t \in x \,\&\, \exists r \in y : (t, r) \in R\},$$
$$\mathrm{Ran}(R) = \{r \mid r \in y \,\&\, \exists t \in x : (t, r) \in R\}$$

to denote the domain and range of R, respectively. If f is a mapping between sets (analogously for classes), then $\mathrm{Dom}(f)$ and $\mathrm{Ran}(f)$ denote the domain and range of f, respectively. We shall use $|x|$ to denote the cardinality of a set x.

3 Fuzzy sets and classes in universes of sets

Let us suppose that a complete residuated lattice $\mathbf{L}$ is given. The following concept of a universe of sets is motivated by the definition of Grothendieck universe (see e.g. [9]).

Definition 3.1. *A universe of sets over* $\mathbf{L}$ is a class $\mathcal{U}$ of sets having the following properties:

(U1) $x \in y$ and $y \in \mathcal{U}$, then $x \in \mathcal{U}$,

(U2) $x, y \in \mathcal{U}$, then $\{x, y\} \in \mathcal{U}$,

(U3) $x \in \mathcal{U}$, then $\mathrm{P}(x) \in \mathcal{U}$,

(U4) $x \in \mathcal{U}$ and $y_i \in \mathcal{U}$ for any $i \in x$, then $\bigcup_{i \in x} y_i \in \mathcal{U}$,

(U5) $x \in \mathcal{U}$ and $f : x \to L$, then $\mathrm{Ran}(f) \in \mathcal{U}$,

where L is the support of the complete residuated lattice $\mathbf{L}$.

Simple examples of such universes are the classes of all, finite, or countable sets. Under certain conditions (see e.g. [9]), a Grothendieck universe may be also a universes of sets over $\mathbf{L}$. One could notice that L need not be included in $\mathcal{U}$. In fact, it is sufficient to consider the class of all finite set and an infinite complete residuated lattice $\mathbf{L}$. Notice that if $\mathcal{U}$ is a proper class, then $x \subseteq \mathcal{U}$ is not necessary a set. The following theorem shows some constructions inside the universe $\mathcal{U}$.

Theorem 3.1. *Let $x, y \in \mathcal{U}$ and $y_i \in \mathcal{U}$ for any $i \in x$. Then we have*

(i) $\emptyset$ and $\{x\}$ belong to $\mathcal{U}$,

(ii) $x \times y$, $x \sqcup y$, $x \cap y$ and y^x belong to $\mathcal{U}$,

(iii) if $z \in \mathcal{U} \cup \{L\}$ and $f : x \to z$, then f and $\mathrm{Ran}(f)$ belong to $\mathcal{U}$,

(iv) if $z \subseteq \mathcal{U}$ and $|z| \leq |x|$, then z belongs to $\mathcal{U}$,

(v) $\prod_{i \in x} y_i$, $\bigsqcup_{i \in x} y_i$ and $\bigcap_{i \in x} y_i$ belong to $\mathcal{U}$.

We can see that the whole set theory (i.e., unions, intersections, products of sets, singletons, the empty set) can be formed inside a universe $\mathcal{U}$, where, moreover, we may consider mappings from sets of $\mathcal{U}$ to the support of $\mathbf{L}$. Let us put $0 = \emptyset$, $1 = 0 \cup \{0\}$, $2 = 1 \cup \{1\}$, etc. Obviously, the natural numbers $0, 1, 2, \ldots$ belong to $\mathcal{U}$ and, according to (U4), an arbitrary finite set of natural numbers is also contained in $\mathcal{U}$. Nevertheless, the set of all natural numbers need not be an element of $\mathcal{U}$ (consider the universe of all finite sets) and it is only a subset of $\mathcal{U}$, in general.

The following theorem is very important in our investigation of equipollence of fuzzy sets (see Definition 4.4) and states that a set of $\mathcal{U}$ may be extended to a suitable greater sets of $\mathcal{U}$. Note that $\mathcal{U} \notin \mathcal{U}$.

Theorem 3.2. *Let $x \in \mathcal{U}$. Then there exists $y \in \mathcal{U}$ such that $|x| \leq |y \setminus x|$.*

Now, we can define the concept of fuzzy sets in $\mathcal{U}$ as follows.

Definition 3.2. Let $\mathcal{U}$ be a universe of sets over $\mathbf{L}$. A mapping $A : x \to L$ is called a fuzzy set in $\mathcal{U}$, if $x \in \mathcal{U}$.

ISBN: 978-989-95079-6-8

A consequence of (U2), (U4) and (U5) is that each fuzzy set (perceived as a relation $A \subseteq x \times L$) belongs to $\mathcal{U}$. Hence, all fuzzy sets form a subclass of $\mathcal{U}$. It is easy to see that fuzzy sets of higher order, i.e., fuzzy sets over sets of fuzzy sets, may be established inside of $\mathcal{U}$.

Let A be a fuzzy set in $\mathcal{U}$. The set $\mathrm{Dom}(A)$ is called a *universe of discourse of* A. We shall use $\mathcal{F}(\mathcal{U})$ to denote the class of all fuzzy sets in $\mathcal{U}$. Obviously, we have $\mathcal{F}(\mathcal{U}) \subseteq \mathcal{U}$, where $\mathcal{F}(\mathcal{U})$ is a proper class, whenever $\mathcal{U}$ is a proper class. The set $\mathrm{Supp}(A) = \{x \in \mathrm{Dom}(A) \mid A(x) > \bot\}$ is called *a support of fuzzy set* A. We shall say that A *is equivalent with* B, if $\mathrm{Supp}(A) = \mathrm{Supp}(B)$ and $A(x) = B(x)$ for any $x \in \mathrm{Supp}(A)$. For example, $A = \{0.9/a, 0/b\}$ is equivalent with $B = \{0.9, /a\}$. Obviously, the relation "to be equivalent" is an equivalence on $\mathcal{F}(\mathcal{U})$. We shall use $\mathrm{cls}(A)$ to denote the class of all equivalent fuzzy sets with A.

Definition 3.3. Let $A, B \in \mathcal{F}(\mathcal{U})$, $x = \mathrm{Dom}(A) \cup \mathrm{Dom}(B)$ and $A' \in \mathrm{cls}(A)$, $B' \in \mathrm{cls}(B)$ such that $\mathrm{Dom}(A') = \mathrm{Dom}(B') = x$. Then

- *the union* of A and B is a mapping $A \cup B : x \to L$ defined by
$$(A \cup B)(a) = A'(a) \vee B'(a) \tag{5}$$
for any $a \in x$,
- *the intersection* of A and B is a mapping $A \cap B : x \to L$ defined by
$$(A \cap B)(a) = A'(a) \wedge B'(a) \tag{6}$$
for any $a \in x$.

Definition 3.4. Let $A, B \in \mathcal{F}(\mathcal{U})$, $x = \mathrm{Dom}(A) \times \mathrm{Dom}(B)$ and $y = \mathrm{Dom}(A) \sqcup \mathrm{Dom}(B)$ (the disjoint union). Then

- *the product* of A, B is a mapping $A \times B : x \to L$ defined by
$$(A \times B)(a, b) = A(a) \wedge B(b) \tag{7}$$
for any $(a, b) \in x$,
- *the disjoint union* of A, B is a mapping $A \sqcup B : y \to L$ defined by
$$(A \sqcup B)(a, i) = \begin{cases} A(a, i), & \text{if } i = 1, \\ B(a, i), & \text{if } i = 2, \end{cases} \tag{8}$$
for any $(a, i) \in y$.

Definition 3.5. Let $A \in \mathcal{F}(\mathcal{U})$ and $A : x \to L$. Then *the complement of* A is a mapping $\overline{A} : x \to L$ defined by
$$\overline{A}(a) = A(a) \to \bot \tag{9}$$
for any $a \in x$.

Let us illustrate the proposed definitions by an example.

Example 3.1. Let $\mathbf{L}$ be the Łukasiewicz algebra and $A = \{1/a, 0.4/b\}$ and $B = \{0.6/a, 0.2/c\}$. Then we have

$$A \cup B = \{1/a, 0.4/b, 0.2/c\}$$
$$A \cap B = \{0.6/a, 0/b, 0/c\}$$
$$A \times B = \{0.6/(a, a), 0.2/(a, c), 0.4/(b, a), 0.2/(b, c)\}$$
$$A \sqcup B = \{1/(a, 1), 0.4/(b, 1), 0.6/(a, 2), 0.2/(c, 2)\}$$
$$\overline{A} = \{0/a, 0.6/b\}.$$

The following theorem shows that the relation "being equivalent" is a congruence for all mentioned operations except of the complement.

Theorem 3.3. *Let* $A, B \in \mathcal{F}(\mathcal{U})$ *and* $\circledast \in \{\cap, \cup, \times, \sqcup\}$. *If* $A' \in \mathrm{cls}(A)$ *and* $B' \in \mathrm{cls}(B)$, *then* $A' \circledast B' \in \mathrm{cls}(A \circledast B)$.

Although, fuzzy set in $\mathcal{U}$ is the most important concept in our theory, it seems to be useful (analogously to the classical set theory) to introduce the concept of fuzzy class in $\mathcal{U}$.

Definition 3.6. Let $\mathcal{U}$ be a universe of sets over $\mathbf{L}$. A mapping $\mathcal{A} : \mathcal{X} \to L$ is called a *fuzzy class in* $\mathcal{U}$, if $\mathcal{X} \subseteq \mathcal{U}$.

We shall use $\mathcal{C}(\mathcal{U})$ to denote the class of all fuzzy classes in $\mathcal{U}$. Obviously, we have $\mathcal{C}(\mathcal{U}) \not\subseteq \mathcal{U}$. From (U1) of Definition 3.1, it is easy to see that $x \subseteq \mathcal{U}$, whenever $x \in \mathcal{U}$. Hence, each fuzzy set is also a fuzzy class in $\mathcal{U}$, i.e., $\mathcal{F}(\mathcal{U}) \subseteq \mathcal{C}(\mathcal{U})$. Now there is a question: When can a fuzzy class be understood as a fuzzy set? One may observe that to each fuzzy set A there is a fuzzy class $\mathcal{A}$ such that $\mathrm{Dom}(\mathcal{A}) \notin \mathcal{U}$. However, such fuzzy class could be regarded as a fuzzy set in $\mathcal{U}$. This motivates us to establish the following condition of being a fuzzy set.

Definition 3.7. We shall say that a fuzzy class $\mathcal{A}$ in $\mathcal{U}$ is a fuzzy set, if there exists a fuzzy set A such that $A \in \mathrm{cls}(\mathcal{A})$. Otherwise, we say that it is *proper*.

4 Graded equipollence of fuzzy sets

As we have mentioned in Introduction, in the classical set theory, the relation of equipollence (equipotency or to have the same cardinality) on the class of all sets is defined using a one-to-one correspondence between sets. In order to introduce an equipollence for fuzzy sets in $\mathcal{U}$, we shall follow the classical idea and use "one-to-one mappings between fuzzy sets". Nevertheless, contrary to the classical approach, it seems to be reasonable to distinguish what one-to-one mapping is better than another one. This motivates us to introduce a degree in which a one-to-one mapping between suitable sets is a one-to-one mapping between fuzzy sets as follows.

Definition 4.1. Let $A, B \in \mathcal{F}(\mathcal{U})$, $x, y \in \mathcal{U}$ and $f : x \to y$ be a one-to-one mapping of x onto y in $\mathcal{U}$. We say that f is a *one-to-one mapping of* A *onto* B *in the degree* α, if $\mathrm{Supp}(A) \subseteq x \subseteq \mathrm{Dom}(A)$ and $\mathrm{Supp}(B) \subseteq y \subseteq \mathrm{Dom}(B)$ and
$$\alpha = \bigwedge_{z \in x} (A(z) \leftrightarrow B(f(z))). \tag{10}$$

On Fig. 1 we can see an example of a one-to-one mapping

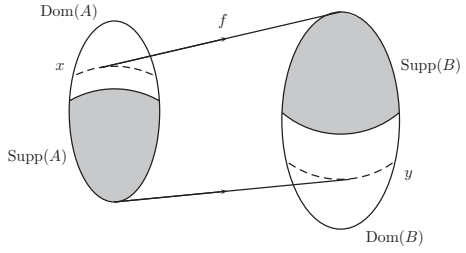


Figure 1: One-to-one mapping of A onto B

ISBN: 978-989-95079-6-8

f between fuzzy sets A and B, where not all elements of domains of A and B are respected. Obviously, $\mathrm{Dom}(A)\backslash x$ (analogously for B) contains the elements for which we have zero membership degree and thus they are unimportant elements of the fuzzy set A from this point of view. Moreover, there exists a fuzzy set A' with $\mathrm{Dom}(A') = x$ such that $A' \in \mathrm{cls}(A)$. This entitles us to omit such elements in our definition.

We shall write $[A \sim_f B] = \alpha$, if f is a one-to-one mapping of A onto B in the degree α. The following three examples demonstrate the notion of one-to-one mapping between fuzzy sets. For simplicity, we suppose that $\mathbf{L}$ is the Łukasiewicz algebra (see Example 2.2).

Example 4.1. Let $A = \{0.9/a, 0.5/b, 0/c\}$ and $B = \{0.5/1, 1/2, 0.2/3, 0/4\}$. If $f : \{a, b, c\} \to \{1, 2, 3\}$ is defined by $f(a) = 1$, $f(b) = 2$ and $f(c) = 3$, then we obtain $[A \sim_f B] = 0.5$, or by $f(a) = 2$, $f(b) = 1$ and $f(c) = 3$, then we obtain $[A \sim_f B] = 0.8$.

Example 4.2. Let A be the same as in Example 4.1 and $B' = \{0.5/1, 1/2, 0.2/3, 0.1/4\}$. Then there is no one-to-one mapping between A and B' satisfying the properties $\mathrm{Supp}(A) \subseteq x \subseteq \mathrm{Dom}(A)$ and $\mathrm{Supp}(B') \subseteq y \subseteq \mathrm{Dom}(B')$.

Example 4.3. Let us suppose that the set of all natural numbers N belongs to $\mathcal{U}$ and define $A, B : N \to [0, 1]$ by $A(n) = 1$ and

$$B(n) = \begin{cases} 1, & \text{if } n \text{ is an odd number,} \\ 0.1, & \text{otherwise.} \end{cases}$$

for any $n \in N$. Obviously, $[A \sim_f B] = 0.1$ for any one-to-one mapping between A and B. Notice that if we put 0 instead of 0.1 in the definition of B, then $[A \sim_f B] = 1$ for a suitable one-to-one mapping f.

From the latter two examples, we can see that there are some problems to derive "correct" degrees in which there exist one-to-one mappings between these fuzzy sets. More precisely, in Example 4.2, we cannot construct any one-to-one mapping between fuzzy sets, and in Example 4.3, the elements with a lower non-zero membership degree significantly influenced the degree of a one-to-one mapping between fuzzy sets. A natural idea how to solve these problems is to consider not just the given fuzzy sets, but the classes of all equivalent fuzzy sets with them. An example may be seen on Fig. 2,

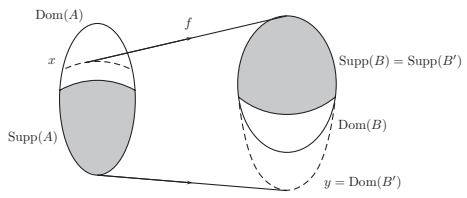


Figure 2: Extension of $\mathrm{Dom}(B)$ to $\mathrm{Dom}(B')$ for $B' \in \mathrm{cls}(B)$

where an extension of the original domain of fuzzy set B to a new domain of fuzzy set B' which is equivalent to B, i.e., $B' \in \mathrm{cls}(B)$, is given. Before we introduce the definition of being equipollent at least in the degree α, let us establish the following set of one-to-one correspondences.

Definition 4.2. Let $A, B \in \mathcal{F}(\mathcal{U})$. A mapping $f : x \to y$ belongs to the set $\mathrm{Bij}(A, B)$, if

(i) f is a one-to-one mapping of x onto y,

(ii) $\mathrm{Supp}(A) \subseteq x \subseteq \mathrm{Dom}(A)$, and

(iii) $\mathrm{Supp}(B) \subseteq y \subseteq \mathrm{Dom}(B)$.

The definition of being equipollent at least in some degree is following.

Definition 4.3. Let $A, B \in \mathcal{F}(\mathcal{U})$. We say that A *is equipollent with* B (or A has the same cardinality as B) *at least in the degree* α, if there exists $A' \in \mathrm{cls}(A)$, $B' \in \mathrm{cls}(B)$ and $f \in \mathrm{Bij}(A', B')$ such that $[A' \sim_f B'] \geq \alpha$.

Example 4.4. Let us suppose A an B' from Examples 4.1 and 4.2. Then A is equipollent with B' at least in the degree 0.8, since $[A' \sim_f B'] = 0.8$, where $A' : \{a, b, c, d\} \to [0, 1]$ with $A'(x) = A(x)$ for $x \in \{a, b, c\}$ and $A(d) = 0$, and

$$f(a) = 2,\ f(b) = 1,\ f(c) = 3,\ f(d) = 4.$$

The following result is clearly expected.

Lemma 4.1. *Let $A \in \mathcal{F}(\mathcal{U})$ and $A' \in \mathrm{cls}(A)$. Then A is equipollent with A' at least in the degree $\top$.*

If we can find a degree α in which some fuzzy sets A and B are "at least equipollent", i.e, there exists a one-to-one mapping f and $A' \in \mathrm{cls}(A)$, $B' \in \mathrm{cls}(B)$ with $[A' \sim_f B'] \geq \alpha$, a natural question arises, how to define a degree in which fuzzy sets A and B are equipollent. Obviously, one way, how to do it, is to use the supremum from all degrees of possible one-to-one mappings that can be created between all couples of fuzzy sets that are equivalent with original ones. However, here is one problem. If $\mathcal{U}$ is a proper class, then the class $\mathrm{cls}(A)$ of all equivalent fuzzy sets with A is proper. Hence, we cannot use the supremum in the mentioned sense. The following definition proposes a way, how to introduce a degree in which two fuzzy sets are equipollent and to avoid the proper classes.

Definition 4.4. Let $A, B \in \mathcal{F}(\mathcal{U})$. We shall say that A *is equipollent with* B (or A has the same cardinality as B) *in the degree* α, if there exist fuzzy sets $C \in \mathrm{cls}(A)$ and $D \in \mathrm{cls}(B)$ such that

$$\alpha = \bigvee_{f \in \mathrm{Bij}(C,D)} [C \sim_f D] \tag{11}$$

and, for each $A' \in \mathrm{cls}(A)$, $B' \in \mathrm{cls}(B)$ and $f \in \mathrm{Bij}(A', B')$, there is $[A' \sim_f B'] \leq \alpha$.

If A is equipollent with B in the degree α, then we shall write $[A \sim B] = \alpha$. One may be surprised, that it is sufficient to restrict ourselves to suitable fuzzy sets, and may ask, if this definition is really correct. That means, if, for arbitrary fuzzy sets A and B, there exists the degree in which they are equipollent. The correctness of our definition follows from the following theorem.

Theorem 4.2. *Let $A, B \in \mathcal{F}(\mathcal{U})$. Then, for any $C \in \mathrm{cls}(A)$ and $D \in \mathrm{cls}(B)$ such that $|\mathrm{Supp}(A)| \leq |\mathrm{Dom}(D)\backslash\mathrm{Supp}(B)|$ and $|\mathrm{Supp}(B)| \leq |\mathrm{Dom}(C) \setminus \mathrm{Supp}(A)|$, there is*

$$[A \sim B] = \bigvee_{f \in \mathrm{Bij}(C,D)} [C \sim_f D]. \tag{12}$$

ISBN: 978-989-95079-6-8

Example 4.5. Let A, B' be fuzzy sets from Example 4.1 and 4.2 and define $C, D : \{a, b, c, 1, 2, 3, 4\} \to [0, 1]$ such that $C \in \mathrm{cls}(A)$ and $D \in \mathrm{cls}(B')$. According to previous theorem, we can write

$$[A \sim B] = \bigvee_{f \in \mathrm{Bij}(C,D)} [C \sim_f D] = [C \sim_{f_0} D] = 0.8,$$

where $f_0 : \{a, b, c, 1\} \to \{1, 2, 3, 4\}$ is defined by $f_0(a) = 1$, $f_0(b) = 2$, $f_0(c) = 3$ and $f_0(1) = 4$ (compare with Example 4.4).

Example 4.6. Let A, B be fuzzy sets from Example 4.3. Put $M = \{0, 0.5, 1, 1.5, \dots\}$, where e.g. $0.5 = \{\{0, 1\}\}$, $1.5 = \{\{1, 2\}\}$, etc. and define $C, D : M \to [0, 1]$ such that $C \in \mathrm{cls}(A)$ and $D \in \mathrm{cls}(B)$. One checks easily that $[A \sim B] = [C \sim_f D] = 0.9$, where $f(m) = 2m + 1$ for any $m \in M$.

Lemma 4.3. *Let $f \in \mathrm{Bij}(A, B)$ and $g \in \mathrm{Bij}(B', C)$, where $B' \in \mathrm{cls}(B)$. Then there exist $C' \in \mathrm{cls}(C)$ and $h \in \mathrm{Bij}(A, C')$ such that*

$$[A \sim_f B] \otimes [B' \sim_g C] \le [A \sim_h C']. \tag{13}$$

Recall that $\mathcal{F}(\mathcal{U}) \times \mathcal{F}(\mathcal{U})$ is a subclass of $\mathcal{U}$. Hence, we can define the following fuzzy class (or also fuzzy class relation) in $\mathcal{U}$, which is analogous to the equivalence relation "to be equipollent" on the class of all sets.

Theorem 4.4. *The fuzzy class relation $\sim : \mathcal{F}(\mathcal{U}) \times \mathcal{F}(\mathcal{U}) \to L$ is a similarity relation ($\otimes$-equivalence) on the class $\mathcal{F}(\mathcal{U})$, i.e.,*

(i) $[A \sim A] = \top$,

(ii) $[A \sim B] = [B \sim A]$,

(iii) $[A \sim B] \otimes [B \sim C] \le [A \sim C]$

hold for arbitrary fuzzy sets $A, B, C \in \mathcal{F}(\mathcal{U})$.

The fuzzy class relation $\sim$ is called the *equipollence of fuzzy sets* and $[A \sim B]$ is the *degree of equipollence* of fuzzy sets A and B. The following statement is straightforward.

Corolarry 4.5. *Let $A, B \in \mathcal{F}(\mathcal{U})$ and $A' \in \mathrm{cls}(A)$. Then $[A \sim B] = [A' \sim B]$.*

In classical set theory, if $f : x \to z$ is a one-to-one mapping of x to z and $|x| = |y|$, then also $|f(x)| = |y|$, where $f(x) = \mathrm{Ran}(f)$. Analogously, we can give the following statement for fuzzy sets.

Theorem 4.6. *Let $A, B \in \mathcal{F}(\mathcal{U})$ and $f : \mathrm{Dom}(A) \to z$ be a one-to-one mapping in $\mathcal{U}$. If we put*

$$f(A)(y) = \begin{cases} A(x), & \text{if } f(x) = y, \\ \bot, & \text{otherwise,} \end{cases} \tag{14}$$

for any $y \in z$, then $[A \sim B] = [f(A) \sim B]$.

A simple consequence of Theorems 4.4 and 4.6 is the following statement.

Corolarry 4.7. *Let $A \in \mathcal{F}(\mathcal{U})$ and $f : \mathrm{Dom}(A) \to z$ be a one-to-one mapping in $\mathcal{U}$. Then $[A \sim f(A)] = \top$.*

Now there is a question, if the definition of graded equipollence, which is more complicated and practically useless, could be simplified. The remaining part of this section will give a partial positive answer.

Let $A, B \in \mathcal{F}(\mathcal{U})$ and $\mathrm{Dom}(A) = x$, $\mathrm{Dom}(B) = y$. Put

$$A^*(z, i) = \begin{cases} A(z), & \text{if } i = 1, \\ \bot, & \text{if } i = 2, \end{cases} \tag{15}$$

$$B^*(z, i) = \begin{cases} \bot, & \text{if } i = 1, \\ B(z), & \text{if } i = 2 \end{cases} \tag{16}$$

for any $(z, i) \in x \sqcup y$. According to Corollary 4.7, we have $[A \sim A^*] = \top$ and $[A \sim B] = [A^* \sim B^*]$ and, hence, it does not matter, whether we consider the original fuzzy sets A, B or their images A^*, B^* to obtain the right degree in which A and B are equipollent. A straightforward consequence of Theorem 4.2 is the following lemma.

Lemma 4.8. *Let $A, B \in \mathcal{F}(\mathcal{U})$. Then*

$$[A^* \sim B^*] = \bigvee_{f \in \mathrm{Bij}(A^*, B^*)} [A^* \sim_f B^*]. \tag{17}$$

The following lemma shows that we need not consider all one-to-one mappings to obtain the right degree of equipollence.

Lemma 4.9. *Let $A, B \in \mathcal{F}(\mathcal{U})$ and $f \in \mathrm{Bij}(A^*, B^*)$. Then for any $z \in \mathcal{U}$ with*

$$\mathrm{Dom}(f) \subseteq z \subseteq \mathrm{Dom}(A) \sqcup \mathrm{Dom}(B) \tag{18}$$

there exists $f' \in \mathrm{Bij}(A^*, B^*)$ with $\mathrm{Dom}(f') = z$ such that

$$[A^* \sim_f B^*] = [A^* \sim_{f'} B^*]. \tag{19}$$

From the previous lemma, one could notice that the set $\mathrm{Bij}(A^*, B^*)$ is too large and it is sufficient to consider the one-to-one mappings with the same domain $\mathrm{Dom}(A^*)$ or conversely with the same range $\mathrm{Dom}(B^*)$. Hence, we may ask, whether we can restrict ourselves to one-to-one correspondences between $\mathrm{Dom}(A^*)$ and $\mathrm{Dom}(B^*)$. The answer is positive as the following theorem shows and this fact is a consequence of Cantor-Bernstein theorem. We shall use $\mathrm{Perm}(A, B)$ to denote the set of all one-to-one mappings of $\mathrm{Dom}(A)$ onto $\mathrm{Dom}(B)$.

Theorem 4.10. *Let $A, B \in \mathcal{F}(\mathcal{U})$. Then*

$$[A^* \sim B^*] = \bigvee_{f \in \mathrm{Perm}(A^*, B^*)} [A^* \sim_f B^*]. \tag{20}$$

A generalization of this theorem to a wider scale of fuzzy sets is as follows (cf. Theorem 4.2).

Theorem 4.11. *Let $A, B \in \mathcal{F}(\mathcal{U})$. Then, for any $C \in \mathrm{cls}(A)$ and $D \in \mathrm{cls}(B)$ such that $|\mathrm{Supp}(A)| = |\mathrm{Dom}(D) \setminus \mathrm{Supp}(B)|$ and $|\mathrm{Supp}(B)| = |\mathrm{Dom}(C) \setminus \mathrm{Supp}(A)|$, there is*

$$[A \sim B] = \bigvee_{f \in \mathrm{Perm}(C, D)} [C \sim_f D]. \tag{21}$$

If we deal with fuzzy sets in $\mathcal{U}$ with finite supports, then it is sufficient to consider fuzzy sets which domains have the same cardinality as the following theorem shows.

ISBN: 978-989-95079-6-8

Theorem 4.12. *Let $A, B \in \mathcal{F}(\mathcal{U})$ and $C \in \mathrm{cls}(A)$, $D \in \mathrm{cls}(B)$ be such that $|\mathrm{Dom}(C)| = |\mathrm{Dom}(D)| = n$. Then*

$$[A \sim B] = \bigvee_{f \in \mathrm{Perm}(C,D)} [C \sim_f D]. \tag{22}$$

Now we may ask, when there exists just one one-to-one mapping between two fuzzy sets which determines the degree of their equipollence. If the given residuated lattice is not linearly ordered or is a complete linearly ordered residuated lattice with an infinite support, then the following examples show simple constructions of fuzzy sets, when more than one one-to-one mapping between them have to be used to find the degrees of their equipollence.

Example 4.7. Let $\mathbf{L}$ be a residuated lattice from Example 2.2, $A = \{\bot/0, b/1, \top/2\}$ and $B = \{a/0, b/1, b/2\}$. One checks easily that $[A \sim B] = \top$ but there is no permutation on x giving the degree of equipollence of A and B equal to $\top$.

Example 4.8. Let $\mathbf{L}$ be the Łukasiewicz algebra and let us suppose that the set of all natural numbers N belongs to $\mathcal{U}$. Define $A, B : N \to L$ by $A(n) = 0.5 - \frac{1}{n+2}$ for any $n \in N$ and $B(0) = 1$, $B(n) = 0$ for any $n \in N$, where $n \neq 0$. Since the presumptions of Theorem 4.11 are satisfied, we can use (21) to compute the degree of equipollence. If f is a permutation on N such that $f(n) = 0$, then one checks easily that

$$[A \sim_f B] = A(n) \leftrightarrow B(0) = 0.5 - \frac{1}{n+2}. \tag{23}$$

Note that $A(n) \leftrightarrow 0 = 0.5 + \frac{1}{n+2}$. Hence, we obtain

$$[A \sim B] = \bigvee_{f \in \mathrm{Perm}(A,B)} [A \sim_f B] = 0.5,$$

but, due to (23), no permutation on N gives such value.

Let N be the set of all natural numbers and $A \in \mathcal{F}(\mathcal{U})$ be a fuzzy set with finite support. Define a fuzzy class $p_A : N \to L$ by

$$p_A(i) = \bigvee_{\substack{z \subseteq \mathrm{Supp}(A),\\ |z| = i}} \bigwedge_{x \in z} A(x) \tag{24}$$

for any $i \in N$.

Theorem 4.13. *Let $\mathbf{L}$ be a complete linearly ordered residuated lattice and $A, B \in \mathcal{F}(\mathcal{U})$ be fuzzy sets with finite supports. Then*

$$[A \sim B] = \bigwedge_{i \in N} (p_A(i) \leftrightarrow p_B(i)).$$

Corolarry 4.14. *Let $\mathbf{L}$ be a linearly ordered residuated lattice and $A, B \in \mathcal{F}(\mathcal{U})$ be fuzzy sets with $|\mathrm{Dom}(A)| = |\mathrm{Dom}(B)| = n$. Then there exists $f \in \mathrm{Perm}(A, B)$ such that $[A \sim B] = [A \sim_f B]$.*

Example 4.9. Let $\mathbf{L}$ be the Łukasiewicz algebra and $x = \{c_1 \ldots, c_5\}$ be a set of criterion. Let us suppose that the following fuzzy sets

$$A = \{0.3/c_1, 0.8/c_2, 0.4/c_3, 0/c_4, 0/c_5\}$$
$$B = \{0/c_1, 0.6/c_2, 0.9/c_3, 0/c_4, 0.7/c_5\}$$

express the satisfaction of all criterion by two applicants and our goal is to determine, whether the numbers of satisfied criterion are more or less the same for both applicants. According to (24), we can establish

$$p_A = \{1/0, 0.8/1, 0.4/2, 0.3/3, 0/4, 0/5, \ldots\},$$
$$p_B = \{1/0, 0.9/1, 0.7/2, 0.6/3, 0/4, 0/5, \ldots\}$$

and, using Theorem 4.13, we obtain

$$[A \sim B] = \bigwedge_{i \in N} (p_A(i) \leftrightarrow p_B(i)) = 0.7.$$

Hence, we can conclude that both applicants have more or less the same number of satisfied criterion.

The following theorem demonstrates a way how we can investigate the well-known relations in cardinal set theory for fuzzy sets in $\mathcal{U}$.

Theorem 4.15. *Let $A, B \in \mathcal{F}(\mathcal{U})$. Then we have*

(i) $[A \sim B] \leq [\overline{A} \sim \overline{B}]$

(ii) $[A \sim B] \otimes [C \sim D] \leq [A \times C \sim B \times D]$,

(iii) if $\mathrm{Supp}(A \cap B) = \mathrm{Supp}(C \cap D) = \emptyset$, *then*

$$[A \sim C] \otimes [B \sim D] \leq [A \cup C \sim B \cup D].$$

5 Conclusion

Let us note that the fuzzy set cardinality is not widely developed in a comparison to the cardinal theory for sets giving many nice result. The proposed definitions of the universe, operations and equipollence of fuzzy sets are a starting point in a deeper fuzzy sets investigation from the cardinality point of view. We think that the results could be interesting for the fuzzy set theory as well as for the practical application as e.g. a determination of "number" of well satisfied criteria.

Acknowledgment

This paper has been supported by the Grant IAA108270901 of the GA AV ČR.

References

[1] D. Klaua. Zum kardinalzahlbegriff in der mehrwertigen mengenlehre. In G. Asser, J. Flashmayers, and W. Rinow, editors, *Theory of Sets and Topology*, pages 313–325. Deutsher Verlag der Wissenshaften, Berlin, 1972.

[2] S. Gottwald. Zahlbereichskonstruktionen in einer mehrwertigen mengenlehre. *Z. Math. Logic Grundl. Math.*, 17:145–188, 1971.

[3] S. Gottwald. Fuzzy uniqueness of fuzzy mappings. *Fuzzy Sets and Systems*, 3:49–74, 1980.

[4] M. Wygralak. Generalized cardinal numbers and operations on them. *Fuzzy Sets and Systems*, 53(1):49–85, 1993.

[5] M. Wygralak. An axiomatic approach to scalar cardinalities of fuzzy sets. *Fuzzy Sets and Systems*, 110(2):175–179, 2000.

[6] M. Wygralak. Fuzzy sets with triangular norms and their cardinality theory. *Fuzzy Sets and Systems*, 124(1):1–24, 2001.

[7] M. Wygralak. *Cardinalities of Fuzzy Sets*. Kluwer Academic Publisher, Berlin, 2003. isbn: 3-5400-0337-1.

[8] E.P. Klement, R. Mesiar, and E. Pap. *Triangular norms*, volume 8 of *Trends in Logic*. Kluwer Academic Publisher, Dordrecht, 2000.

[9] C. McLarty. *Elementary Categories, Elementary Toposes*. Oxford University Press, 1992.

Managing the Absence of Items in Fuzzy Association Mining

C. Molina[1] D. Sanchez[2] J.M. Serrano[1] M.A. Vila[2]

1.Department of Computer Science, University of Jaen
Jaen, Spain
2.Department of Computer Science and Artificial Intelligence, University of Granada
Granada, Spain
Email: carlosmo@ujaen.es, daniel@decsai.ugr.es, jschica@ujaen.es, vila@decsai.ugr.es

Abstract— *One of the most well-known and extended data mining techniques is that of association rule mining, a helpful tool to discover relations between items present in sets of transactions. Nevertheless, in some other scenarios, another interesting issue is that of considering not only the possible relations involving presence of items, but the absence of them. The problem gets more complex when it is necessary to represent also imprecision and/or uncertainty in the information. In this paper, we introduce a methodology to obtain fuzzy association rules involving absent items. Additionally, our proposal is based on restriction level sets, a recent representation of imprecision that extends that of fuzzy sets, and introduces some new operators, covering some misleading results obtained from usual fuzzy operators as, for example, negation. In our methodology, we define new measures of interest and accuracy for fuzzy association rules as RL-numbers, as well as we propose a new way of summarizing the resulting set of fuzzy association rules, distributed in restriction levels.*

Keywords— Absence of items, fuzzy association rules, restriction levels

1 Introduction

As the amount of information stored in databases grows continuously day after day, it is desirable the development of tools, not only to properly manage all this knowledge, but to get a profit from it. Over the past two decades, considerable efforts have been devoted to the study and development of techniques in data mining and knowledge discovery in databases. One of the most well-known and extended data mining techniques is that of association rule mining, a helpful tool to discover implicit, non-trivial and potentially useful relations between items in sets of transactions. The most known example if that of market baskets, but association rules can be also applied on relational databases. Since the original association rule definition was proposed in [1], many studies and related methodologies have been devoted to extend this tool in order to manage different scenarios and knowledge representations.

One interesting issue, approached in works like [2], and [3], is that of considering both present and absent items. In these cases, it is not only interesting to detect relations between items included in transactions, but also if the absence of given items in the transaction can be related to the presence of some others. One possible application can be the discovery of conflicts or complements between items, as well as the establishment of constraints in the sets of items involved in the data mining procedure. Some other authors have shown interest in this problem, and additional information and efficient algorithms can be found in [4],[5],[6], and [7].

On the other hand, another interesting issue is that of applying the theory of fuzzy sets [8] in the modeling of how items are included in transactions, leading to the definition of fuzzy association rules. In the literature, some interesting approaches are discussed in [9], [10], [11], [12], and [13], among others.

But, as far as we know, none of the cited works manage properly the absence of items in the fuzzy case. In particular, consider an item i that belongs to a fuzzy transaction with a membership degree of 0.5. According to the usual fuzzy extension of the complement operator, the complementary item $\neg i$ should belong to the same transaction with a membership degree of $1 - 0.5 = 0.5$, leading us to counterintuitive results. Restriction level sets [14], a recently introduced representation of fuzzy quantities that extends the representation by means of fuzzy sets, manage to solve scenarios like these.

The representation via restriction level sets can be seen as equivalent to that of fuzzy sets seen as collections of alpha cuts, and hence it offers a alternative approach to the representation and management of fuzzy quantities, differing from fuzzy sets in the sense that operations are defined on levels. However, some operations as negation result in a representation that may not correspond to the alpha cuts of a fuzzy set.

Actually, although a parallel definition of a logic model for association rules based also on restriction level sets can be found in [15], our proposal, in this paper, is a methodology for mining fuzzy association rules via restriction level sets, instead of fuzzy sets, extending the definition introduced in [16], and, in addition, considering both present and absent items.

The paper is organized as follows. Sections 2 and 3 are devoted to introduce the reader in the followed representations and previous methodologies. Next, we introduce our proposal in section 4, where we define new measures for fuzzy association rules, based on RL-numbers, and we describe some derived properties from these. Next, some aspects with respect to the implementation and our first experimentation results are discussed in section 5. Finally, we present our concluding remarks as well as propose some interesting open issues in the conclusions sections.

2 Data mining tools

2.1 Association rules

Given a set I ("set of items") and a set of transactions T (also called T-set), each transaction being a subset of I, association

ISBN: 978-989-95079-6-8

rules [1] are "implications" of the form $A \Rightarrow C$ that relate the presence of itemsets (sets of items) A and C in transactions of T, assuming $A, C \subseteq I$, $A \cap C = \emptyset$ and $A, C \neq \emptyset$.

In the case of relational databases, it is usual to consider that items are pairs $\langle attribute, value \rangle$, and transactions are tuples in a table. For example, the item $\langle X, x_0 \rangle$ is in the transaction associated to a tuple t iff $t[X] = x_0$.

The ordinary measures proposed in [1] to assess association rules are *confidence* (the conditional probability $p(C|A)$) and *support* (the joint probability $p(A \cup C)$). An alternative framework was proposed in [17]. In this framework, accuracy is measured by means of Shortliffe and Buchanan's certainty factors [18], in the following way: the certainty factor of the rule $A \Rightarrow C$ is

$$CF(A \Rightarrow C) = \frac{(Conf(A \Rightarrow C)) - S(C)}{1 - S(C)} \quad (1)$$

if $Conf(A \Rightarrow C) > S(C)$, and

$$CF(A \Rightarrow C) = \frac{(Conf(A \Rightarrow C)) - S(C)}{S(C)} \quad (2)$$

if $Conf(A \Rightarrow C) < S(C)$, and 0 otherwise.

Certainty factors take values in $[-1, 1]$, indicating the extent to which our belief that the consequent is true varies when the antecedent is also true. It ranges from 1, meaning maximum increment (i.e., when A is true then C is true) to -1, meaning maximum decrement.

2.2 Fuzzy association rules

In [16], the model for association rules is extended in order to manage fuzzy values in databases. The approach is based on the definition of fuzzy transactions as fuzzy subsets of items. Let $I = \{i_1, \ldots, i_m\}$ be a set of items and T' be a set of fuzzy transactions, where each fuzzy transaction is a fuzzy subset of I. Let $\tilde{\tau} \in T'$ be a fuzzy transaction, we note $\tilde{\tau}(i_k)$ the membership degree of i_k in $\tilde{\tau}$. A fuzzy association rule is an implication of the form $A \Rightarrow C$ such that $A, C \subset I$ and $A \cap C = \emptyset$.

It is immediate that the set of transactions where a given item appears is a fuzzy set. We call it *representation* of the item. For item i_k in T' we have the following fuzzy subset of T':

$$\tilde{\Gamma}_{i_k} = \sum_{\tilde{\tau} \in T'} \tilde{\tau}(i_k)/\tilde{\tau} \quad (3)$$

This representation can be extended to itemsets as follows: let $I_0 \subset I$ be an itemset, its representation is the following subset of T':

$$\tilde{\Gamma}_{I_0} = \bigcap_{i \in I_0} \tilde{\Gamma}_i = min_{i \in I_0} \tilde{\Gamma}_i \quad (4)$$

In order to measure the interest and accuracy of a fuzzy association rule, we must use approximate reasoning tools, because of the imprecision that affects fuzzy transactions and, consequently, the representation of itemsets. In [16], a semantic approach is proposed based on the evaluation of quantified sentences (see [19]). Let Q be a fuzzy coherent quantifier:

- The support of an itemset $\tilde{\Gamma}_{I_0}$ is equal to the result of evaluating the quantified sentence Q of T' are $\tilde{\Gamma}_{I_0}$.
- The support of the fuzzy association rule $A \Rightarrow C$ in the FT-set T', $Supp(A \Rightarrow C)$, is the evaluation of the quantified sentence Q of T are $\tilde{\Gamma}_{A \cup C}$ = Q of T are $(\tilde{\Gamma}_A \cap \tilde{\Gamma}_C)$.
- The confidence of the fuzzy association rule $A \Rightarrow C$ in the FT-set T', $Conf(A \Rightarrow C)$, is the evaluation of the quantified sentence Q of $\tilde{\Gamma}_A$ are $\tilde{\Gamma}_C$.

As seen in [16], the proposed method is a generalization of the ordinary association rule assessment framework in the crisp case.

3 Restriction-level representation

3.1 Representation

The RL-representation of an imprecise property is a collection of crisp sets, each crisp set corresponding to a crisp realization of the property under a *restriction* rule. We distinguish between *atomic* and *derived* properties. Atomic properties are those that cannot be defined in terms of other properties in our problem. Derived properties are defined by logical operations on other properties.

In [14] we consider that atomic imprecise properties are represented by fuzzy sets, and hence restrictions are of the form $degree \geq \alpha$ with $\alpha \in (0, 1]$, and restriction levels are associated to values $\alpha \in (0, 1]$. In the same case, the crisp realization of an atomic imprecise property represented by a fuzzy set A in the restriction level α corresponds to the α-cut A_α.

For every property we assume that there is a finite set of restriction levels $\Lambda = \{\alpha_1, \ldots, \alpha_m\}$ verifying that $1 = \alpha_1 > \alpha_2 > \cdots > \alpha_m > \alpha_{m+1} = 0$, $m \geq 1$. We call such sets *RL-sets*. The consideration that a RL-set is finite is not a practical limitation since humans are able to distinguish a limited number of restriction or precision levels and, in practice, the limit in precision and storage of computers allows us to work with a finite number of degrees (and consequently, of levels) only. In practice, the RL-set for an atomic property represented by a fuzzy set A on an universe X is

$$\Lambda_A = \{A(x) \mid x \in support(A)\} \cup \{1\} \quad (5)$$

The RL-set employed to represent a derived property is obtained as the union of the RL-sets of the atomic properties in terms of which the property is defined. Finally, the RL-representation of an imprecise property on X is defined in [14] as follows:

Definition 1 *A RL-representation is a pair (Λ, ρ) where Λ is a RL-set and ρ is a function*

$$\rho : \Lambda \rightarrow \mathcal{P}(X) \quad (6)$$

The function ρ indicates the crisp realization that represents the imprecise property for each restriction level. As an example, the RL-representation for an atomic imprecise property defined by a fuzzy set A is the pair (Λ_A, ρ_A), where Λ_A is obtained using equation (5), and $\rho_A(\alpha) = A_\alpha \; \forall \alpha \in \Lambda_A$.

Given an imprecise property P represented by (Λ_P, ρ_P), we define the set of crisp representatives of P, Ω_P, as

$$\Omega_P = \{\rho_P(\alpha) \mid \alpha \in \Lambda_P\} \quad (7)$$

ISBN: 978-989-95079-6-8

For an atomic property A, the set of crisp representatives Ω_A is the set of significant α-cuts of A, as we have seen. However, notice that in definition 1 there is no restriction about the possible crisp representatives for non-atomic properties. In particular, as a consequence of operations, they don't need to be nested, so the final RL-representation of a derived property is not always equivalent to the α-cut representation of fuzzy sets.

In order to define properties by operations, it is convenient to extend the function ρ to any RL $\alpha \in (0,1]$. Let (Λ, ρ) be a RL-representation with $\Lambda = \{\alpha_1, \ldots, \alpha_m\}$ verifying $1 = \alpha_1 > \alpha_2 > \cdots > \alpha_m > \alpha_{m+1} = 0$. Let $\alpha \in (0,1]$ and $\alpha_i, \alpha_{i+1} \in \Lambda$ such that $\alpha_i \geq \alpha > \alpha_{i+1}$. Then

$$\rho(\alpha) = \rho(\alpha_i) \tag{8}$$

Finally, let us remark that a RL-representation (Λ, ρ) on X is a crisp set $A \subseteq X$ iff $\forall \alpha \in \Lambda$, $\rho(\alpha) = A$.

3.2 Interpretation in terms of evidence

Given a RL-representation (Λ_A, ρ_A) for an atomic property A, the values of Λ_A can be interpreted as values of possibility of a possibility measure defined $\forall \rho_A(\alpha_i) \in \Omega_A$ as

$$Pos(\rho_A(\alpha_i)) = \alpha_i. \tag{9}$$

Following this interpretation we define a basic probability assignment in the usual way:

Definition 2 *Let (Λ, ρ) be a RL-representation with crisp representatives Ω. The associated probability distribution $m : \Omega \to [0,1]$ is*

$$m(Y) = \sum_{\alpha_i \mid Y = \rho(\alpha_i)} \alpha_i - \alpha_{i+1}. \tag{10}$$

The basic probability assignment m_F gives us information about how representative of the property F is each crisp set in Ω_F. For each $Y \in \Omega_F$, the value $m_F(Y)$ represents the proportion to which the available evidence supports the claim that the property F is represented by Y. From this point of view, a RL-representation can be seen as a basic probability assignment in the sense of the theory of evidence, *plus a structure indicating dependencies between the possible representations of different properties*.

3.3 RL-numbers

On the basis of RL-representations and operations, RL-numbers were introduced in [20] as a representation of imprecise quantities.

Definition 3 *A RL-real number is a pair $(\Lambda, \mathcal{R})$ where Λ is a RL-set and $\mathcal{R} : (0,1] \to \mathbb{R}$.*

We shall note $\mathbb{R}_{RL}$ the set of RL-real numbers. The RL-real number R_x is the representation of a (precise) real number x iff $\forall \alpha \in \Lambda_{R_x}$, $\mathcal{R}_{R_x}(\alpha) = x$. We shall denote such RL-real number as R_x or, equivalently, x, since in the crisp case, the set Λ_{R_x} is unimportant. Operations are extended as follows:

Definition 4 *Let $f : \mathbb{R}^n \to \mathbb{R}$ and let $R_1 \ldots R_n$ be RL-real numbers. Then $f(R_1, \ldots, R_n)$ is a RL-real number with*

$$\Lambda_{f(R_1,\ldots,R_n)} = \bigcup_{1 \leq i \leq n} \Lambda_{R_i} \tag{11}$$

and, $\forall \alpha \in \Lambda_{f(R_1,\ldots,R_n)}$

$$\mathcal{R}_{f(R_1,\ldots,R_n)}(\alpha) = f(\mathcal{R}_{R_1}(\alpha), \ldots, \mathcal{R}_{R_n}(\alpha)) \tag{12}$$

This approach offers two main advantages:

- RL-numbers are representations of fuzzy quantities that can be easily obtained by extending usual crisp measurements to fuzzy sets.
- Arithmetic and logical operations on RL-numbers are straightforward and unique extensions of the operations on crisp numbers, verifying the following:
  - They verify all the usual properties of crisp arithmetic and logical operations.
  - The imprecision does not necessarily increase through operations, and can even diminish. The maximum imprecision is related to the number of restriction levels employed.

4 Fuzzy association rules via restriction levels

As we noted before, restriction level sets can be viewed as sets of alpha cuts in fuzzy sets. Actually, if the restriction level set represents an atomic property, the representation is equal to that of fuzzy sets. Following this idea, we extend the definition of fuzzy association rules introduced in [16], using restriction level sets instead. Moreover, since each restriction level corresponds to a crisp set, and crisp operators hold for each level, we can define a fuzzy association rule via restriction levels as an aggregation of all the related (crisp) association rules extracted at each restriction level.

In our approach, we represent items via restriction levels as follows. Let $I = \{i_1, \ldots, i_m\}$ be a set of items and T' be a set of fuzzy transactions, where each fuzzy transaction is a restriction level set of I. Let $\tilde{\tau} \in T'$ be a fuzzy transaction, we note $\tilde{\tau}(i_k)$ the membership degree of i_k in $\tilde{\tau}$. Let $supp_\alpha(i_k)$ be the support of i_k (i.e., the probability that $\tilde{\tau}(i_k) > \alpha, \forall \tilde{\tau} \in T'$) at the restriction level α.

On the other hand, let $\neg i_k$ be the negation (absence) of item i_k, and let $supp_\alpha(\neg i_k) = 1 - supp_\alpha(i_k)$ be the support of $\neg i_k$ at the restriction level α. Note how, in the case of negated items, the subsequent restriction level set may not correspond necessarily to a fuzzy set.

Let $\bar{I} = \{\neg i_1, \ldots, \neg i_m\}$ be the set of negated items from I, and $A \subset \{I \cup \bar{I}\}, C \in \{I \cup \bar{I}\}, A \cap C = \emptyset$, the antecedent and the consequent of the fuzzy association rule $R : A \Rightarrow C$. If we see this fuzzy association rule R under its RL-representation as the pair (Λ_R, ρ_R), we define the set of crisp representations of R, Ω_R, as

$$\Omega_R = \{\rho_R(\alpha) \mid \alpha \in \Lambda_R\} \tag{13}$$

where $\rho_R)(\alpha) = (R)_\alpha$, that is, ρ represents the association rules at each restriction level. In other words, this fuzzy association rule can be viewed as the aggregation of all (crisp)

ISBN: 978-989-95079-6-8

association rules of the form $A \Rightarrow C$ that can be obtained at each restriction level α.

Let $Supp_\alpha(A \Rightarrow C)$, $Conf_\alpha(A \Rightarrow C)$, and $CF_\alpha(A \Rightarrow C)$ be, respectively, the support, confidence and certainty factor of the crisp association rule $Supp_\alpha(A \Rightarrow C)$ at restriction level α.

Following the definitions in section 2.2, we must define the measures of interest and accuracy of a fuzzy association rule in terms of evaluation of quantified sentences. Hence, we define the support, confidence and certainty factor of the fuzzy association rule $A \Rightarrow C$ as RL-numbers, according to the RL-evaluation process proposed in [21].

Definition 5 *Let m_R be the basic probability assignment for association rule $R : A \Rightarrow C$ at each restriction level. The support of the fuzzy association rule $A \Rightarrow C$, RL-Supp$(A \Rightarrow C)$, is given by*

$$RLSupp(A \Rightarrow C) = \sum_{(A \Rightarrow C)_\alpha \in \Omega_R} m_R((A \Rightarrow C)_\alpha) \cdot Supp_\alpha(A \Rightarrow C) \quad (14)$$

Definition 6 *The confidence of the fuzzy association rule $R : A \Rightarrow C$, RL-Conf$(A \Rightarrow C)$, is given by*

$$RLConf(A \Rightarrow C) = \sum_{(A \Rightarrow C)_\alpha \in \Omega_R} m_R((A \Rightarrow C)_\alpha) \cdot Conf_\alpha(A \Rightarrow C) \quad (15)$$

Definition 7 *The certainty factor of the fuzzy association rule $R : A \Rightarrow C$, RL-CF$(A \Rightarrow C)$, is given by*

$$RLCF(A \Rightarrow C) = \sum_{(A \Rightarrow C)_\alpha \in \Omega_R} m_R((A \Rightarrow C)_\alpha) \cdot CF_\alpha(A \Rightarrow C) \quad (16)$$

These measures, RL-Supp$(A \Rightarrow C)$, RL-Conf$(A \Rightarrow C)$ and RL-CF$(A \Rightarrow C)$, can be viewed as the summarization of, respectively, the support, confidence, and certainty factor of every crisp association rule $A \Rightarrow C$ obtained at each restriction level. In fact, if the considered association rule has no negated items involved, the following propositions (as generalizations of the one proposed in [21]) hold.

Proposition 1 *The RL-support of the fuzzy association rule $A \Rightarrow C$, RL-Supp$(A \Rightarrow C)$, is equal to the support of the rule as computed in [16], using the method GD proposed in [22].*

Proposition 2 *The RL-confidence of the fuzzy association rule $A \Rightarrow C$, RL-Conf$(A \Rightarrow C)$, is equal to the confidence of the rule as computed in [16], using the method GD proposed in [22].*

Let us remark that these same propositions for RL-Supp and RL-Conf are not applicable for RL-CF since, again according to [16], the certainty factor for a fuzzy association rule was still computed in the same way as for a crisp one. Thus, we are defining a new measure of accuracy for fuzzy association rules, via restriction levels. The derived properties for this new measure will be the object of study in a separate paper.

Finally, let us notice that, due to the followed RL-representation, and opposite to the fuzzy sets case, it is not possible to find itemsets containing both items i and $\neg i$, which not only is desirable but also sounds logical.

5 Experiments

5.1 Implementation aspects

In our implementation, we start from the same Apriori-based algorithm proposed in [16], with the exception that we must compute the measures of interest and accuracy in terms of evidence in restriction levels, following definitions 5, 6, and 7.

Additionally, we must take into account that now we consider also absent/negated items, which result in a larger number of itemsets to be computed. I.e., if the original Apriori algorithm [1] takes into account up to 2^n itemsets, in our proposal we must consider all the combinations of items and negated items (except those impossible cases containing both i and $\neg i$), resulting in a total number of $(n+1) \cdot 2^n \approx n \cdot 2^n$ itemsets. In each iteration, we must count the occurrences of items in the set of transactions at each restriction level, but also the absences of them. This results in that when considering combinations of items and negated items, we must correctly compute, for each restriction level, the correct support for each itemset. I.e., $supp(\{\neg i_1, \neg i_2\}) = 1 - (supp(i_1) + supp(i_2) - supp(\{i_1, i_2\}))$, $supp(\{i_1, \neg i_2\}) = supp(i_1) - supp(\{i_1, i_2\})$, and so on. One important remark is that we need to keep all the generated itemsets, in order to have access to their support values when necessary. This fact increases the amount of necessary memory, and delays the threshold pruning (by minimum support, etc.) until the second stage of the algorithm (association rule extraction).

According to this, the final efficiency of the algorithm is approximately that of classic Apriori algorithm, but multiplied by $(n+1)$ (due to negated items consideration) and by k (being k the number of restriction levels considered). One of our pending tasks will be the study and development of a more efficient algorithm. In this sense, one choice is that of parallelize the overall process. Since each one of the restriction levels is a crisp set, we can reduce the problem to that of concurrently extract crisp association rules at each level, and then aggregate the obtained measures. We will address this aspect in a future paper.

5.2 Experimentation results

In order to test the resulting algorithm, we first performed our experiments over artificially generated set of transactions. We randomly generated a set of transactions, involving 100 items (and their negations), considered $k = 10$ restriction levels, and obtained a total number of 39600 fuzzy association rules. For space saving purposes, we restrict the number of items in a rule to 2, 1 in the antecendent and 1 in the consequent. In forthcoming papers, we will afford the study of more complex rules.

Each fuzzy association rule has a related set of (crisp) association rules, present in each restriction level. Table 2 shows an example subset of the obtained rules, after establishing a threshold of minimum support and minimum certainty factor of 0.5. Note that, after establishing these thresholds, a rule

ISBN: 978-989-95079-6-8

Table 1: Example of extracted fuzzy association rules

$id\#$	Rule	Evidence	RL-Supp	RL-CF
#39283	$\neg i_{90} \Rightarrow \neg i_{92}$	0.4	0.31	0.31
#39295	$\neg i_{90} \Rightarrow \neg i_{95}$	0.4	0.30	0.25
#39347	$\neg i_{91} \Rightarrow \neg i_{92}$	0.4	0.31	0.31
#39359	$\neg i_{91} \Rightarrow \neg i_{95}$	0.1	0.05	0.05
#39463	$\neg i_{93} \Rightarrow \neg i_{95}$	0.5	0.38	0.36
#39503	$\neg i_{94} \Rightarrow \neg i_{95}$	0.5	0.38	0.36
#39519	$\neg i_{94} \Rightarrow \neg i_{99}$	0.6	0.45	0.38
#39591	$\neg i_{97} \Rightarrow \neg i_{99}$	0.5	0.38	0.37
#39599	$\neg i_{98} \Rightarrow \neg i_{99}$	0.5	0.38	0.37

may not appear in all restriction levels. For sake of simplicity and space saving matters, only the rule number is shown. See Table 1 for a more complete description of some of the rules.

Applying definitions 5 and 7, we can summarize the support and certainty factor of the rules in the restriction levels, and compute the measures for the fuzzy association rules. Table 1 shows some of the obtained fuzzy association rules. The rule numbers correspond to the referred ones in previous Table 2.

In addition, another interesting issue of our proposed methodology is that, in terms of the restriction levels evidence, we can summarize also the obtained set of fuzzy association rules. Actually, we can reduce the number of rules present at each restriction level, by discarding those with low support and/or low certainty factor. After this pruning, we can compute the associated probability distribution (see definition 10) for each rule, and then interpret the resulting set of association rules in terms of evidence. For example, the association rules shown in Table 2 could be summarize according to their evidence in the following expression,

$$\begin{aligned} Ruleset = \{\#39519\}/0.6 + \\ \{\#39463, \#39503, \#39591, \#39599\}/0.5 + \\ \{\#39283, \#39295, \#39347\}/0.4 \end{aligned} \quad (17)$$

Again, for space saving purposes, we only refer to the rules by its rule number. Let us remark that this expression is not actually a fuzzy set, but still can be very helpful in order to interpret the set of results, as we relate a relevance degree, the basic probability assignment, to each rule or set of rules.

Finally, we applied our proposed methodology on real data, obtained from a database containing soil information (see [23] for a more complete description). We reduce the set of data to those attributes modeled as fuzzy quantities ($Averagerainfall$, $Altitude$, $Depth$, PH, and attributes describing the soil chemical composition). A set of linguistic labels $\{Low, Medium, High\}$ was defined for every attribute. Table 3 shows some of the obtained fuzzy association rules, all having $Evidence = 1.0$, restraining the thresholds of minimum support to 0.5 and minimum certainty factor to 0.7, for space saving purposes.

Let us notice that considering absent items allows us to improve the semantics of the rules. I.e., considering item $\neg Avg.Rainfall = Low$ allows us to represent both $Avg.Rainfall = Medium$ and $Avg.Rainfall = High$ concepts. This aspect will be addressed in more detail in a forthcoming paper.

Table 3: Example of fuzzy association rules on real data

Rule	RL-Supp	RL-CF
$PH = High \Rightarrow \neg Avg.Rain. = Low$	0.616	0.97
$PH = Low \Rightarrow \neg Avg.Rain. = Medium$	0.616	0.97
$PH = High \Rightarrow \neg Avg.Rain. = Medium$	0.616	0.97
$PH = High \Rightarrow \neg Avg.Rain. = High$	0.616	0.97
$PH = Medium \Rightarrow \neg Avg.Rain. = Low$	0.616	0.97
$PH = Low \Rightarrow \neg Avg.Rain. = Low$	0.616	0.97
$PH = Medium \Rightarrow \neg Avg.Rain. = Medium$	0.616	0.97
$PH = Low \Rightarrow \neg Avg.Rain. = High$	0.616	0.97
$PH = Medium \Rightarrow \neg Avg.Rain. = High$	0.616	0.97
$\neg Alt. = Low \Rightarrow \neg Avg.Rain. = High$	0.810	0.83
$\neg Alt. = Medium \Rightarrow \neg Avg.Rain. = High$	0.810	0.83
$\neg Alt. = High \Rightarrow \neg Avg.Rain. = Low$	0.810	0.83
$\neg Alt. = High \Rightarrow \neg Avg.Rain. = High$	0.810	0.83
$\neg Alt. = Medium \Rightarrow \neg Avg.Rain. = Low$	0.810	0.83
$\neg Alt. = High \Rightarrow \neg Avg.Rain. = Medium$	0.810	0.83
$\neg Alt. = Low \Rightarrow \neg Avg.Rain. = Medium$	0.810	0.83
$\neg Alt. = Medium \Rightarrow \neg Avg.Rain. = Medium$	0.810	0.83
$\neg Alt. = Low \Rightarrow \neg Avg.Rain. = Low$	0.810	0.83

6 Concluding remarks

We have introduced a new methodology to represent fuzzy association rules in terms of RL-sets, as an alternative representation to that of fuzzy sets. This methodology offers some new advantages as, for example, it allows us to properly consider issues as fuzzy association rules involving absent items. As far as we know, ours is a novel methodology, combining both presence and absence of items with imprecision and uncertainty management. Considering both presence and absence of items can lead us to discern possible conflicts between items or detect when a set of items complements another.

In addition, it also offers a method of summarizing the set of obtained rules, representing them in a format similar to that of fuzzy sets, and thus allowing us to help the user to discern the relevance of a given rule or set of rules, along with the usual measures of interest and accuracy. With respect to these measures, we have seen how the resulting measures values are the same to those obtained in the fuzzy set approach. In this sense, our approach is still coherent with respect to previous ones. Nevertheless, in the general case, and specially in the case of the measure of certainty factor, new challenges appears abroad, and the derived properties for these new definitions for measuring the accuracy in rules will be approached in a future paper.

Finally, we have extended an existing Apriori-based algorithm in order to consider a fuzzy association rule as the aggregation (RL-representation) of the related association rules

ISBN: 978-989-95079-6-8

Table 2: Example of (crisp) association rules distributed through restriction levels

RL \ *Rule*	#39283	#39295	#39347	#39359	#39463	#39503	#39519	#39591	#39599
1.0	×		×		×	×	×	×	×
0.9	×	×	×		×	×	×	×	×
0.8	×	×	×		×	×	×	×	×
0.7	×	×	×		×	×	×	×	×
0.6		×			×	×	×	×	×
0.5							×		
0.4				×					

at each restriction level. As the overall performance is affected by the higher number of itemsets to be taken into account, one pending task will be the optimization of this algorithm in order to improve its efficiency.

Acknowledgment

This work was developed as part of the research project TIN2006-07262.

References

[1] R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. In *Proc. Of the 1993 ACM SIGMOD Conference*, pages 207–216, 1993.

[2] Craig Silverstein, Sergey Brin, and Rajeev Motwani. Beyond market baskets: Generalizing association rules to dependence rules. *Data Min. Knowl. Discov*, 2(1):39–68, 1998.

[3] Maria-Luiza Antonie and Osmar R. Zaane. Mining positive and negative association rules: an approach for confined rules. In *In European PKDD Conf*, pages 27–38, 2004.

[4] Xindong Wu, Chengqi Zhang, and Shichao Zhang. Efficient mining of both positive and negative association rules. *ACM Transactions on Information Systems*, 22(3):381–405, July 2004.

[5] Wei-Guang Teng, Ming-Jyh Hsieh, and Ming-Syan Chen. On the mining of substitution rules for statistically dependent items. In *ICDM*, pages 442–449. IEEE Computer Society, 2002.

[6] Ashok Savasere, Edward Omiecinski, and Shamkant Navathe. Mining for strong negative associations in a large database of customer transactions. pages 494–502. IEEE Computer Society, 1998.

[7] Xiaohui Yuan, Bill P. Buckles, Zhaoshan Yuan, and Jian Zhang. Mining negative association rules. *Computers and Communications, IEEE Symposium on*, 0:623, 2002.

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

[9] W.H. Au and K.C.C. Chan. Mining fuzzy association rules. In *Proc. Of 6th Int. Conf. On Information and Knowledge Management. Las Vegas, NV, USA*, pages 209–215, 1997.

[10] A.W.C. Fu, M.H. Wong, S.C. Sze, W.C. Wong, W.L. Wong, and W.K. Yu. Finding fuzzy sets for the mining of fuzzy association rules for numerical attributes. In *Proc. Int. Symp. On Intelligent Data Engineering and Learning (IDEAL'98), Hong Kong*, pages 263–268, 1998.

[11] S. Ben-Yahia and A. Jaoua. A top-down approach for mining fuzzy association rules. In *Proc. 8th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU'2000*, pages 952–959, 2000.

[12] Shu-Yue-J, Tsang-E, Yenng-D, and Daming-Shi. Mining fuzzy association rules with weighted items. In *Proc. IEEE Int. Conf. On Systems, Man and Cybernetics*, 2000.

[13] D. Dubois, H. Prade, and T. Sudkamp. A discussion of indices for the evaluation of fuzzy associations in relational databases. In T. Bilgic, B. De Baets, and O. Kaynak, editors, *Fuzzy Sets and Systems - Proceedings IFSA'03, LNAI 2715*, pages 111–118. Springer, 2003.

[14] D. Sánchez, M. Delgado, and M.A. Vila. A restriction level approach to the representation of imprecise properties. In *IPMU 2008*, pages 153–159, 2008.

[15] M. Delgado, M.D. Ruiz, and D. Sánchez. A restriction level approach for the representation and evaluation of fuzzy association rules. Submitted to IFSA'2009, 2009.

[16] M. Delgado, N. Marín, D. Sánchez, and M.A. Vila. fuzzy association rules: General model and applications. *IEEE Transactions on Fuzzy Systems 11(2)*, pages 214–225, 2003.

[17] F. Berzal, I. Blanco, D. Sánchez, and M.A. Vila. Measuring the accuracy and interest of association rules: A new framework. *Intelligent Data Analysis*, 6:221–235, 2002.

[18] E. Shortliffe and B. Buchanan. A model of inexact reasoning in medicine. *Mathematical Biosciences*, 23:351–379, 1975.

[19] L.A. Zadeh. A computational approach to fuzzy quantifiers in natural languages. *Computing and Mathematics with Applications*, 9(1):149–184, 1983.

[20] D. Sánchez, M. Delgado, and M.A. Vila. RL-numbers: An alternative to fuzzy numbers for the representation of imprecise quantities. In *FUZZ-IEEE 2008*, pages 2058–2065, 2008.

[21] D. Sánchez, Delgado M., and M.A. Vila. Fuzzy quantification using restriction levels. Submitted to WILF'2009, 2009.

[22] M. Delgado, D. Sánchez, and M.A. Vila. Fuzzy cardinality based evaluation of quantified sentences. *International Journal of Approximate Reasoning*, 23:23–66, 2000.

[23] J. Calero, G. Delgado, M. Sánchez-Marañón, D. Sanchez, MA Vila, and J. Serrano. An experience in management of imprecise soil databases by means of fuzzy association rules and fuzzy approximate dependencies. In *Proceedings of the 6th international conference on enterprise information systems, ICEIS*, volume 4, pages 138–146. Springer, 2004.

ISBN: 978-989-95079-6-8

Towards an interpretation of the connectivity analysis in the frame of fuzzy equivalence relations for medical diagnoses

Tatiana Kiseliova

Institute of Computer Sciences
Faculty of Exact and Natural Sciences
Javakhishvili Tbilisi State University
Email: t.kiseliova@web.de

Abstract— *Connectivity analysis methodology is suitable to find representative symptoms of a disease. This methodology describes connections between symptoms in particular way and then chooses the group of symptoms that have the high level of connection, or in other words, have strong interconnections between elements of a group. In this paper we investigate the analogy between connectivity analysis and cluster analysis based on fuzzy equivalence relations. A comparison of two approaches, one of which has strong theoretical background (cluster analysis based on fuzzy equivalence relations) and more practically oriented connectivity analysis assures more convincing and accurate connectivity analysis from one side and applicability of fuzzy equivalence relations for medical diagnoses from another. Connectivity analysis, as shown in the paper, is one of the clustering methods, can be used in many applications where feature selection and extraction problem is considered, in particular, in pattern recognition and image processing. The results of the comparison are demonstrated on the examples.*

Keywords— connectivity, equivalence relations, fuzzy cluster analysis.

1 Introduction

When speaking about the physician intuition one means the ability of a doctor to establish a diagnosis for a patient. This process can be separated into different stages. One of them, considered in this paper, distinguishes an experienced doctor from a beginner: the expert uses a small amount of observations to make a conclusion about a patient's state. Calling these observations representative symptoms one highlights their high discrimination power between this and other diseases, their best characterization for a disease. "The best characterization" means: if a patient has this representative group of symptoms, his diagnosis is almost sure. Fuzzy notations such as "small amount", "almost sure" show - as was already mentioned in many scientific papers devoted to applications of fuzzy logic in medical domains - that definitions, reasoning and conclusions in medicine are usually done with uncertainty. And a task - to find the representative symptoms - is not an exception.

For reasons of simplification we use *symptoms* as a general concept for clinical parameters, different manifestations (signal, clinical), meanwhile *diseases* include disorders, different diagnostic hypotheses.

The problem of finding the representative symptoms can be solved within a computer-assisted medical diagnosis system. Ideally, an established set of representative symptoms for a particular disease allows to only partly examine a patient suspected for this disease, that, in turn, reduces costs, time, etc. Different groups of representative symptoms can be useful at the different stages of a chronic disease. Representative symptoms can help to establish an overall risk for a patient attacked from a particular disaster. For example, in life insurance medicine [16] the mortality of applicants within the period of insurance is assessed on the basis of present risk factors. Another possible application of representative symptoms is an optimization of questionaries, e.g., in screening [5].

There are at least two ways to get the representative symptoms: to ask an experienced physician or to use some methods to extract this set from the available information.

As was already mentioned, symptoms cannot normally be described in simple "black and white" terms with "yes/no" or "present/not present" answers. An expert seeking to comprehensively describe the multiple influencing factors would very quickly reach the limit of his capacity [7]. For example, if an expert should estimate a group of three symptoms, each of which has three values, he will face 27 possible states. If the number of symptoms and the number of values rise to five, the number of states increase to 3125. Moreover, some disease categories overlap. For example, no consensus exists on clinical or epidemiologic measures that can be used to classify patients based on asthma symptoms [17]. Due to the highly variable nature of asthma, classification of patients into mild, moderate and severe disease categories is necessarily imperfect. That categories may overlap means, that trying to find the representative symptoms for three categories of a disease (mild asthma, moderate asthma and severe asthma) we do not have a unique description for each of them. The USA National Asthma Education Program of guidelines suggest symptoms (wheeze, dyspnoea, cough, sputum production, allergy characteristics, etc.), the degree of airflow obstruction and frequency of use of oral glucocorticoids (never, infrequently for attacks, frequently for attacks, and daily use) to validate three asthma categories. Due to the symptom-disease-patient terminology proposed by us the representative group of symptoms could be chosen from *symptoms* (the medication use, spirometry, current symptoms) for three *diseases* (indicators of asthma severity) for this example. Therefore it is not an easy task to get representative symptoms from an expert for building a computer-assisted medical system. A specialist uses his unconscious knowledge, due to the professional intuition and acumen.

To rely on data from the patient records only (or other sources containing information about patients), using statistical methods, seems also to be not a doubtless way. Widely recognized drawbacks of statistical estimations is the demand on high amounts of data. In other situations, conclusions de-

ISBN: 978-989-95079-6-8

duced from records do not correspond to results from other sources. For example, as was reported in [17], a severity of asthma induced from oral glucocorticoids use and spirometry does not correlate with current asthma symptoms from medical records.

A combination of approaches is a matter of investigations since years [2]. Especially since the theory of fuzzy sets was proven to be useful in modeling uncertainties, in particular, in medicine. Expert knowledge is composed of the evaluations of the observed symptoms (the influencing factors) and the rules for the combinations of these evaluations. The expert formulates and describes his decision in a linguistic form in order to render it comprehensible to a layperson [16, 20]. Often, he/she demonstrates only certain key values of the influencing factors, the representative symptoms.

The complexity of a method partially depends on the initial information available. For example, if we have a collection of rules, where all symptoms and combinations of symptoms are already estimated by experts as was done in [18] for pulmonary tuberculosis with some degree of confirmation and exclusion, the task to find the representative symptoms may be easily solved. Experts are asked to establish thresholds for an exclusion and a confirmation of symptoms and this will allow to classify the symptoms to the corresponding representative groups.

The medical records is a point of departure for connectivity analysis as well, but values in these records represent some degree of expressiveness of a symptom, or the degree of compatibility between patients and symptoms. For example, for "overweight" the expert may put a number between 0 and 1 to reflect his opinion about this patient's symptom. The expert may know the exact weight of a patient or not, his estimation 0 shows that the patient has no overweight, 1 - the patient is very thick, 0.2 -the patient is almost in norm. The "ruddy complexion" is another example for the scale representation, e.g., $0.2, 0.4, 0.8$ would mean different degrees of a skin face color estimated by an expert.

The connectivity analysis has been intensively applied, in particular, in medicine, management, geology [4, 10, 11, 12, 13, 22]. In most cases its practical usefulness was proven by examples. But in general, the connectivity analysis can be considered as a part of pattern recognition problem, in particular, cluster analysis. Till now such relations were not established. Our intention in this paper is to show that there exists an analogy between the connectivity analysis and fuzzy cluster analysis based on fuzzy equivalence relations. This comparison (or, in some sense, an interpretation of connectivity analysis in terms of well-established cluster analysis) brings more insights into the connectivity analysis which is practically used but has some weaknesses in the theoretical background from one side and applicability of fuzzy equivalence relations for medical diagnoses from another.

To realize our intentions, the paper is organized as follows. In the next section we describe what we understand under connectivity analysis and how it is applied to the representative symptoms mining problem. Section 3 describes elements of a fuzzy cluster analysis based on fuzzy equivalence relations. In Section 4 the analogy between two approaches are considered and the paper is concluded by final remarks.

2 Connectivity between symptoms

Assume from the database of already diagnosed patients, *suffering from a particular disease*, the table $R_{SP} = \{r_{ij}\}$, $i = 1 \dots n$, $j = 1 \dots m$ is obtained where only presence or absence of symptoms are ticked. For example, in Table 1 information about 5 patients who have been suffering from a particular disease with 4 corresponding predefined symptoms is presented. Compared with a real practical situation, the description of e.g., *pulmonary tuberculosis (PT)* [18] consists of 30 symptoms.

Table 1: An initial table for a symptom-patient connection

	p_1	p_2	p_3	p_4	p_5
s_1	1	0	1	0	0
s_2	1	1	0	1	0
s_3	1	1	1	1	1
s_4	0	1	0	1	1

2.1 The positive level of connection

Assume, Table 1 defines the relation of connectivity between a set of patients $P = \{p_1, p_2, p_3, p_4, p_5\}$ and a set of symptoms $S = \{s_1, s_2, s_3, s_4\}$. $r_{ij} = 1$ denotes that person p_j has the symptom s_i and $r_{ik} = 0$ denotes that person p_k does not have it.

Dealing with these data, the intuitive proposal to get a group of representative symptoms is to highlight those symptoms that often meet under description of a patients' condition. For example, from Table 1 it can be seen that symptom s_3 is present by all patients, symptoms s_1 and s_2 have in common only p_1. Speaking language of connectivity, the representative symptoms can be naturally chosen based on the level of connection among symptoms that is equal to the number of common patients affected by those symptoms. For example, symptoms s_1 and s_2 have in common only p_1 and hence they are connected at level 1, meanwhile symptoms s_2 and s_3 have in common patients p_1, p_2 and p_4 , i.e., they are connected at level 3. Formula (1) describes pairwise connectivities between symptoms.

$$Q_s = R_{SP} R_{SP}^{\mathsf{T}} \tag{1}$$

In (1) ordinary multiplication of matrices is used.

The result of (1) is matrix (2). Clearly, Q_s is diagonal symmetric and can be represented in upper triangular form: the order of symptoms is not important.

$$Q_s = \begin{pmatrix} 2 & 1 & 2 & 0 \\ 1 & 3 & 3 & 2 \\ 2 & 3 & 5 & 3 \\ 0 & 2 & 3 & 3 \end{pmatrix} \tag{2}$$

Connectivity results can be represented as shown in Table 2.

0-connection is not considered. The symptom s_3 can be taken as the representative one, but the groups of symptoms are of interest from the physician point of view: one symptom, for instance, a cough, can be present due to many diseases, and cannot be distinguishable for them.

ISBN: 978-989-95079-6-8

Table 2: Pairwise connections from matrix (2)

Levels of connectivity, q	Groups of symptoms
5	$\{s_3\}$
3	$\{s_2\}\{s_2, s_3\}\{s_3, s_4\}, \{s_4\}$
2	$\{s_1\}\{s_1, s_3\}\{s_2, s_4\}$
1	$\{s_1, s_2\}$

Sets $\{s_2, s_3\}$ or $\{s_3, s_4\}$ can be considered as appropriate ones. They have, however, in common one element, s_3. A fusion of these two sets into $\{s_2, s_3, s_4\}$ would mean, that the same number of patients (i.e., 3) have in common subsets of this set.

Therefore, let us transform Table 2 into Table 3 due to the following observations:

1. if n patients have in common a symptom (or symptoms), $n-1$ patients have this symptom(s), too. The same is for $n-1$, $n-2$, etc. This property we call inheritance.
2. if the same symptoms belong to different groups at the same level of connectivity, these groups can be fused.

Table 3: After fusion and inheritance in Table 2

Levels of connectivity, q	Groups of symptoms
5	$\{s_3\}$
4	$\{s_3\}$
3	$\{s_2, s_3, s_4\}$
2	$\{s_1, s_2, s_3, s_4\}$
1	$\{s_1, s_2, s_3, s_4\}$

Information from Table 3 can be interpreted. For instance, symptoms $\{s_2, s_3, s_4\}$ are more representative for the considered disease, because they are at a high level of connectivity i.e., have in common many patients and the cardinality of this set is not small. This intuitive procedure can lead, for example, to similar groups of representative symptoms, i.e., at the same level, two groups and the same number of different symptoms in each group. In this case both groups can be considered as representative ones or a consultation with a physician is supposed.

So far so good, but even in this example part of the information used is lost: for instance, the absence of the symptoms are not taken into account. In Table 4,

Table 4: s_1 is more like s_2 than s_3

	p_1	p_2	p_3	p_4
s_1	1	0	1	1
s_2	1	0	1	1
s_3	1	1	1	1

s_1 and s_3 are connected (via p_1, p_3 and p_4) at the same level ($q = 3$) as s_1 and s_2 (via p_1, p_3 and p_4). But obviously, s_1 is more like s_2 than s_3.

2.2 *The negative level of connection*

Due to the tradition of expert systems in medical diagnoses [3, 6, 8, 18, 19, 21] it is important to consider not only an occurrence of symptoms, but a non-occurrence as well.

Let us now get information from the non-occurrence of symptoms in Table 1. For each pair of symptoms the number of patients, that do *not* have in common these symptoms is calculated. Formally, formula (1) can be used here, where matrix R_{SP} is substituted by $\Omega - R_{SP}$, where Ω is a matrix with all elements unity. The results of negative connectivity is represented in Table 5, where q^- are levels of connectivity.

Table 5: Pairwise connections of symptoms that patients have *no* in common from Table 2

q^-	Groups of symptoms
3	$\{s_1\}$
2	$\{s_2\}, \{s_4\}$
1	$\{s_1, s_2\}\{s_2, s_4\}$
0	$\{s_3\}, \{s_3, s_4\}, \{s_2, s_3\}, \{s_1, s_3\}, \{s_1, s_4\}$

The results of this table can be interpreted. For example, the lowest row shows, that there are no patients, that do not have in common $\{s_3\}$, or $\{s_3, s_4\}$, or $\{s_2, s_3\}$, or $\{s_1, s_3\}$, or $\{s_1, s_4\}$, i.e., all patients have at least one of these sets in common. This leads to the assumption, that these groups of symptoms can be considered as candidates to be representative ones. $\{s_1\}$ does not occur by 3 patients, and therefore $\{s_1\}$ can be considered as a representative non-occurrence symptom for the particular disease (if we assume, of course, that one symptom can be representative for non-disease; this case, as a corresponding one from the positive connectivity should be discussed with an expert).

Intuitively it is also clear that if $\{s_1\}$ does not occur by each of three patients, it does not occur by two or one of them. The procedures of inheritance and fusion are the same as in the case of positive connectivities.

The final set of representative symptoms can be differently established. A possible way is to combine representative symptoms for a disease and for non this disease. For example, if a patient has $\{s_2, s_3, s_4\}$ and does not have $\{s_1\}$ we can conclude that the patient has this disease.

2.3 *Some remarks, concerning connectivity analysis*

We did not consider in the previous sections cases when the elements of the matrices R_{SP} take the values from $[0, 1]$ or, moreover, are represented by linguistic terms. In this paper we restrict ourselves to the "crisp" cases to concentrate on the explanation of the idea of the connectivity analysis to demonstrate a way to find similarities between symptoms due to their presence by patients. If, however, the elements of R_{SP} are numbers from $[0, 1]$, they are interpreted as a strength of connection (or compatibility) between a patient and a disease.Q_s is a composition of fuzzy relations. Elements of Q_s can be interpreted as a pairwise strength between symptoms for patients.

It is easy to see that absolute values of the connectivity levels are not decisive for choice of the group of representative

ISBN: 978-989-95079-6-8

symptoms. A corresponding ordinal scale can bear the necessary information for connectivity. It means that for particular calculations we can simply linearly transform obtained connectivity levels into corresponding numbers from $[0, 1]$.

Above we have used a fairly fuzzy procedure to find the representative symptoms from positive and negative connectivities. The main demands are "high level of connection" and rather "wide" set of symptoms, connected at this level. In some sense, more formal algorithm would be preferable.

In the next section we consider if it is possible to formulate the connectivity analysis by similarity relations and corresponding fuzzy cluster analysis to assure more convincing and accurate connectivity analysis.

This idea appears by an intuitive analogy between levels of connections and α-cuts of fuzzy relations. Let us first recall several definitions used to build such analogy.

3 Cluster analysis based on fuzzy equivalence relations

It is known that every fuzzy equivalence relation (sometime called a similarity relation) induces a crisp partition of its α-cuts. And therefore, fuzzy clustering problem can be viewed as the problem of identifying the appropriate fuzzy equivalence relation on given data.

Definition 3.1 *Let R be a fuzzy relation, $R : X \times Y \to [0, 1]$, i.e., $R = \{((x, y), R(x, y))|(x, y) \in X \times Y\}$, the α-cut matrix ${}^{\alpha}R$ is denoted by*

$${}^{\alpha}R = \{(((x, y), {}^{\alpha}R(x, y))|{}^{\alpha}R(x, y) = 1 \text{ if } R(x, y) \geq \alpha;$$
$${}^{\alpha}R(x, y) = 0 \text{ if } R(x, y) < \alpha, (x, y) \in X \times Y, \alpha \in [0, 1]\}$$

Definition 3.2 *Binary fuzzy relation $R : X \times X \to [0, 1]$ is a fuzzy equivalence relation iff it is reflexive, i. e., $R(x,\ x) = 1$; symmetric, i. e., $R(x,\ y) = R(y,\ x)$, and max-min transitive: $R^{(2)} = R \circ R \subset R$ or, more explicitly $R(x,\ z) \geq \max_y \left\{ \min_{x,z}\{R(x,\ y),\ R(y,\ z)\} \right\}$, $\forall x, y, z \in X$*

For practical tasks it is easier first to build a fuzzy compatibility relation and then calculate a transitive closure of this compatibility relation and this way to complete identification of a fuzzy equivalence relation.

Definition 3.3 *A fuzzy relation R on $X \times X$ is called a fuzzy compatibility relation if it satisfies reflexive and symmetric conditions.*

Definition 3.4 *The transitive closure R_T of a fuzzy relation R is defined as the relation that is transitive, contains R ($R_T \supset R$) and has the smallest possible membership grades.*

Theorem 3.1 *[14] Let R be a fuzzy compatibility relation on a finite universal set X with $|X| = n$, then the $\max - \min$ transitive closure of R is the relation $R^{(n-1)}$.*

4 Connectivity analysis in the frame of fuzzy equivalence relations

Ideally, if Q_s (see (1)) would be a similarity relation one can easy see that levels of connections correspond to α-cuts. Thus, the analogy between connectivity analysis and a fuzzy clustering based on fuzzy equivalence relations is established.

NormallyQ_s is symmetric, but its reflexivity and transitivity are questionable. To reach these two properties one can do as follows.

First, as was already told in the section 2.3, we can substitute the elements of Q_s by the corresponding ones from interval $[0, 1]$. Due to the definition of ε-reflexivity, i.e. $R(x,\ x) \geq \varepsilon\ x \in X$, $0 < \varepsilon < 1$ [14], Q_s is ε-reflexive. Thus, Q_s is a fuzzy compatibility relation.

Transitive closure of Q_s can be find due to the following procedure [15] based on the results of the Theorem 3.1: find the transitive closure Q_T of fuzzy compatibility relation Q_s.

> Step 1: Calculate $Q^{(2)}$ if $Q^{(2)} = Q_s$ or $Q^{(2)} \subset Q_s$, then transitive closure $Q_T = Q_s$ and stop. Otherwise, $k = 2$, go to step 2.
>
> Step 2 : If $2^k \geq n - 1$, then $Q_T = Q^{(n-1)}$ and stop. Otherwise, calculate $Q^{(2^k)} = Q^{(2^{k-1})} \circ Q^{(2^{k-1})}$, if $Q^{(2^k)} = Q^{(2^{k-1})}$, then transitive closure $Q_T = Q^{(2^k)}$ and stop. Otherwise, go to step 3.
>
> Step 3: $k = k + 1$, go to step 2.

Due to this procedure a transitive closure of Q_s is as follows (mention a reverse transformation of the previous scale):

$$Q_T = \begin{pmatrix} 2 & 2 & 2 & 2 \\ 2 & 3 & 3 & 3 \\ 2 & 3 & 5 & 3 \\ 2 & 3 & 3 & 3 \end{pmatrix} \tag{3}$$

Following Definition 3.1 the relation (3) induces the partitions of its α-cuts represented in the Table 6:

Table 6:

α-cuts	Groups of symptoms
5	$\{s_3\}$
3	$\{s_2, s_3, s_4\}$
2	$\{s_1, s_2, s_3, s_4\}$
1	$\{s_1, s_2, s_3, s_4\}$

It can be easy see, that these results coincide with results of connectivity analysis represented in the Table 3. Relation (3) represents a transitive closure of Q_s, and the elements of this relation differ from corresponding elements of Q_s. A similar procedure can be done for "negative connectivity" analogy.

Notice, that recently there exist other methods, that transform the initial fuzzy binary relation into reflexive, symmetric and T-transitive fuzzy binary relation, i.e., a similarity relation, that contains the initial relation [9].

Although results from Table 6 satisfy our initial intuition, "the transitive closure method carries a number of major problems, such as the need of storage and computer time required; ... there is no way of controlling the distortion that this method produce on the data sample"[23].

Several possibilities exist to find a more efficient way than this transitive closure method does. For instance, an application of the representation theorem [23] no longer requires reflexive and symmetric fuzzy relation as initial data.

ISBN: 978-989-95079-6-8

Thus, to build a similarity relation from an initial fuzzy relation, can be done differently. Our approach, based on the ε-reflexivity, and transitive closure, return a similarity relation, which α-cuts coincides with group of symptoms from the connectivity analysis. It was checked for data considered in different applications [4, 12, 13, 22] and gives the same results. This is the approach used thus far to establish the validity of the proposed analogy between connectivity analysis and the fuzzy cluster analysis based on fuzzy equivalence relations. Notice that this analogy was found for low dimensional data, large-scale data is a subject of future investigations.

5 Concluding remarks

Let us summarize what we have done. We have shown how connectivity analysis can work in the frame of fuzzy equivalence relation. This means, in particular, that due to the above described analogy, levels of connection in the connectivity analysis can be represented as α-cuts. In turn, the methodology to build clusters based on the α-cuts from fuzzy relations is already established. Therefore, the connectivity analysis can be considered as a fuzzy clustering problem based on the fuzzy compatibility relations.

The question is how to obtain a compatibility relation from initial table (see, for instance, Table 1). In our discussions we used Q_S. As was already told, the connectivity analysis belongs to the family of cluster methods, and particularly, it can be considered as one of the hierarchical methods [1]. Therefore other dissimilarity measures between patterns and features (patients and symptoms) can be used and the problem to find the representative symptoms becomes a problem of the feature selection and feature extraction problem in the cluster analysis. A fuzzy compatibility relation can be defined in terms of an appropriate distance function of, e.g., the Minkowski class [1]. The question is in interpretation of this relation. Therefore the next investigation can be dedicated to the construction of Q_S that allows to inherit useful information from the initial table R_{SP} to build representative symptoms.

References

[1] J. Abonyi and B. Feil. *Cluster Analysis for Data Mining and System Identification*. Birkhäuser Verlag AG, 2007.

[2] K.-P. Adlassnig and G. Kolarz. CADIAG-II: computer-assisted medical diagnosis using fuzzy subsets. In M.M. Gupta and E. Sanchez, editors, *Approximate Reasoning in Decision Analysis*, pages 495–505. North Holland Publishing Company, 1982.

[3] K. Bögl, H. Leitich, G. Kolousek, G. Rothenfluh, and K.-P. Adlassnig. Clinical data interpretation in MedFrame/CADIAG-4 using fuzzy sets. *Biomedical Engineering: Applications, Basis and Communications*, 8(6):488–495, 1996.

[4] F. Criado, T. Gachechiladze, N. Jorjiashvili, Z. Khvedelidze, H. Meladze, J. M. Saacutenchez, G. Sirbiladze, and G. Tsertsvadze. Theory of connectivity and apportionment of representative chains in the problem of decision-making concerning earthquake possibility. *International Journal of General Systems*, 32(2):103–121, 2003.

[5] F. Georgsson and P. Eklund. An identification of handling uncertainties within medical screening: a case study within screening for the breast cancer. In *Fuzzy and Neuro-Fuzzy Systems in Medicine*, pages 161–188. CRC Press, Boca Raton, FL, 1998.

[6] L. Godo, R. López de Mántaras, C. Sierra, and A. Verdaguer. MILORD: The architecture and management of linguistically expressed uncertainty. In D. Dubois, H. Prade, and R.R. Yager, editors, *Fuzzy Information Engineering: A Guided Tour of Applications*, pages 357–387. Wiley Computing Publishing, John Wiley and Sons, Inc, 1990.

[7] P.-J. Horgby, R. Lohse, and N.-A. Sittaro. Fuzzy underwriting: an approach of fuzzy logic to medical underwriting. *Actuarial Practice*, 5(1), 1997.

[8] P. R. Innocent and R. I. John. Computer aided medical diagnoses. *Information Sciences*, 162(2004):81–104.

[9] P. Julián-Iranzo. A procedure for the construction of a similarity relation. In *IPMU2008*, 2008.

[10] T. Kiseliova. A theoretical comparison of Disco and CADIAG-II-like systems for medical diagnoses. *Kybernetika*, 42(6):723–748, 2006.

[11] T. Kiseliova and S. Krajči. Generation of representative symptoms based on fuzzy concept lattices. In B. Reusch, editor, *Computational Intelligence. Theory and Applications: International Conference on Computational Intelligence, 8. Fuzzy Days in Dortmund*, pages 349–356, Dortmund, Germany, September 30 - October 2, 2004. Springer-Verlag Heidelberg.

[12] T. Kiseliova and K. Pagava. New approach to estimate different drugs and/or other medical interventions effectiveness based on fuzzy logic principles. *Georgia Medical News*, 3(156):65–68, March 2008.

[13] T. Kiseliova, K. Pagava, H. Pagava, and A. Pagava. The influence of socio-economical factors on health and health-related behaviour of adolescents in Georgia. In *4th International Conference on Children's Health and the Enviroment. Reducing environmental risks for our children*, volume 10 of *in Book of abstracts in: "MEDYCYNA SRODONOWISKOWA" (Environmental Medicine)*, page 135, Vienna, June, 10-12 2007.

[14] G. J. Klir and B. Yuan. *Fuzzy Sets and Fuzzy Logic. Theory and Applications*. Prentice Hall PTR, 1995.

[15] Gin-Shuh Liang, Tsung-Yu Chou, and Tzeu-Chen Han. Cluster analysis based on fuzzy equivalence relation. *European Journal of Operational Research*, 166(2005):160–171, 2005.

[16] R. Lohse. Fuzzy logic in life underwriting as illustrated by the cardiovascular risk assessment of diabetes melitus type I. In *Hannover Rück*, number 11. http://www.hannover-rueck.de, 2003.

ISBN: 978-989-95079-6-8

[17] M. L. Osborne, W. M. Vollmer, K. L. Pedula, J. Wilkins, A. S. Buist, and M. O'Hollaren. Lack of correlation of symptoms with specialist-assessed long-term asthma severity. *Chest*, 115:85–91, 1999.

[18] N. H. Phuong and D. H. Hung. Tuberdiag: an expert system for pulmonary tuberculosis diagnosis. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 7(4):371–381, 1999.

[19] F. Puppe. *Problemlösungsmethoden in Expertensystemen*. Springer-Verlag, Berlin, 1990.

[20] M. A. M. Reis, P. S. P. Silveira, and N. R. S. Orteg. Fuzzy expert system in prediction of neonatal resuscitation. *Brazilian Journal of Medical and Biological Research*, 37(5):755–764, May 2004.

[21] E. H. Shortliffe. *Computer-Based Medical Consultations, MYCIN*. Elsevier, North-Holland, New York, 1976.

[22] A. Sikharulidze. Decision-making aiding fuzzy information analysis in investments. *Proc. of Javakhishvili Tbilisi State University, Applied Mathematics, Computer Sciences,*, 353:77–94, 2003.

[23] L. Varverde and S. Ovchinnikov. Representation of t-similarity relations. *Fuzzy Sets and Systems*, 159(2008):2211–2220, 2008.

ISBN: 978-989-95079-6-8

Identification of Web Information using Concept Hierarchies and On-line Updates of Concept Importance

Zhan Li[1], Ronald R. Yager[2] and Marek Reformat[1]

[1]thinkS[2]: thinking software and system laboratory
Electrical and Computer Engineering
University of Alberta, Edmonton, Canada, T6G 2V4

[2]Machine Intelligence Institute
Iona College, New Rochelle, NY 10801

Abstract—*The Internet users have to perform a lot of search to find web pages with relevant information. The paper proposes an approach for utilization of a hierarchy of concepts to perform identification of web pages in the environment of the Semantic Web. A user provides a hierarchy of concepts that can only partially "cover" their domain of interest. Ontologies related to that domain are used to instantiate the hierarchy with concrete information, as well as to enhance it with new concepts initially unknown to the user. A web page is checked against concepts from the hierarchy and activation levels of those concepts are aggregated using Ordered Weighted Averaging (OWA) operators that are part of the hierarchy of concepts. Mechanisms of aggregations embedded in OWA operators are determined using linguistic quantifiers and importance of concepts. The importance of concepts is constantly changing, and in order to automatically assign importance values and "keep track" of changes a new algorithm for Adaptive Assigning of Term Importance (AATI) is used.*

Keywords—*hierarchy of concepts, ontology, ordered weighted aggregation, text identification*

1 Introduction

The enormous growth of the WWW (World Wide Web) creates a serious challenge in discovering relevant information among all web documents. Those documents vary in quality of information they contain, as well as in their timeliness. It is not easy to find the most relevant documents when the repository is constantly changing.

In general, there are two approaches used for finding relevant web pages, i.e., "Information retrieval" (IR) [4] and "Text Categorization" (TC) [8]. IR retrieves documents based on a query using a "keyword-matching" mechanism without considering different meanings of keywords under different contexts. In TC, there is no query, and the knowledge about a given category is embedded in models developed using machine learning (ML) techniques based on documents representing the category.

In real world scenarios, people relay on Search Engines (SEs) to retrieve required information. SEs accept users' queries and return sets of related web documents applying one of two approaches: building a full-text index-based search (IR-like approach), like Google and Altavista; or constructing taxonomies with web documents populating categories (TC-like approach), such as Yahoo.

People identify documents as belonging to a specific category based on a set of keywords. A person looks at a document and "searches" for keywords and associated with them concepts related to a given category. Moreover, for a person all keywords and concepts are interconnected and constitute a network. Activation of one keyword or concept initiates activation of related concepts. Relations between them can be of different nature: some can "point" to more *specific* concepts, some can relate to concepts that are *parts of* the original concept, and some can relate to concepts that *contain* the original one. More keywords and associated concepts are found in the document more support is collected towards assigning this document to a category described by the keywords.

An introduction of the Semantic Web [1] has led to the increased popularity of ontology as a means of representing knowledge in a way that a human being as well as a machine can understand. An ontology is a formal, explicit specification of a shared conceptualization [2]. It is a set of well-defined classes that describe data models in a specific domain. Together with their individuals (instances of classes), ontologies work as knowledge characters to express individual facts [7]. This new representation of knowledge introduced to the web environment brings new possibilities of utilization of information [5,6].

A different format of knowledge representation has been proposed in [10]. The proposed format, called Hierarchy of Concepts (*HOFC*), represents concepts with atomic attributes or other concepts. As the result a graph-like structure is established where each vertex is a concept, and terminal vertices are attributes. The edges of the hierarchy represent relationships that define concepts with a set of other concepts and/or attributes. An *HOFC* can be used for representing any human-like structure of concepts, for example queries [10].

ISBN: 978-989-95079-6-8

The paper introduces a new idea for constructing a category identifier using an *HOFC* combined with ontology and an algorithm for automatic updating of concept importance. In a nutshell, a user provides a simple hierarchy describing their category of interest. This hierarchy contains concepts that are the most obvious for the user. The concepts are linked together via different types of relationships. The concepts that constitute the hierarchy are generic – they are just definitions of concepts. The specific pieces of information are extracted form ontologies and "attached" to the equivalent concepts of the hierarchy as attributes – in other words ontologies are used to populate the hierarchy. Relations existing between ontology classes are used to find new concepts that are added to the hierarchy as well. This provides a mechanism for dynamic enhancement of hierarchies.

The dynamic character of the proposed approach is evident in a process of determining importance of *HOFC* concepts. Concept importance changes accordingly to changes in the contents of web pages – newly posted documents and texts affect importance of keywords in their ability to "identify" a given category. In order to "follow" these changes we equipped our *HOFC* with *AATI* [3]. The *AATI* is an un-supervised algorithm with ability to determine and update concept importance values automatically.

2 Background

2.1 Hierarchy of Concepts

In a nutshell, an *HOFC* uses a notion of representing concepts with atomic attributes or other concepts. Edges of *HOFC* are of significant importance to the whole idea of *HOFC*. If we assume that a concept C_1 is defined, described by or related to other two concepts C_2 and C_3, then the hierarchy will have two edges connecting C_2 with C_1, and C_3 with C_1. The concept C_1 is called a higher-level concept, and C_2 and C_3 are lower-level concepts. This also means that activation of concepts C_2 and C_3 leads to activation of C_1.

*HOFC*s introduce a very important feature – the process of activation of higher-level concepts by lower-level concepts is fully controlled by a user. There are two controlling components: importance vector $\boldsymbol{M}$ and linguistic quantifier Q. The vector $\boldsymbol{M}$ is an indicator of significance of each lower-level concept in defining a higher-level concept. In other words, $\boldsymbol{M}$ determines a weight of each of the participating (lower-level) concepts in a process of identifying an activation level of higher-level concept. The linguistic quantifier Q guides the aggregation of lower-level concept activations levels. We can have such quantifiers as *some*, *most*, *at least half*, *about 1/3*, or *for all*. Both $\boldsymbol{M}$ and Q determine how activation levels of lower-level concepts should be combined using the Ordered Weighted Averaging (*OWA*) operator [9].

In a formal representation, the *OWA* operator, defined on the unit interval I and having dimension n, is a mapping F_w: $I^n \rightarrow I$ such that:

$$F_w(a_1,\ldots,a_n) = \sum_{j=1}^{n} w_j * b_j \qquad (1)$$

where b_j is the j^{th} largest of all arguments $a_1, a_2,\ldots, a_n$, and w_j is a weight such that w_j is in $[0, 1]$ and the sum of w_j is equal 1. If $\boldsymbol{W}$ is an n-dimensional vector whose j^{th} component is w_j and $\boldsymbol{B}$ is an n-dimensional vector whose j^{th} component is b_j then $F_w(a_1, a_2,\ldots, a_n) = \boldsymbol{W}^T\boldsymbol{B}$. In this formulation $\boldsymbol{W}$ is referred to as the *OWA* weighting vector and $\boldsymbol{B}$ is called the ordered argument vector.

2.2 Ontology

The most important aspect of the ontology used for Semantic Web applications is related to identifying two ontology layers: the ontology definition layer, and the ontology instance layer[1].

The *ontology definition layer* represents a framework used for establishing a structure of ontology and for defining classes (concepts)[2] existing in a given domain. A structure of ontology is built based on a relation *is-a* between classes. This relation represents a *subClassOf* connection between a superclass and a subclass.

The ontology definition contains descriptions of classes. Classes are defined using two types of the properties:

- *datatype property* – this type of property focuses on describing features of a class; this property can be expressed as values of such data types as boolean, float, integer, string, and many more (for example, byte, date, decimal, time);
- *object property* – this property defines other than *is-a* relationships among classes (nodes); these relationships follow the notion of Resource Description Framework (RDF) [11] that is based on a triple *subject-predicate-object*, where: *subject* identifies what object the triple is describing; *predicate* (property) defines the piece of data in the object a value is given to; and *object* is the actual value of the property; for example, the triple "John likes books" has "John" as subject, "likes" as predicate and "books" as object.

Once the ontology definition is constructed, it can be instantiated. The properties of classes are filled out with real data – values are assigned to datatype properties, and links to instances of other classes are assigned to object properties.

2.3 AATI

Knowledge about a specific category is understood as the importance of keywords describing a category. The keywords and their importance are used to identify documents that belong to the category. The primary characteristic of a new algorithm for *Adaptive Assigning of Terms Importance (AATI)* is that it updates weights of keywords – called here terms – at the same time it categorizes web documents. It means that a training phase is not required [3]. The proposed algorithm is able to handle changes of terms and their importance. Knowledge collected by *AATI* is constantly updated, and there is no need to restart the algorithm when terms and their importance change.

AATI works on two sets of terms. One set contains terms that represent a given category, which is called *target* category,

[1] According to the terminology adopted by the Semantic Web community, the term "instance" has been replaced by the term "individual". For the purpose of clarity, the term instance – similar to the term *instantiated* – is used throughout the paper

[2] The term class is used to distinguish ontology concepts from *HOFC* concepts.

ISBN: 978-989-95079-6-8

while the other set contains terms that do not belong to a *target* category. *AATI* updates importance of the terms via changing weights associated with the terms called Term Weights (*TW*). The importance of terms reflects their discriminatory power in identifying a *target* category. Each web document is evaluated by summing *TW* of all terms that the page contains. The sum is called Page Value (*PV*). Please note that *TW* and *PV* are recursively defined. That is, *TW* and *PV* depend on each other.
Let a set $\boldsymbol{P} = \{p_1,p_2,p_3,...,p_m\}$ denotes *PV* of web documents, where each element p_i denotes *PV* of the *i*th document. Let a set $\boldsymbol{T} = \{t_1,t_2,t_3,...,t_n\}$ denotes terms in web documents, where each element t_j represents *TW* of the term *j*. In *AATI*, *TW* defines *PV*:

$$p_i = \sum_{j=0}^{n} \frac{f_{ij}}{N_i} t_j \tag{2}$$

where N_i is the number of words on a page *i*; f_{ij} is the number of occurrences of term *j* on the page *i*. If a term *j* does not occur on page *i*, f_{ij} is zero.
At the same time, *PV* defines *TW*:

$$t_j = \sum_{i=0}^{m} \frac{f_{ij}}{N'_j} p_i \tag{3}$$

where N'_j is the number of occurrences of term *j* in the document set containing *m* documents. If a term does not occur in the page *i*, f_{ij} is zero.
Therefore in the language of linear algebra, the definition between *PV* and *TW* can be written as

$$\begin{matrix} P = A \cdot T \\ T = B \cdot P \end{matrix} \Rightarrow T = (B \cdot A) \cdot T \Rightarrow T = C \cdot T \tag{4}$$

Thus, $\boldsymbol{T}$ is the eigenvector of matrix $\boldsymbol{C}$ with the eigenvalue one. At the same time, because both the sum of each row in the matrix $\boldsymbol{A}$ and that of each row in the matrix $\boldsymbol{B}$ are one, the sum of each row in the matrix $\boldsymbol{C}$ is also one. So, the eigenvector *T* as the solution of (4) exists and is unique.
Therefore, we define the algorithm *AATI* to evaluate *TW* :

- ***Step1***: build a list of terms for a *target* category, and assign zero as *TW* for all terms;
- ***Step2***: take a new web document; if there is no more web documents, ***STOP***;
- ***Step3***: parse the document based on the list of terms;
- ***Step4***: for each term occurring in the document do one of the following:
  (a) If the term value is not zero (it means the term has already appeared in documents), take this value as *TW*;
  (b) If the term value is zero (the term has not been found in any documents), randomly generate a number between 0 and 1, and assume it as *TW* for this term;
- ***Step5***: Calculate the *PV* for this web document by Equation (2)
- ***Step6***: Update the *TW* of those terms found in the web document (3);
- ***Step7***: Normalize TW across all terms;
- ***Step8***: Go back to Step2.

3 HOFC as Keyword Structure for Text Identification

3.1 Description of Domain of Interest

Text categorization uses keywords as representation of a category, and a text identification process checks if a given text contains those keywords. If all or fraction of keywords (depending of the applied technique) are found in a given document then the document is considered as belonging to a category described by the keywords. Usually, all keywords are not equally important and they are associated with weights identifying their importance. The keywords are represented as a flat structure, and there is no indication that they are "related" to each other, and that some of them depend on others.
A human, as it has been already indicated, looks at a document and uses not only keywords but also related keywords to identify a category the document belongs to. The presence of one keyword in the document activates other connected keywords. More keywords are found in the document more support is collected towards assigning this document to a category described by those keywords.
This human-like approach to use a network of keywords describing a specific category can be implemented with an *HOFC* (Section 2.1). An *HOFC* defines a concept with a set of other concepts or attributes[3]. The relations between a defined (higher-level) concept and defining (lower-level) concepts and attributes represent different types of associations that are meaningful for a user. Therefore, a structure of *HOFC* can be treated as a user-defined description of a given category.
This analogy goes even further when we take a closer look at the process of activation of *HOFC* concepts. In a nutshell, an activation level of higher-level concepts depends on activation levels of lower-level concepts. In other words, activation of a higher-level concept emerges via activation of more specific lower-level concepts and attributes. A level of activation of higher-level concepts is determined via an aggregation operator *OWA* defined with a linguistic quantifier (Q) and an importance vector ($\boldsymbol{M}$).

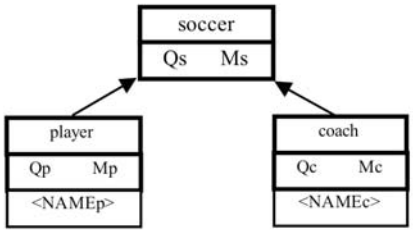


Figure 1: A simple structure of *HOFC* representing the category *soccer*

Fig. 1 shows a simple *HOFC* representing a *soccer* concept defined by two lower-level concepts: *player*, and *coach*. Each

[3] To ensure clarity it should be noted that among all phrases used in the paper the following relations hold: concepts are equivalent to classes; keywords, attributes and terms are equivalent to each other.

ISBN: 978-989-95079-6-8

of them has a single attribute: <NAMEp> and <NAMEc>, respectively. Concepts *soccer*, *player* and *coach* have their own linguistic quantifiers: *Qs*, *Qp*, and *Qc*, and importance vectors ***Ms***, ***Mp***, and ***Mc***. The vector ***Ms*** is of dimensionality 2, dimensionalities of ***Mp*** and ***Mc*** depend on a number of instances of each concept. Instances, that represent concrete pieces of information, are values of attributes <NAMEp> and <NAMEc>.

3.2 HOFC and Ontology

The application of *HOFC* to represent a user's view of the domain of interests brings a number of benefits. The *HOFC* allows us to incorporate different types of relations between concepts and attributes, and build quite complicated and well-interconnected nets of keywords/concepts describing the domain of interest.

However, users who provide *HOFC*s do not have to be experts in the domain. They can provide a very simple hierarchy – called a basic *HOFC* – that contains just a few concepts. This basic *HOFC* is treated as a starting point for building a more comprehensive *HOFC* better suited for identifying a category a given web page belongs to. The process of enriching the user-provided hierarchy is performed with ontology.

An ontology as proposed by the Semantic Web community (Section 2.2) represents a very attractive and powerful approach for knowledge representation. The richness of information that can be expressed with an ontology means that an ontology can be used to provide concrete and additional information for *HOFC*s. In the first case, we talk about instantiation of *HOFC*, and about enrichment of a hierarchy in the second case.

The process of instantiation can be explained with the help of a basic *HOFC*, Fig. 1. This hierarchy is "instantiated" with specific information obtained from the *soccer* category. The ontology is queried about *player*. The result – three instances of the class *player* – is presented in Fig. 2. The process of assigning values to *Q* and ***M*** is described in Section 3.3 and 4.

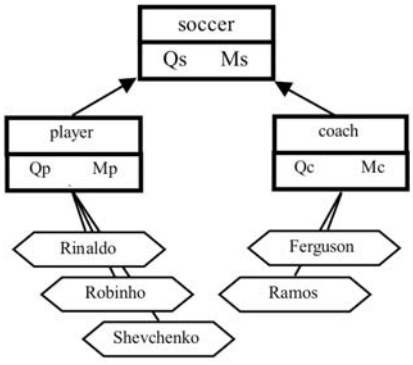


Figure 2: The HOFC of the category *soccer* with instances

An ontology can also be treated as a repository of new concepts that can be added to the concepts already existing in a *HOFC*. This enhancement can be twofold: in the form of addition of lower-level concepts, and in the form of providing concepts related in a "non-trivial" way to concepts already held in the hierarchy.

The concepts provided by a user are used to find corresponding classes in an ontology. The sub-classes of these classes are taken from the ontology and attached to the concepts of *HOFC* as new lower-level concepts. For an *HOFC* presented in Fig. 1, it would mean lower-level concepts: *forward*, *defender*, and *midfielder* are attached to the concept *player*.

Classes defined in an ontology possess object properties (Section 2.2). Object properties are definitions of "non-trivial" relations that exist between pairs of ontology classes. They are used to bring new concepts to an *HOFC*.

For example, the ontology class *player* has an object property *playsFor* that "binds" any instance of *player* with an instance of the ontology class *team*. If we assume that there is an instance of the class *player* called *Rinaldo*. *Rinaldo* has an object property *playsFor* that binds *Rinaldo* with an instance of the class *team – Manchester_United*, Fig. 3.

3.3 Text Identification process

A web page identification process corresponds to an "activation" of the top concept of an *HOFC*. The activation level of the top concept depends on activation levels of lower-level concepts and attributes. The activation of *HOFC* is determined in a bottom-up manner. Attributes at the bottom of *HOFC* are "matched" against words that appear in a web page. Presence of those words means activation of attributes, and activation of attributes means activation of upper-level concepts. This approach requires aggregation of activation levels of all attributes/concepts that define (are attached to) a concept. The aggregation process is performed using *OWA* (Section 2.1). The weights of *OWA* are determined based on *Q* and ***M***.

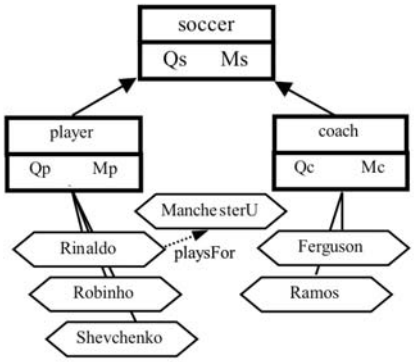


Figure 3: The *HOFC* of the category *soccer* with individuals and "enhancement"

Let $m_i \in [0,1]$ be a value associated with an attribute or concept a_i indicating its importance. If n concepts define a single higher-level concept then ***M*** of this concept is a n-dimensional importance vector $[m_1, m_2, \ldots, m_n]$, and the weighing vector ***W*** has to be calculated based upon both *Q* and ***M***. The first step is to determine the ordered argument vector ***B***, such that b_j is the j^{th} largest of all arguments $a_1, a_2, \ldots, a_n$. The arguments are numbers of occurrences of attributes in a

ISBN: 978-989-95079-6-8

text. Furthermore, we assume μ_j to denote the importance value associated with the attribute or concept that has the j^{th} largest value. Thus if a_5 is the largest value, then $b_1=a_5$ and $\mu_1=m_5$. The next step is to calculate the *OWA* weighing vector $\boldsymbol{W}$ using:

$$w_j = Q(\frac{S_j}{T}) - Q(\frac{S_{j-1}}{T})$$

where

$$S_j = \sum_{k=1}^{j} \mu_k \quad \text{and} \quad T = S_n = \sum_{k=1}^{n} \mu_k \cdot$$

So, S_j is the sum of the importances of the j^{th} most satisfied attributes/concpets, and T is the sum of all importances.
The linguistic quantifier Q is provided by a user at the time they provides their basic *HOFC*. During processes of enhancement, the values of Q for lower-level attributes and concepts are "inherited" from the upper-level concepts. In the case this is not possible a default quantifier *OR* is used.
The values of $\boldsymbol{M}$ representing importance of attributes and concepts of *HOFC* need to reflect a true significance of keywords in their capability to identify a category. The *AATI* algorithm capable of finding those values is used.

4 Importance of HofC Concepts

4.1 HOFC and AATI

Importance levels associated with each concept, keyword/attribute are considered as the knowledge about a domain of interest. Usage of the most accurate and relevant values is important for a good performance of the proposed *HOFC*-based approach. Therefore, the features of *AATI* (Section 2.3) make it a very good candidate for a process of updating and "tracking" changes in importance values of keywords. For the *AATI* algorithm, the importance values $\boldsymbol{M}$ are represented by *TW*.
The *AATI* algorithm modifies *TW* at the same time when it evaluates *PV* for each web document. This process is "controlled" by upcoming pages. If the content of stream of web documents changes, the *TW*s change too.
All concepts/keywords of *HOFC* are input to *AATI*. According to the algorithm their importance is randomly initiated but when time progresses and web pages are being processed – the values of *TW* are being updated and reflect strength of a connection of a term with a domain – strength of a term to "identify" the domain. These *TW* values are fed back to *HOFC* as $\boldsymbol{M}$ values.

4.2 Continuous Changes in Concept Importance

The ability of *AATI* to "follow" the web trends in popularity of a given keyword (popularity associated with a frequency of occurrence) can be illustrated with two examples. The first example represents a change in importance values when a source of web pages changes, while he second focuses on changes in the contents of web page over time.
Fig. 4 and 5 present examples of the adaptability of the *AATI* algorithm. Fig. 4 shows *TW* values of the term "Manchester United" (soccer ontology) as a function of a number of "processed" web pages. The first 1000 pages are from the BBC website (http://www.bbc.co.uk). The next 600 pages are from the CCN website (http://www.cnn.com). We see a decrease in the TW value (solid line after 1000 pages). If the next 600 were from the BBC website the values would follow the dashed line.

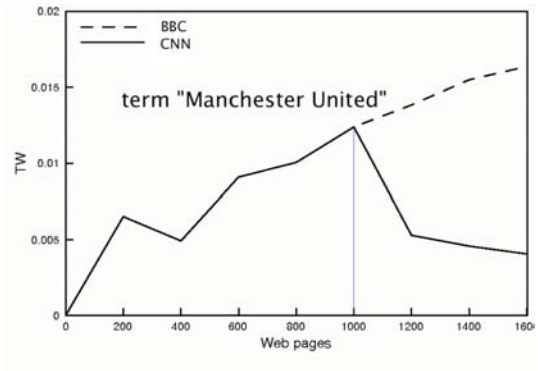


Figure 4: TW changes when the source of upcoming documents changes from BBC to CNN

The term "Manchester United" represents a name of a famous soccer club in England, and its *TW* goes up when the documents are downloaded form BBC. However, when the source of web pages changes and the upcoming documents are pages collected from the CNN site, and the term "Manchester United" is less popular. This decrease in popularity of the term "Manchester United" translates into decrease in its *TW*.
A different kind of change is seen in Fig. 5. The plots represent changes in importance levels when web pages collected at different moments in times are fed into the *AATI* algorithm. The first 2300 pages were collected from the BBC site in February 2007, while the last 2500 pages are from BBC site from January 2009. It can be seen that the importance levels for different terms behave differently: the importance of the player Ronaldo stays more or less at same level – indication that Ronaldo was and still is popular; the importance of the player Robinho increased and then slowly decreased, while the popularity of the player Shevchenko decreased.

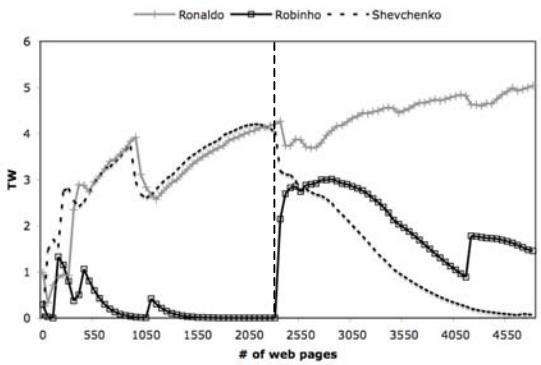


Figure 5: Changes of *TW* when the set of upcoming documents change from Feb07 to Jan09: for Ronaldo, Robinho, and Shevchenko

ISBN: 978-989-95079-6-8

5 Classification System

5.1 Architecture

An implementation of the proposed approach for web content identification requires integration of multiple components executing different tasks. A simplified architecture of a system performing the identification task is shown in Fig. 6. The system contains the following three parts: *HOFC* Preparation and Enhancement Unit, *AATI* Execution Unit, and Identification Unit. The *HOFC* Preperation and Enhancment Unit is responsible for setting up a basic *HOFC* representing user's query, adding new concepts from a domain ontology, and populating an enhanced *HOFC* with concrete pieces of information from the ontology. The *AATI* Unit is invoked on regular basis (based on time interval or a number of incoming web pages) to update values of importance of concepts. The Identification Unit checks the incoming web page against the concepts from the *HOFC*. This process starts with concepts located at the bottom of the *HOFC* – and it progresses towards the top. It is checked if the *HOFC* concepts are "activated". The activation level of concepts is the result of a number of their occurrences in a text, as well as Q and $\boldsymbol{M}$ (Section 3.3).
A prototype of the system has been developed in Java. The Protégé API is used to access ontology classes, their properties and instances. Moreover, API functions allow for creating a new enhanced hierarchy of concepts based on classes and individuals obtained from the domain ontology and the hierarchy of concepts given by a user.

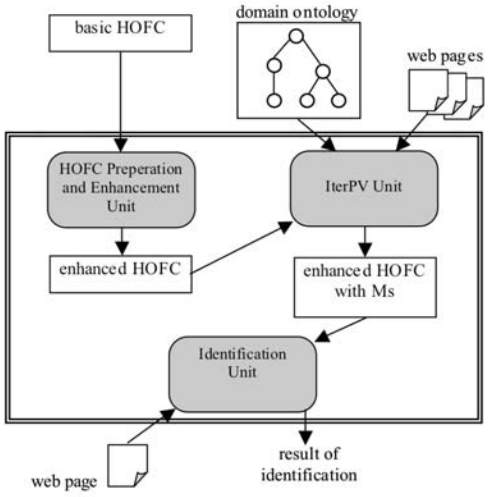


Figure 6: Text Identification System

5.2 Experimental Results

A simple example that illustrates application of the prototype system is presented. The *HOFC* used in the experiment is shown in Fig. 1. It is a very simple hierarchy representing the category *soccer*. A user is a novice to this category, and the only concepts that define soccer are *players* and *coach*. This *HOFC* is later instantiated and enhanced, Fig 3. The enhanced *HOFC* represents the following structure of concepts:

> The concept *soccer* is activated if there are "MOST" of *players* and *coaches*. The concept *players* is activated if there exist "OR" Ronaldo "OR" Robinho "OR" Shevchenko. The concept *coach* is activated when "OR" Ferguson "OR" Ramos are present.

This simple *HOFC* has been checked against a number of different web pages. In all cases the activation levels of the concept *soccer* reflected a human evaluation of pages. The activation levels have been calculated based on aggregations of activations of keywords – Ronaldo, Robinho, Shevchenko, Ferguson, Ramos – using simple linguistic quantifiers MOST and OR, and values of $\boldsymbol{Ms}$ provided by *AATI*.

6 Conclusions

The paper proposes a concept-based structure suitable for text identification. The method uses hierarchies of concepts as category identifiers. The hierarchies use *OWA* operators with linguistic quantifiers and concept importance to control conditions that have to be satisfied by lower-level concepts and attributes in order to activate a higher-level concept.
The concept importance levels are updated using the *AATI* algorithm. This algorithm is able to modify concept and attribute importance values in on-line fashion in an un-supervised manner.
A prototype of the system implementing the proposed approach is briefly described in the paper.

Reference

[1] T. Berners-Lee, J. Hendler, and O. Lassila, The Semantic Web, *Scientific American*, 284: 34-43, May 2001.

[2] T.E. Gruber, A translation approach to portable ontology specifications, *Knowledge Acquisition*, 5: 199-220, 1993.

[3] Z. Li and M. Reformat, An Iterative Algorithm for Text Categorization, submitted to *Information Processing Letters*.

[4] C. Manning, P. Raghavan, and H. Schutze, *Introduction to information retrieval*, Cambridge University Press, New York, NY, 2008

[5] T. Martin, Searching and smushing on the semantic web-challenges for soft computing, in M. Nikravesh, B. Azvine, R. R. Yager, & L. A. Zadeh (Ed.), *Enhancing the Power of the Web*, Heidelberg: Springer, 167-188, 2004.

[6] E. Sanchez, and T. Yamanoi, Fuzzy Ontologies for the Semantic Web, in *Flexible Query Answering Systems*, 691-699, 2006.

[7] F. Scott, W.D. Lewis, and D.T. Langendoen, An ontology for Linguistic Annotation 1-2, in Proceedings of *14th Innovative Applications of AI Conference*, Edmonton, Canada, 11-19, 2002.

[8] F. Sebastiani, Machine learning in automated text categorization, *ACM Computing Survey(CSUR)*, 34(1): 1-47, 2002.

[9] R.R. Yager, On ordered weighted averaging aggregation operators in multi-criteria decision making. *IEEE Transactions on Systems, Man and Cybernetics*, 18: 183-190, 1988.

[10] R.R. Yager, A Hierarchical Document Retrieval Language, *Information Retrieval*, 3: 357-377, 2000.

[11] RDF: http://www.w3.org/RDF/

ISBN: 978-989-95079-6-8

Possibility distribution: a unified representation for parameter estimation

Gilles P. Mauris

LISTIC, Polytech'Savoie, Université de Savoie
Annecy, France
Email: mauris@univ-savoie.fr

Abstract— *the paper presents a possibility formulation of one-parameter estimation that unifies some usual probability formulations. Point and confidence interval estimation are unified in a single theoretical formulation and incorporated into estimators of an omnibus form: a possibility distribution. New relationships between continuous possibility distribution and probability concepts are established. The concept of specificity ordering of a possibility distribution is then used for comparing the efficiency of different estimators. The usefulness of the approach is illustrated on mean and median estimators from data sample of different size and of different probability distribution.*

Keywords— efficiency, parameter estimation, possibility, probability, stochastic ordering, uncertainty.

1 Introduction

Systems designed by engineers are often meant to influence their environment: to manipulate it, to move it, to control it and so on. To enable such actions, these systems need information, e.g. values of physical quantities describing their environment. Generally, two types of information sources are available: prior knowledge and empirical knowledge. The latter is obtained by sensor observations. Prior knowledge is the knowledge that was already there before a given observation became available. The combination of prior knowledge and empirical knowledge leads to posterior knowledge useful for the task at hand. Parameter Estimation deals with this problem to infer a parametric description for an object, a physical process or an event, given measurements that always come with uncertainties due to variability of influence quantities or phenomena. Therefore an assessment of the parameter estimate uncertainty is also required by the application, and prior knowledge is often used to reduce the uncertainty of the final estimate. To tackle such problems a lot of developments have been done in the statistical science [1].

In statistics, a parameter estimator $T_\theta(X)$ is defined as a function of the observable sample data $X_1, X_2, ..., X_n$ that is used to estimate an unknown parameter θ; an estimate $\hat{\theta}$ is the result from the actual application of the function to a particular set of data. Prior knowledge can be that $X_1, X_2, ..., X_n$ is an independent identical distributed (*iid*) sample from a continuous random variable X having a distribution function F. Many different estimators are possible for any given parameter. Two main types of estimators are worldwide used: point estimators and interval estimators. The selection of an estimator of one of the above kinds for purposes of informative inference, including typical applications in scientific research is generally admitted to involve elements of choice which are in some degree arbitrary. Such elements include the choice of a particular confidence level for an interval estimator, and the choice of a location function for a point estimator for a given confidence coefficient. In addition, a point estimate is often desired along with an interval. Such considerations and related ones have led to proposals for simultaneously use of a point confidence estimator and a set of confidence limit or interval estimation having various confidence coefficients [2]. Such estimators may be regarded as a modern formulation of a long standing practice of reporting estimates in the form $\hat{\theta} \pm k\sigma$, where k is some constant and σ the standard deviation of the observations. Note that this expression is recommended by the Guide of Uncertainty in Measurement (GUM [3]) edited by international legal metrology institutions. This form may be interpreted as an ordered set of three point estimators. Tukey proposed in [4] that for general purposes, it would be advantageous to use a set of five points estimators at standard confidence level *2,5%,16,66%, 50%, 83,33% and 97,5%*. Cox proposed to use the full continuous family of confidence levels varying between *0* and *1* by introducing a so called confidence curve estimator [5].

The problem of choosing a good estimator, i.e. an estimator which tends to take values close to the true unknown value, is formulated mathematical by introducing criteria of closeness. The latter has been specified by the introduction of specific loss function (called sometimes risk function). For example, the absolute error criterion was introduced by Laplace, Gauss replaced this by the squared error criterion which proved mathematically much more tractable [2]. Each such definite specification of closeness can be criticized as somewhat arbitrary, except in a context when one postulates the reality of the indicated cost of errors of each possible kind. This point put forth by Galileo leads to comparisons between estimators on the basis of all their probabilities rather than on the basis of certain summaries such as the variance [2]. Birnbaum extended this concept to continuous random variables by introducing the notion of peakedness of a probability distribution, which characterizes the concentration of values around the estimate [6].

ISBN: 978-989-95079-6-8

In this paper, we propose a possibility formulation of one parameter estimation, which unifies the usual previous probability formulation. Our proposition is based on a possibility/probability transformation we previously proposed [7][8], and on the notion of possibility distribution specificity ordering [9]. In section 2, this transformation is recalled and new relationships with the above mentioned probability estimation concepts are exhibited. In particular, the efficiency of estimators is characterized by its stochastic concentration defined by the possibility specificity. The approach is illustrated in section 3 on the problems of mean and median estimation. The advantages of the possibility formulation, which leads to an intuitive graphic representation, are highlighted, particularly when the sample size varies and the probability distribution as well.

2 Possibility versus probability estimation

2.1 Basics of possibility theory

Possibility theory is one of the modern theories available to represent uncertainty when information is scarce and/or imprecise [10]. The basic notion is the possibility distribution [11], denoted π. Here, we consider possibility distributions defined on the real line, i.e. π is an upper semi-continuous mapping from the real line to the unit interval. Thus π is a fuzzy subset but with specific semantics for the membership function. Indeed, a possibility distribution describes the more or less plausible values of some uncertain variable *X*. The possibility theory provides two evaluations of the likelihood of an event, for instance whether the value of a real variable *X* does lie within a certain interval: the possibility Π and the necessity N. The normalized measures of possibility Π and necessity N are defined from the possibility distribution $\pi : R \to [0,1]$ such that $\sup_{x \in R} \pi(x) = 1$ as follows:

$\forall A \subset R, \Pi(A) = \sup_{x \in A} \pi(x)$ and

$\forall A \subset R, N(A) = 1 - \Pi(\bar{A}) = \inf_{x \notin A}(1 - \pi(x))$

The possibility measure Π satisfies:

$\forall A, B \subset R, \Pi(A \cup B) = \max(\Pi(A), \Pi(B))$

The necessity measure N satisfies:

$\forall A, B \subset R, N(A \cap B) = \min(N(A), N(B))$

In fact, possibility measures are set functions similar to probability measures, but they rely on axioms which involve the operations " maximum " and " minimum " instead of the operations "addition" and "product" (if the measures are decomposable).

Moreover we can interpret any pair of dual functions necessity/possibility $[N, \Pi]$ as upper and lower probabilities induced from specific probability families. Let π be a possibility distribution inducing a pair of functions $[N, \Pi]$. We define the probability family $\mathcal{P}(\pi) = \{P, \forall A, N(A) \le P(A)\} = \{P, \forall A, P(A) \le \Pi(A)\}$. In this case, $\sup_{P \in \mathcal{P}(\pi)} P(A) = \Pi(A)$ and $\inf_{P \in \mathcal{P}(\pi)} P(A) = N(A)$ [12]. In other words, the family $\mathcal{P}(\pi)$ is entirely determined by the probability intervals it generates. In the estimation context, this fact can be useful when it is not possible to identify one specific probability distribution for the observations. The concept of specificity can be used to qualify the informativeness of a possibility distribution. Indeed, a possibility distribution π_1 is said more specific than π_2 as soon as $\forall x, \pi_1(x) \le \pi_2(x)$ [9](it is the usual definition of inclusion of fuzzy sets), i.e. π_1 is more informative than π_2.

2.2 Background on probability/possibility transformation

Let us assume that the sample data are issued from a continuous random variable X defined on the set of reals and described by a probability density function f, F being its corresponding cumulative distribution function, with F^{-1} its inverse function if it exists (otherwise the pseudo-inverse function can be considered [13]). For every possible confidence level $\beta \in [0,1]$, the corresponding confidence interval is defined as an interval that contains the parameter to be estimated, i.e. θ, with probability $\ge \beta$. In other words, a confidence interval of confidence level β (denoted I_β) is defined as an interval for which the probability P_{out} to be outside this interval I_β does not exceed $\alpha \overset{def}{=} 1 - \beta$, i.e. $P(\theta \notin I_\beta) = \alpha$.

It is possible to link confidence intervals and possibility distribution in the following way. A unimodal numerical possibility distribution may be viewed as a nested set of confidence intervals, which are the α cuts $[\underline{x}_\alpha, \overline{x}_\alpha] = \{x, \pi(x) \ge \alpha\}$ of π. The degree of certainty that $[\underline{x}_\alpha, \overline{x}_\alpha]$ contains μ is $N([\underline{x}_\alpha, \overline{x}_\alpha])$ (if π continuous). Obviously, the confidence intervals built around the same point x^* are nested. It has been proven in [7] that stacking confidence intervals of a probability distribution on top of one another leads to a possibility distribution (denoted π^* having x^* as modal value). In fact, in this way, the α-cuts of π^*, i.e. $A_\alpha = \{x | \pi^*(x) \ge \alpha\}$ are identified with the confidence interval I^*_β of confidence level $\beta = 1 - \alpha$ around the nominal value x^*. Thus, the possibility distribution π^* encodes the whole set of confidence intervals in its membership function. Moreover, this possibility distribution satisfies:

$\forall A \subset R, \Pi^*(A) \ge P(A)$, with Π^* and P the possibility and probability measures associated respectively to π^* and f.

A closed form expression of the possibility distribution $\pi^M(x)$ induced by confidence intervals around the median $x^* = M$ is obtained for symmetric unimodal continuous probability densities $f(x)$ strictly increasing on the left and decreasing on the right of M [7]:

$$\begin{aligned} &\forall x \in [-\infty, M], \pi^M(x) = 2F(x) \\ &\forall x \in [M, +\infty], \pi^M(x) = 2(1 - F(x)) \end{aligned} \quad (1)$$

Thus, the possibility distribution is closely related to the cumulative distribution function.

ISBN: 978-989-95079-6-8

2.3 Relationships with probability estimation notions

According to the way a possibility distribution is built from the probability distribution F from which the sample data come from, the link with confidence intervals is immediate. Indeed, generally the parameter to be estimated is a location parameter of a distribution function, e.g. the mean. The link with Cox confidence curve $c_\theta(x)$ is also obvious. Indeed the latter is defined by [5]:

$\forall \theta \in [-\infty, x], c_\theta(x) = F(\theta - x)$

$\forall \theta \in [x, +\infty], c_\theta(x) = 1 - F(\theta - x)$

Thus, $c_\theta(x) = \pi_F(x-\theta)/2$.

Other new links concern the quality of estimator that is related to the concentration of values around the estimate, quantified for example by the variance or by the absolute mean deviation. Hereafter a few proposition formalizing for continuous symmetric random variables the links with the absolute mean deviation, the Levy concentration function, and other stochastic orderings.

Proposition 1:

The specificity index $sp(\pi) = \int_{-\infty}^{+\infty} \pi^M(x)dx$ of the possibility $\pi^M(x)$ distribution equivalent to a continuous symmetric random variable X is equal to twice the absolute mean deviation:

$$sp(\pi_X^M) = 2.E|X - M| .$$

Proof:

Assume without loss of generality that *M=0*. Due to symmetry:

$\int_{-\infty}^{+\infty} \pi_X^M(x)dx = \int_{-\infty}^{0} \pi_X^M(x)dx + \int_{0}^{+\infty} \pi_X^M(x)dx =$

$2\int_{-\infty}^{0} \pi_X^M(x)dx = 4\int_{-\infty}^{O} F(x)dx$

In other respect, as $F(-\infty) = 0$:

$2\int_{-\infty}^{0} F(x)dx = 2[xF(x)]_{-\infty}^{0} - 2\int_{-\infty}^{0} xf(x)dx$

$= 0 - \int_{-\infty}^{\circ\infty} xf(x)dx = E(|X|)$

Therefore: $sp(\pi_X^M) = \int_{-\infty}^{+\infty} \pi_X^M(x)dx = 2.E|X - M|$.

Thus the possibility distribution is closely related to the absolute mean deviation, (and not to the conventional standard deviation), of course if the latter is defined. Indeed, when the term $\int_{-\infty}^{0} xf(x)dx$ is infinite, the possibility distribution can be computed but the associated specificity is infinite.

Proposition 2:

The possibility distribution $\pi_X^M(x)$ equivalent to a continuous unimodal symmetric random variable X is related to the Lévy concentration function $Q_X(x)$ by the relation:

$$\forall x \geq 0, Q_X(x) = 1 - \pi_X^M(x + M)$$

Proof:

The concentration function introduced by P. Lévy in 1935 has been first defined by the expression [14, page 44]:

$\forall x \geq 0, Q_X(x) = \sup_{x_0} [F(x_0 + x) - F(x_0 - x)]$

Note that, in a latter definition [14, page 90], which is the one considered usually, the expression has became: $Q_X(x) = \sup_{x_0} [F(x_0 + x) - F(x_0)]$. Here we will consider the first definition, more close to our approach.

If F is a symmetric and unimodal distribution about M then, the supremum in equation is attained for $x_0 = M$ (see [8] for a demonstration). Thus we have:

$\sup_{x_0} [F(x_0 + x) - F(x_0 - x)] = F(M + x) - F(M - x)$.

The symmetry of F gives for $x \geq 0$:

$F(M + x) - F(M - x) = F(M + x) - (1 - F(M + x))$

$= 2F(M + x) - 1 = 1 - 2(1 - F(M + x)) = 1 - \pi(x)$.

Therefore, $Q_X(x) = 1 - \pi_X^M(x + M)$.

In fact, the possibility distribution gathers probability dispersion intervals and it is complementary to the Lévy concentration function for continuous unimodal random variables.

Proposition 3:

The specificity order (defined by the fuzzy subset inclusion) defines a stochastic concentration ordering equivalent for continuous unimodal symmetric random variables both to the peakedness ordering of Birnbaum [6] and to the majorization ordering of Hickey [15]:

$$\pi_X^M(x) \leq \pi_Y^M(x) \Leftrightarrow X \geq^{peaked} Y \Leftrightarrow X \leq^{maj} Y$$

Proof:

According to Birnbaum [6], X is said to be more peaked about θ , than Y about θ if and only if:

$\Pr(|X - \theta| \geq t) \leq \Pr(|Y - \theta| \geq t)$ holds for all $t \geq 0$.

It is clear that we have:

$\pi_Y(\theta - t) = \pi_Y(\theta + t) = 1 - \Pr(\theta - t, \theta + t) = \Pr(|X - \theta| \geq t$

Thus: $\pi_X^M(x) \leq \pi_Y^M(x) \Leftrightarrow X \geq^{peaked} Y$

Note also that peakenedness ordering is equivalent to conventional stochastic ordering of absolute variables:

$X \geq^{peaked} Y \Leftrightarrow |X - \theta| \leq^{sto} |Y - \theta| \overset{def}{\Leftrightarrow} F_{|X-\theta|}(x) \leq F_{|Y-\theta|}(x)$

The majorization order introduced by Hickey is used to compare continuous distribution in terms of randomness. Y is said at least as random as X in the majorization sense if [15]:

$\int_o^t g^*(y)dy \leq \int_o^t f^*(y)dy, \forall t$,

where f^* and g^* are the decreasing rearrangement of f and g respectively; that is:

$f^*(x) = \sup\{c : m(c) > x\}, x > 0$ with,

$m(c) = \mu\{x : f(x) > c\}$, μ denoting Lebesgue measure and f the corresponding density function of F .

ISBN: 978-989-95079-6-8

It is proved in [16] that for unimodal continuous random variables:

$\int_o^t f^*(y)dy = Q_X(t)$.

Thus: $\pi_X^M(x) \le \pi_Y^M(x) \Leftrightarrow Q_X(x) \ge Q_Y(x) \Leftrightarrow X \le^{maj} Y$.

Note also that a similar proposition as the proposition 3, but for discrete probability and possibility distributions has been stated by Dubois and Hüllermeier [17].

Thus given two estimators $T_\theta(X)$ and $U_\theta(X)$ of a parameter θ, T is said to have greater stochastic concentration about than U does if T is more peaked than U . An interesting consequence of greater peakedness proved by Hwang [18] is that if L is any loss function L satisfying: $L(x,\theta) = h(|x-\theta|)$,

where for each θ, $h:[0,\infty) \to [0,\infty)$ is a non decreasing function , we have:

$\pi_{T_\theta(X)}(x) \le \pi_{U_\theta(X)}(x) \Leftrightarrow E_\vartheta(L(T_\theta(X)) \le E_\vartheta(L(U_\theta(X))$

In other words not only is T at least as likely as U to be within a neighborhood $[\theta - x, \theta + x]$ of the true parameter, it is at least as good with regard to a broad class of loss functions which includes absolute mean deviation ($L(x,\theta) = |x-\theta|$), variance ($L(x,\theta) = |x-\theta|^2$), and any other symmetric loss function where the loss is at least as large when x is farther away from θ than when it is closer to θ . It includes also the continuous entropy. Indeed, the latter is defined as: $H(X) = -\int_X f(x)Ln(f(x))$, and the majorization order implies the entropy order [19]. Therefore, for continuous unimodal distribution, the specificity order implies the entropy order.

Thus, the preference ordering of estimators based on greater possibility specificity around the parameter is indeed a strong preference ordering.

2.4 Relationships with other notions

The specificity order (defined by the fuzzy subset inclusion) is also related to second order stochastic dominance (SOSD) in connection with risk-aversion in decision theory. In this context, the notion of dispersion is translated to a notion of risk. Given two symmetric distributions with the same expected value, G is riskier than F if every risk-averse individual prefers F to G. One mathematical definition is:

$F \le_{SOSD} G \Leftrightarrow \int_{-\infty}^t F(x)dx \le \int_{-\infty}^t G(x)dx, \forall t$

We have: $\pi_X^M(x) \le \pi_Y^M(x) \Rightarrow X \le_{SOSD} Y$.

Proof: $\int_{-\infty}^t F(x)dx \le \int_{-\infty}^t G(x)dx$ is equivalent to:

$U(t) = \int_{-\infty}^t (F(x) - G(x))dx \le 0, \forall t$

We have: $U'(t) = F(t) - G(t) \le 0, \forall t \le 0$

due to $\pi_X^M(x) \le \pi_Y^M(x)$. Thus U is decreasing, and since $U(-\infty) = 0$, then $\forall t \le 0, U(t) \le 0$.

By a symmetric reasoning, $\forall t \ge 0, U(t) \le 0$, which completes the proof.

A link exists also with the Value at Risk notion $VaR_\alpha(X)$ that is defined by the 1-α quantile of X, i.e.:

$P(X > VaR_\alpha(X)) = 1 - \alpha$.

It is easy to see that:

$\pi_X^M(x) \le \pi_Y^M(x) \Rightarrow VaR_\alpha(X) \ge VaR_\alpha(Y), \forall \alpha \in [0,1]$

The specificity order is also related to notions of social science, i.e. the Lorenz curve and the Gini index defined by:

$L_X(t) = \frac{1}{E(X)} \int_O^t F(x)dx$ and $G(X) = 1 - 2\int_O^1 L_X(t)dt$

In fact, Lorenz dominance order is equivalent to second order stochastic dominance for distributions with equals means [19]. Thus, we have:

$\pi_X^M(x) \le \pi_Y^M(x) \Rightarrow L_X(x) \le L_Y(x), \forall x$ and,

$\pi_X^M(x) \le \pi_Y^M(x) \Rightarrow G(X) \ge G(Y)$.

The Gini index is a measure of statistical dispersion most prominently used as a measure of inequality of wealth distribution. A low Gini index indicates more equal wealth distribution, while a high Gini index indicates more unequal distribution.

3 Mean and median estimators

Let us consider the case where the parameter to be estimated is a location parameter of a continuous symmetric distribution function. This situation is often encountered in physical entity measurements where the variability of observations is due to variability of influence quantities or phenomena [3].

We considered hereafter the parameter estimation by using the mean and the median estimators.

3.1 Mean estimator

Suppose $X_1, X_2, ..., X_n$ are an *iid* sample from a distribution F . The mean estimator is defined: $\hat{\theta}_n = \left(\sum_{i=1}^n X_i\right)/n$.

Generally, the Gaussian distribution with a deviation σ is used for F , because, in this case the distribution of the mean estimator is also Gaussian but with a standard deviation $\sigma/\sqrt{n}$. Hereafter are plotted the possibility distributions associated to such mean estimators for $n = 1, 3, 30$ and $\sigma = 1$. As expected, the larger the sample size the more specific the possibility distribution.

ISBN: 978-989-95079-6-8

Let us now consider that the sample data are issued from a Cauchy distribution with a scaling parameter of *1*: $f(x) = \frac{1}{\pi(1+x^2)}$.

This distribution has no finite moments, but the median exists and can be used as a measure of central tendency. According to equation (1), the equivalent possibility distribution is defined by:

$$\forall x \le 0, \pi(x) = 1 + \frac{2}{\pi} arctg(x)$$

$$\forall x \ge 0, \pi(x) = 1 - \frac{2}{\pi} arctg(x)$$

The probability distribution of the mean estimator is also a Cauchy distribution with a scale parameter of *1* [1]. Thus the possibility distribution of the mean estimator of any number of data is the same as the one obtained for one datum. Therefore the specificity is not improved by increasing the sample size. Note that this is not in contradiction with the central limit theorem though the latter considers only distribution with finite variance (which is not the case of the Cauchy distribution).

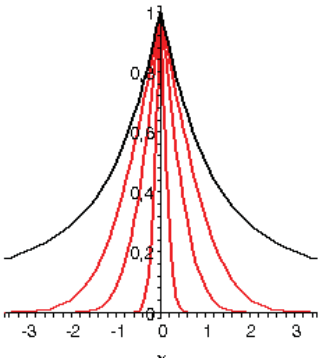


Figure 1: Possibility mean estimators for Normal (red) and Cauchy (black) probability distributions (*n=1, 3, 30*).

3.2 Median estimator

Let M_n be the median of $X_1, X_2, ..., X_n$, θ being the median of the distribution F from which the sample data are issued. For n odd, we have the following relationship [20]:

$$P(M_n - \theta \ge x) = \sum_{i=0}^{(n-1)/2} C_n^i F(x)^i (1 - F(x))^{n-i}.$$

Therefore the expression of the possibility distribution of median estimator is:

$$\forall x \le \theta, \pi_{T_\theta(X)}(x) = 2(1 - \sum_{i=0}^{(n-1)/2} C_n^i F(x)^i (1 - F(x))^{n-i})$$

$$\forall x \ge \theta, \pi_{T_\theta(X)}(x) = 2 \sum_{i=0}^{(n-1)/2} C_n^i F(x)^i (1 - F(x))^{n-i}$$

The possibility median estimators for a Gaussian and a Cauchy distribution are plotted in figure 2 for $n = 1, 3$.

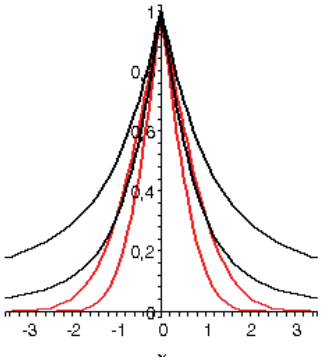


Figure 2: Possibility median estimators for Cauchy (black) and Normal (red) distributions (n=1, 3).

Note that the specificity increases with the sample size in both cases, which is remarkable for the Cauchy distribution.

3.3 Comparison

The preceding results show that the specificity of the possibility distribution associated to a median estimator increases with the sample size (with odd data) even for a long tail probability distribution such as the Cauchy distribution. It is not the case for the mean estimators. This seems indicate that the median estimator is better. Hereafter are plotted the possibility mean and median estimators for a few observations from a normal distribution.

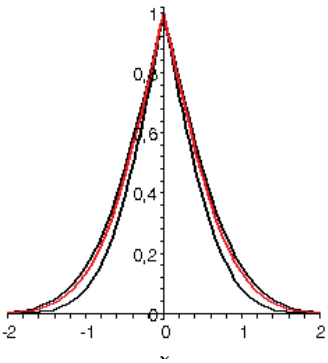


Figure 3: Possibility mean (black, n=2, 3) and median (red, n=3) estimators for a Normal distribution.

The possibility distribution of the median estimators with 3 observations is more specific than the possibility distribution of the mean estimator for 2 observations but less specific than the possibility distribution of the mean estimator for 3 observations. It seems that in general the possibility median estimator for *2n+1* data is more specific than the mean estimator for *2n* data.

ISBN: 978-989-95079-6-8

4 Conclusions

The paper has presented new links between continuous symmetric probability and possibility distributions. The interest of the possibility approach in the context of parameter estimation has been highlighted through the powerful concept of specificity ordering of possibility distribution. The latter allows sounded and intuitive comparisons between different estimators (e.g. mean and median). Further developments should consider dissymmetric distributions and other estimators.

References

[1] J. Rice, *Mathematical Statistics and Data Analysis*, Duxbury Press, 1995, 597 pages.

[2] A. Birnbaum, "A unified theory of estimation", *The Annals of Mathematical Statistics*, Vol. 32, No 1, 1961, p. 112-135.

[3] *Guide for the expression of uncertainty in measurement*, ISO 1993, 1993.

[4] J.W. Tuckey, "Standard confidence points", *Memorandum report 26,* Princeton University, 1949.

[5] D.R. Cox, "Some problems connected with statistical inference", *The Annals of Mathematical Statistics*, Vol. 29, 1958, p. 357-372.

[6] Z.W. Birnbaum, "On random variables with comparable peakedness", *The Annals of Mathematical Statistics*, Vol. 19, 1948, p. 76-81.

[7] G. Mauris, V. Lasserre and L. Foulloy, "A fuzzy approach for the expression of uncertainty in measurement", *Int. Journal of Measurement*, Vol. 29, No 3, March 2001, pp. 165-177.

[8] D. Dubois, L. Foulloy, G. Mauris and H. Prade, "Probability-possibility transformations, triangular fuzzy sets and probabilistic inequalities", *Reliable Computing*, Vol. 10, No 4, 2004, pp. 273-297.

[9] D. Dubois, H. Prade and S. Sandri, "On possibility/probability transformations", in*: Fuzzy Logic*, R. Lowen, M. Roubens, Eds, 1993, pp. 103-112.

[10] C. Baudrit and D. Dubois, "Practical representation of incomplete probabilistic information", *Computational Statistics & Data Analysis*, Vol. 51, Issue 1, 2006, pp. 86-108.

[11] L.A. Zadeh, "Fuzzy sets as a basis for a theory of possibility", *Fuzzy Sets and Systems*, Vol.1, No1, 1978, pp. 3-28.

[12] D. Dubois and H. Prade," On several representations of an uncertain body of evidence", in: *Fuzzy Information and Decision Processes* (M.M. Gupta, E. Sanchez, eds.), North-Holland, 1982, pp. 167-181.

[13] M. Kendall and A. Stuart, *The advanced theory of statistics*, Ed. Griffin and Co., 1977.

[14] P. Lévy, *Théorie de l'addition des variables aléatoires,* Gauthier-Villars, Paris, 1954, 416 pages.

[15] R.J. Hickey, "Continuous majorization and randomness", *Journal of applied probability*, No. 23, 1984, pp. 924-929.

[16] J.M. Fernandez-Ponce and A. Suarez-Llorens, "An aging concept based on majorization", *Probability in the engineering and informational science*, No 17, 2003, pp. 107-117.

[17] D. Dubois and H. Hüllermeier, "Comparing probability measures using possibility theory: a notion of relative peakedness", *Int. journal of Approximate reasoning,* No 45, 2007, pp. 364-385.

[18] J.T. Hwang, "Universal domination and stochastic domination: estimation simultaneously under a broad class of loss functions", *Annals of Statistics,* No 13, 1985, pp. 295-314.

[19] A. Müller and D. Stoyan, *Comparison methods for stochastic models and risks*, Wiley, 2002, 330 pages.

[20] H.A. David, *Order Statistics*, Wiley, 2003.

ISBN: 978-989-95079-6-8

Cooperative Partner Agent of Seven-Card-Stud Poker

K. Osone[1] T. Onisawa[2]

1,2 Graduate School of Systems and Information Engineering,
University of Tsukuba, 1-1-1, Tennodai, Tsukuba, 305-8573 Japan
Email: 1:kfukuda@fhuman.esys.tsukuba.ac.jp, 2:onisawa@iit.tsukuba.ac.jp

Abstract— *This paper aims at the construction of a cooperative partner agent that plays a seven-card stud poker game with a human partner against an opponent player. If a human player needs some advice on a game, the partner agent gives a human partner player strategy for a game and its grounds. A human player can also ask some questions about the current situation of a game. If a human player tries to take different strategy from the one presented by the partner agent, the partner agent calls human player's attention to the presented strategy with its grounds. Fuzzy theory and case based reasoning are applied to decision-making part in the partner agent. This paper also performs subject experiments to evaluate the effectiveness of the presented partner agent, where each subject plays games with/without the partner agent.*

Keywords— game with imperfect information, poker game, cooperation, fuzzy inferences, case based reasoning

1 Introduction

In our daily life we are often confronted with decision-making problems under uncertain situations. Although we think of many alternatives or various different opinions under such situation [1], it is usually difficult to solve them completely. Therefore, we often get other perspectives or hints to solve problems by exchanging opinions among others or by taking advice from others. An adviser with ripe experience may give better advice on decision-making problems. Exchanging opinions or taking advice may lead to some solution that cannot be found by oneself or to some other understanding on problems. From this point of view, there are many studies on decision support systems [2,3,4,5,6]. However, these systems only support users' decision and do not consider decision-making program from the viewpoint of cooperation with human users.

This paper considers the situation in which human makes decision in cooperation with a computer. In this paper, as an example of decision-making problem, seven-card-stud poker game, a kind of a game with imperfect information [7], is taken because situations estimate is usually difficult under uncertainty of poker game situations. Although there are also some studies about negotiation agent in a game, e.g., a MONOPOLY game [8], or poker playing systems [9,10,11], studies on decision-making from the viewpoint of cooperation with human players in a game are not found.

This paper aims at the construction of not a mere poker playing system but a partner agent, which has a discussion with a human player on game strategy in a seven-card-stud poker game cooperating with a partner player. The presented agent is based on the authors' previous seven-card-stud poker game playing system [12,13,14], and is enlarged from our previous partner agent [15] in a sense that the system gives some advice to a human partner player using case based reasoning. A human player plays poker games with the presented partner agent that cooperates with a human player by not only presenting information on game situations but also performing question and answering of game strategy. Therefore, this paper considers not only gain/loss of point but also the human partner player's evaluation of usefulness/helpfulness.

The organization of the paper is as follows. In Section 2 a seven-card-stud poker game used in this paper is explained. Section 3 presents the partner agent consisting of the decision-making module and the cooperation module. Section 4 mentions subject experiments performed in order to show the usefulness of the presented partner agent and their results. Final section shows conclusions of this paper.

2 Seven-Card-Stud Poker

2.1 Stud Poker

A stud poker game is different from a draw poker one. And although a stud poker game has various kinds of game variations, face-down and face-up cards are dealt to each player in every stud poker game. The card dealt face down is called a hole card that only a card owner can see. On the other hand the card dealt face up is called an up card that all players can see.

2.2 Seven-Card-Stud Poker

This paper considers the situation in which a human player and the partner agent play a poker game together against the previous poker playing system [13,14] as an opponent player. This section describes rules of a seven-card-stud poker game used in this paper.

After both players (a human player and the playing system) pay 5 points as the ante, two hole cards and one up card are dealt to each player, and the first betting round is started. The second player must bet at least the same amount of bet as the first player's one, where one player can raise the betting point only once at each betting round, and the upper limit of betting points is assumed to be total points betted so far. When both players' betting points become even, a next card is dealt to each player and the next betting round is started. The fourth, the fifth and the sixth cards are face-up cards, and the seventh

ISBN: 978-989-95079-6-8

card is a hole card. After all cards are dealt to and betting rounds are ended, each player makes a decision which hole card is revealed. Each player reveals one hole card and the betting round is started. If any players still don't fold a game, the procedure of revealing hole cards and that of betting round are continued. Bluff strategy or slow play is also often taken in the procedure of revealing hole cards. If both players continue a game after the final betting round, they show their final hole cards each other, which is called Showdown, and the player having a stronger poker hand with five cards out of seven is a winner of a game.

2.3 Poker Hands

This paper considers 10 sorts of poker hands including no pair (high card). Scores of these hands are defined as shown in Table 1. In this paper, scores are simply defined as 10 scales from 0 to 9 because the linear scale is simple.

Table 1:Score of poker hands

Poker Hands	Scores	Poker Hands	Scores
Nothing	0	Flush	5
One Pair	1	Full House	6
Two Pairs	2	Four of a Kind	7
Three of a Kind	3	Straight Flush	8
Straight	4	Royal Straight Flush	9

3 Partner Agent

3.1 Relationship between Agent and Players

A human player plays a poker game in cooperation with the partner agent against an opponent player. It is assumed that when the partner agent is called by a human player according to game situations, the agent presents its strategy to a human player. Furthermore, when a human player has some questions on some strategy, the partner agent answers these questions. It is also assumed that the partner agent has same information as the one a human player does. Fig.1 shows the decision-making procedure of the partner agent, which has the decision-making module and the cooperation module.

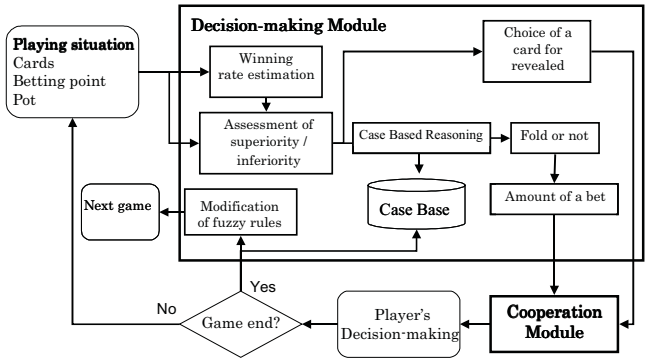


Figure 1: Decision-making procedure of partner agent

3.2 Decision-making Module

The decision-making module of the partner agent consists of six parts, winning rate estimation, assessment of superiority/inferiority, decision-making to fold or not, decision-making on betting points, choice of a card to be revealed and modification of fuzzy rules.

3.2.1 Winning Rate Estimation

In order to estimate the winning rate of a partner player, the agent suppositionally assigns possible combination of cards to human and opponent players' hole cards and to their up cards dealt from now on. The agent makes a judgment on human partner player's winning or losing. The agent repeats this trial and counts the number of the partner player's winning. Let this repeating trial times and the number of human partner player's winning be N_{times} and N_{win}, respectively. The winning rate is defined by

$$P_{win} = \frac{N_{win}}{N_{times}} \cdot \tag{1}$$

In this paper, N_{times} is fixed to be 10000.

3.2.2 Assessment of Superiority/Inferiority

The relative strength of the human partner player's poker hand to the opponent player's one is defined by

$$expect = 2(P_{win} - 0.5) . \tag{2}$$

The *expect* means that the larger the *expect* is, the higher the possibility of the superiority of a human player's hand is. In this paper the obtained *expect* is adjusted by fuzzy inference using fuzzy rules considering the tendency of opponent player's strategy, betting points and the number of turn. Table 2 shows an example of fuzzy rules at the early game stage, where the turn is normalized by 7, which is the maximum number of betting round times in a seven-card-stud poker game, and betting points are normalized by the pot. As Table 2 shows, the fuzzy rules have parameter values $t_{ij} \in [0,100]$ called the tendency values. The tendency value means that if t_{ij} is larger than 50, an opponent player tends to take aggressive strategy, and if t_{ij} is smaller than 50, an opponent player tends to take passive strategy [14]. The initial values of all tendency values are defined to be 50 and the tendency values are changed every game according to the opponent player's strategy.

Table 2: Fuzzy rules for assessment of superiority/inferiority at early game

Adjustment degree (Turn is Early)		Opponent's bet		
		Few	Middle	Much
Hand strength	Weak	Negative t_{11} Positive	Negative t_{12} Positive	Negative t_{13} Positive
	Strong	Negative t_{21} Positive	Negative t_{22} Positive	Negative t_{23} Positive

ISBN: 978-989-95079-6-8

The fuzzy inference using these fuzzy rules is performed as follows. Let the result of fuzzy inference using the Positive consequent part and that using the Negative consequent part be $C_{positive}$ and $C_{negative}$, respectively, where the min-max inference method and the center of gravity are used in the fuzzy inference [18].

$$infer = C_{positive} \times (\frac{t_{ij}}{100}) + C_{negative} \times (1 - \frac{t_{ij}}{100}) . \quad (3)$$

And $expect$ is adjusted in the range of [0,1] using the result of the fuzzy inference (3) by

$$result = \begin{cases} expect + (1 - expect) \times infer \\ \qquad (0 \le infer \le 1) \\ expect + (expect + 1) \times infer \\ \qquad (-1 \le infer \le 0) \end{cases} . \quad (4)$$

The $result$ is the assessment result of the superiority/ inferiority of a human player's poker hand.

3.2.3 Case Based Reasoning

In order to use case examples in past games for decision-making, case based reasoning (CBR) [17] is applied in this paper. The following items are considered as indices of problem features used in CBR: (i) $result$ in Eq. (4), (ii) opponent's bet, (iii) change of $result$ which is the difference between $result$ of the current situation and $result$ of the previous situation, (iv) change of opponent's bet which is the difference between opponent's bet of the current situation and opponent's bet of the previous situation, (v) $potential$ in Eq. (7), (vi) change of potential which is the difference between $potential$ of the current situation and potential of the previous situation, (vii) the number of turn.

The following items are considered as solutions of a problem: (i) a player folds a game or not, (ii) strategy is successful or not. The similarity degree of each item is obtained by

$$Sim_i = 1 - |Past_i - Current_i| , \quad (5)$$

where $Past_i (i = 1,2,...,n)$ is index of the i-th problem feature in a past case, $Current_i (i = 1,2,...,n)$ is the index of the i-th problem feature in a current case, and n is the number of problem features. In this paper, n is fixed at 7. The whole similarity degree is obtained by

$$Sim_{whole} = \frac{\sum_{i=1}^{n} Sim_i}{n} . \quad (6)$$

Case with the largest whole similarity degree larger than 0.9 is defined as the most similar case.

3.2.4 Decision to Fold or Not

From the first through the fifth turns the partner agent has the following decision procedures. When the agent assesses the partner player's hand at superiority and continuing a game is chosen as strategy by CBR, then the partner agent advises to continue a game. However, when folding a game is chosen as strategy by CBR and the whole similarity degree is larger than 0.95, then the partner agent advises to fold a game even if the agent assesses the hand at superiority. When the partner agent assesses the partner player's hand at inferiority, the partner agent makes a decision by the fuzzy inference whether a partner player should fold a game or not. An example of the rules is; if opponent's betting points are large and the number of turn is large, then fold a game with high possibility. If the fuzzy inference result($\in [0,1]$), i.e., the possibility of folding a game, is larger than a uniformly random generated number($\in [0,1]$), the partner agent makes a decision to fold a game. However, when continuing a game is chosen as strategy by CBR and the whole similarity degree is larger than 0.95, then the partner agent advises to continue a game even if the agent assess the partner's hand at inferiority.

In the sixth and the seventh turns, a poker hand is already determined because the seventh card is dealt. However, even if the hand strength is assessed at inferiority, there is sometimes a case where it is possible to take bluff strategy. The presented agent decides whether bluff strategy should be taken or not according to the following procedures.

In order to choose a hole card to be revealed, the potentiality is obtained by the following way. Assuming that one partner player's hole card is revealed, possible combinations of cards are assigned to remaining hole card(s). The expected value of the strength of possible poker hands is obtained by

$$potential = \sum_{i=1}^{m} u_i p(u_i) , \quad (7)$$

where m is the number of partner's possible poker hands, u_i is the score of the possible hand as shown in Table 1, and $p(u_i)$ is the probability of u_i. This expected value is called $potential$ in this paper. In the same way, the agent obtains the expected value of opponent player's hand called $Opp_potential$. If Eq. (8) holds good, the partner agent decides not to fold and to take a bluff strategy, when Eq. (8) is considered by previous work [15]. However, even if the partner agent decides not to fold, when folding a game is chosen as strategy by CBR and the whole similarity degree is larger than 0.95, then the partner agent advises to fold a game.

$$potential \ge Opp_potential + 1.3 \quad (8)$$

3.2.5 Decision on Betting Points

Let the upper limit and the lower limit of betting points be $upper$ and $lower$, respectively, where these limits are dependent on the betting round and are provided by the poker game rules. In this paper the upper limit is assumed to be the total points betted so far and the lower limit is assumed to be 5 points. The bet points are obtained by

$$C_{bet} = lower + a(upper - lower) , \quad (9)$$

where a is defined by

ISBN: 978-989-95079-6-8

$$a = \frac{result + 1}{2} + noise\,, \tag{10}$$

where $noise$ is a normally distributed random number with the mean value 0 and the variance 0.1. If bluff strategy is considered by the agent, C_{bet} in Eq. (9) is obtained using a defined by

$$a = \frac{(potential/9) + 1}{2}, \tag{11}$$

where $potential$ is obtained by Eq.(7) and is normalized by 9, i.e., the maximum score of poker hands defined by Table 1. This means that the agent bets large points in order to make the partner player's hand look strong using bluff strategy.

3.2.6 Choice of a Card to be Revealed

In the fifth and the sixth turns, after the betting round is finished, each player chooses one hole card to be revealed and turns it face up. When the agent assesses a partner's hand at superiority and the hand is already shown in face-up cards, a card to be revealed is chosen at random. When the agent assesses a partner's hand at superiority and the hand is hidden in hole cards, the agent advises to reveal the card at the rate of 80%, of which *potential* is the lowest. The agent advises to reveal the card at the rate of 20%, of which *potential* is the second lowest. When the agent assesses a partner's hand at inferiority, the agent advises to reveal the card of which *potential* is the highest.

3.2.7 Modification of Fuzzy Rules

All cards are shown at the game end and both players' hands are revealed. Every one game the modification part of fuzzy rules estimates the change of the tendency value considering the difference between the estimated relative hand strength and actual hand strength by using fuzzy inference. The modification is performed for the fuzzy rule of which satisfaction degree is the highest in the assessment of superiority/inferiority. The fuzzy rules for modifying the fuzzy rules for assessment of superiority/inferiority have the following form of statement. (1) If the estimated relative poker hand strength is lower than the actual poker hand strength, then the tendency value is changed to be small. (2) If the estimated relative poker hand strength is higher than the actual poker hand strength, then the tendency value is changed to be large. (3) If the estimated relative poker hand strength and the actual one are even, then the tendency value is not changed.

The tendency value is actually changed considering the last five estimated changing values of the tendency value and the latest changing value in order to consider opponent player's tendency in a long period.

3.3 Cooperation Module

The cooperation section has two parts. One part gives a human partner player strategy to take in the current game situation and its grounds. The other gives a human player reply to human player's questions about the current game situation.

3.3.1 Strategy Presentation

Agent's strategy presentation to a human partner player is performed at the decision-making process in each betting round or at the process of the human player's choice of a hole card to be revealed after the fifth turn. When the partner agent is called by a human player, the agent gives strategy for the current game situation.

3.3.1.1 Decision to fold or not to fold

At the beginning of each betting round, when a human player calls the partner agent, the partner agent presents its decision and its grounds to a human player according to the assessment of superiority/inferiority obtained by Eq. (4). If the agent decides to fold according to 3.2.4, then the agent presents its grounds to a human player. And when there are case examples similar to the current situation in a game, the partner agent presents a CBR result to a human player. For example, the partner agent says that you have past similar case examples, you folded games in the past similar situations, and that folding a game is good solution for the current situation.

3.3.1.2 Decision of amount of bed

The partner agent presents the point value obtained by 3.2.5. For example, when the agent decides the betting point to be 10, the agent says that 10 points should be bet for the current situation.

3.3.1.3 Decision to call

After an opponent player bets some points, these decision procedures are performed. The agent decides to fold or not as the same procedures in 3.4.1.1

3.3.1.4 Choice of a card to be revealed

The agent presents a card chosen by the procedure in 3.2.6

3.3.1.5 A human player takes another strategy

Although the agent presents strategy to fold a game to a human player, if a human player tries to take different strategy from the presented one, the agent calls the attention to a human player about the game continuation. For example, the agent presents *you should fold this game, because your hand is weak and opponent player's bet is large*.

3.3.2 Answer to Player's Question

When a human partner player asks the agent some question about the current game situation, the agent replies to it.

Table 3: Linguistic expressions corresponding to t_{ij}

t_{ij}	Linguistic Expressions
100-70	Very aggressive
70-60	Aggressive
60-50	Rather aggressive
50-40	Rather passive
40-30	Passive
30-0	Very passive

ISBN: 978-989-95079-6-8

3.3.2.1 Human player's game playing tendency

When the agent gets a question about human player's game playing tendency, the agent presents linguistic expressions based on the tendency value t_{ij} explained in 3.2.2. Although the tendency value t_{ij} is essentially obtained for an opponent player, the partner agent can obtain not only an opponent player's tendency value but also a partner player's one by the same procedures described in 3.2.7. For example, the agent says that a human partner player tends to be passive in the game situation such as the current one. Table 3 shows the linguistic expressions corresponding to t_{ij} .

3.3.2.2 Estimated expected opponent player's hand

The agent performs simulation repeatedly by assigning possible combination of cards to opponent player's hole cards and/or opponent player's cards dealt from now on. Then, the agent estimates the opponent player's poker hand with the highest hand probability and the poker hand with its probability. The agent presents these poker hands and their estimated probabilities to a human partner player. For example, the agent says that the opponent player's hand with the highest probability is One Pair with 46.9% and that the estimated highest score hand is Straight with 10.5%. If a human partner player asks the agent about other poker hands, the agent also replies them.

3.3.2.3 Estimated expected human player's hand

The agent performs simulation repeatedly by assigning possible combination of cards to a human partner player's cards to be dealt from now on, and obtains the partner player's poker hand with the highest hand probability and the highest score poker hand with its probability. Then, the agent presents these poker hands and their estimated probabilities to a partner player. If a human partner player asks the agent about other poker hands, the agent also replies them.

3.3.2.4 Estimated human player's hand strength from opponent player's side

The agent performs simulation repeatedly by assigning possible combination of cards to a human partner player's hole cards and cards to be dealt from now on, and obtains the partner player's poker hand with the highest hand probability and the highest score poker hand with its probability. Then the agent presents these poker hands and their estimated probabilities to a partner player. These probabilities give much information for decision making to a human player whether bluff strategy can be taken or not. If a human partner player asks the agent about other poker hands, the agent also replies them.

4 Experiments

4.1 Experimental Procedures

Subject experiments are carried out in order to evaluate the effectiveness of the presented agent. Each subject plays 30 poker games with the presented partner agent against the poker playing system that has the winning rate estimation part added to our previous playing system, and other 30 poker games without the partner agent. If all 30 games are finished by a showdown, prepared cards are set to have 15 wins and 15 loses, respectively. Subjects are divided into two groups in order to analyze playing strategy in every game with/without the partner agent. Cards in each card set are dealt to all subjects in the same order. That is, all subjects play games on the same condition in the experiments. However, subjects do not know it.

The number of subjects is 12. Although they are novices of a poker game, they have knowledge on poker game rules.

4.2 Experimental Results

4.2.1 Gain/Loss of subjects

Table 4 shows the overall gain/loss points of each subject with/without the partner agent against the playing system. Wilcoxon signed-rank tests [18] of gain/loss points of all subjects with/without the partner agent are carried out. The observed value is 0.98, and significance level 5% is 1.65 on directional test. The null hypotheses are not rejected. This test result shows that there is no difference between gain/loss points with the partner agent and those without the partner agent. However, it is found that nine subjects out of twelve can get more points with the partner agent than without it. Subjects B and H bet large point in betting rounds because both subject and partner agent assess their poker hand at superiority. Therefore, they lose a lot of points. For example, although player hand is tree of a kind, however, opponent player hand is straight, and player and partner agent assess their poker hand at superiority. It is the reason why there is no difference between gain/loss points with the partner agent and those without the partner agent.

Table 4: Gain/Loss of subjects

Subject	without partner agent	with partner agent
A	-17925	2313
B	18217	-79140
C	8193	13817
D	14789	110104
E	12015	-12367
F	-225734	-87563
G	13817	87509
H	-293177	-400371
I	21123	52980
J	19055	21763
K	-52172	-16293
L	-121285	-31182

4.2.2 Evaluations of partner agent

4.2.2.1 Usefulness of partner agent

After the experiments, subjects evaluate the usefulness of the partner agent with a 5-points grade, 5: very useful, 4: useful, 3: neutral, 2: useless, 1: very useless. Fig. 2 shows the evaluation results. Ten subjects out of twelve mark 4 or 5 and the average score among all subjects is 4.2, i.e., useful. And there are no subjects marking 2 or 1. Then subjects' evaluation of the partner agent is high. As a mater of fact, some subjects have comments that the partner agent is useful for the estimation of

ISBN: 978-989-95079-6-8

opponent player's hand and the estimation of superior/inferior of the partner player's hand. And other subjects have comments that decision of betting points becomes easier with the partner agent than without agent because decision of betting points by himself is very difficult for him.

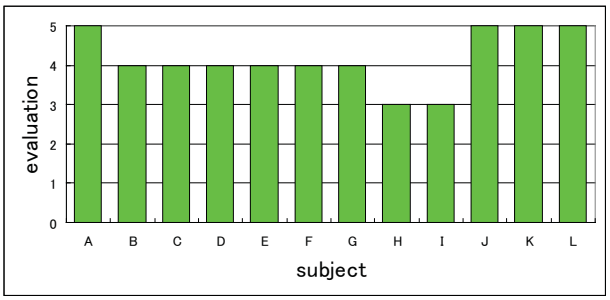


Figure 2: Usefulness of partner agent

4.2.2.2 Learning effect of partner agent

In order to confirm the learning effect of the presented partner agent, after the experiments, the subjects also evaluate whether the partner agent's advices are helpful or not with a 5-points grade, 5: very helpful, 4: helpful, 3: neutral, 2: unhelpful, 1: very unhelpful. Fig. 3 shows the evaluation results. Seven subjects out of twelve mark 4 or 5 and the average score among all subjects is 3.6, i.e., more or less helpful. And there are no subjects marking 1. In this paper, the learning effect is simply evaluated by subjective assessment and subjects' comments analyze the learning effect objectively. Some subjects have comments that they feel that the partner agent's advises are helpful because the partner agent gives them advices referring to case examples them advices are intelligible to them. It is found that the agent gives partner players good advices using CBR. However, some subjects negative comment. Subject H evaluates the partner agent's advices are not helpful because although he wants to play a game more aggressively, the partner agent advises him to fold a game referring to case examples.

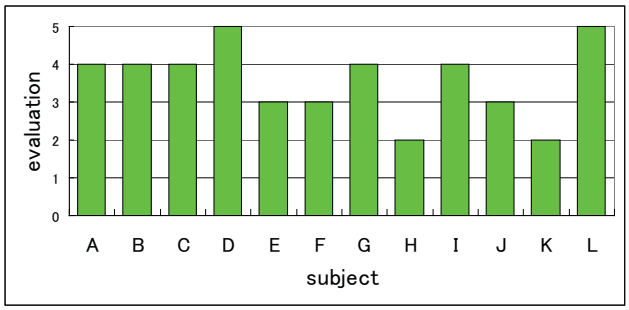


Figure 3: Helpfulness of partner agent's advices

5 Conclusions

This paper constructs a partner agent that supports a human player in playing a seven-card stud poker game. This agent presents its own strategy and/or replies some questions about game strategy in the current game situation when a human player calls it. In order to make use case examples in past games, CBR is also applied. Subject experiments are performed to evaluate the effectiveness of the presented partner agent. Experimental results show that although subjects do not necessarily get more points with the partner agent than without the agent from the viewpoints of gain/loss points, subject's evaluation of usefulness of the partner agent is high. And the evaluation of helpfulness of partner agent's advices is more or less high. It is found that CBR is useful to the presented partner agent.

References

[1] Toshimasa Moriwaki, group, organization, Tokyo Daigaku Shuppan, 2000

[2] Koki Matsura, Negotiation Simulation System Based on Fuzzy Decision-Making, Journal of Japan Society for Fuzzy Theory and System, Vol.8, No.5, pp. 882-890 1996

[3] D.Dubois, M. Grabisch, F. Modave and H. Prade, Relating decision under uncertainty and multicriteria decision making models, Frontiers in Soft Computing and Decision Systems, pp. 6-15, 1997

[4] Gleb Beliakov and Jim Warren, Appropriate Choice of Aggregation Operators in Fuzzy Decision Support System, IEEE TRANSACTION ON FUZZY STSTENS, vol. 9, No. 6, pp. 773-783, 2001

[5] John Zeleznikow, James R. Nolan, Using soft computing to build real intelligent decision support system in uncertain domains, Decision Support 31, pp. 263-285, 2001

[6] Chang-Shing Lee, Chen-Yu Pan, An intelligent fuzzy agent for meeting scheduling decision support system, Fuzzy Set and Systems 142, pp. 467-488, 2004

[7] Norio Sasaki: Introduction of game theory, Nippon Hyoron Sha, 2003

[8] Toshiaki Yasumura, Hidehisa Akiyama, Kunihiko Oguti, Katsumi Nitta: Negotiation Agents in the MONOPOLY Game, Information Processing Society Japan Vol43, No.10, pp. 3048-3055, 200

[9] Darse Billings, Aron Davidson, Jonathan Schaeffer, Duane Szafron, The challenge of poker, Artificial Intelligence 134, pp. 201-240, 2002

[10] Graham Kendall, Mark Willdig, An Investigation of an Adapting Poker Player, Artificial Intelligense, LNAI 2246, pp.189-200, 2001

[11] Arild Sandven and Bjorner Tessem, A Case-Based Learner for Poker, Proceedings of the Ninth Scandinavian Conference on Artificial Intelligence (SCAI 2006) pp. 159-167.

[12] Takehisa ONISAWA, Satoru KAZAMI, Chiharu TAKAHASHI, Construction of Playing System of Game with Imperfect Information -Seven-Card Stud Poke-, Journal of Japan Society for Fuzzy Theory and System, Vol.15, No.1,pp. 127-141, 2003

[13] Takehisa Onisawa and Chiharu Takahashi, Bluff Strategy in Seven-card-Stud Poker Game, Proceedings of the 10th IFSA World Congress, Istanbul, TURKY pp. 491-494, 2003

[14] T. Onisawa and C.Takahashi, Poker Playing system Considering Opponent Player's Strategy, Proceedings of SICE Annual Conference 2004 in Sapporo, pp. 2347-2352, 2004

[15] K. Ohsone, T. Onisawa, Cooperative Partner agent of Imperfect Information Game 2008 IEEE Conference on Soft Computing in Industrial Applications (SMCia/08), June 25-27, Muroran, JAPAN pp. 60-65, 2008

[16] R. Kruse, J. Gebhardt, F.Klawonn Foundations of FUZZY SYSTEMS, JOHN WILEY & SONS, 1994

[17] Yoshio Nakatani, Recent Developments in Case-Based Reasoning, Journal of the Japanese society for Artificial Intelligence, Vol.17 No.1 pp. 28-33, 2002

[18] Ann Hughes, Dennis Grawoig, Statistics: A Foundation for Analysis, ADDISON-WESLEY PUBLISHING COMPANY, 1971

ISBN: 978-989-95079-6-8

Systems of Fuzzy Relation Equations in a Space with Fuzzy Preorder

Irina Perfilieva

Institute for Research and Applications of Fuzzy Modeling, University of Ostrava
Ostrava, Czech Republic
Email: Irina.Perfilieva@osu.cz

Abstract— In this paper, we consider the problem of solving systems of fuzzy relation equations in a space with fuzzy preorder. Two types of these systems with different compositions are considered. New solvability criteria are proposed for systems of both types. The new criteria are weaker than all the known ones that are based on the assumption that fuzzy sets on the left-hand side of a system establish a fuzzy partition of a respective universe.

Keywords— system of fuzzy relation equations, fuzzy preorder, criterion of solvability

1 Preliminaries

Let throughout this contribution $\mathscr{L} = \langle L, \vee, \wedge, *, \rightarrow, 0, 1\rangle$ be an integral, residuated, commutative l-monoid (*a residuated lattice*), X a non-empty set and L^X a set of L-valued functions on X. Fuzzy subsets of X are identified with L-valued functions on X (membership functions).

Let X and Y be two universes, not necessary different, $A_i \in L^X$, $B_i \in L^Y$ arbitrarily chosen fuzzy subsets of respective universes, and $R \in L^{X\times Y}$ a fuzzy subset of $X \times Y$. The latter is called a fuzzy relation. Lattice operations $\vee$ and $\wedge$ induce respective union and intersection of fuzzy sets. Two other binary operations $*, \rightarrow$ of $\mathscr{L}$ are used for compositions - binary operations on $L^{X\times Y}$. We will consider two of them: sup-*-composition that is usually denoted by $\circ$, and $\inf$-$\rightarrow$-composition that is denoted by $\triangleleft$. The first one has been introduced by L. Zadeh [16] and the second one by W. Bandler and L. Kohout [1]. We will demonstrate definitions of both compositions on particular examples of set-relation compositions $A \circ R$ and $A \triangleleft R$ where $A \in L^X$ and $R \in L^{X\times Y}$:

$$(A \circ R)(y) = \bigvee_{x\in X} (A(x) * R(x, y)),$$
$$(A \triangleleft R)(y) = \bigvee_{x\in X} (A(x) \rightarrow R(x, y)).$$

Remark 1
Let us remark that both compositions can be considered as set-set compositions where R is assumed to be replaced by a fuzzy set. In this reduced form they are used in instances of systems of fuzzy relation equations below.

By a *system of fuzzy relation equations with sup-*-composition* (SFRE$*$), we mean the following system of equations

$$A_i \circ R = B_i, \quad 1 \le i \le n, \tag{1}$$

that is considered with respect to unknown fuzzy relation $R \in L^{X\times Y}$. Its counterpart is a *system of fuzzy relation equations with* $\inf$-$\rightarrow$*composition* (SFRE$\rightarrow$)

$$A_i \triangleleft R = B_i, \quad 1 \le i \le n, \tag{2}$$

that is considered with respect to unknown $R \in L^{X\times Y}$ also. System (1) and its potential solutions are well investigated in the literature (see e.g. [3, 2, 4, 5, 12, 8, 13, 15]). On the other hand, investigation of solvability of (2) is not so intensive (see [2, 10]).

Both systems of fuzzy relation equations arise when a system of fuzzy IF-THEN rules is modeled by a fuzzy relation (below in (3) it is denoted by R), and continuity of the model [9] is requested. In order to explain this request, we recall that in relation models, a computation of an output value (B) which relates to a given input $A \in L^X$ is performed with the help of sup-*, respectively $\inf$-$\rightarrow$ composition:

$$B = A \circ R \quad \text{or} \quad B = A \triangleleft R. \tag{3}$$

(3) is a computational realization of the *Generalized Modus Ponens* inference scheme in fuzzy logic (in a broader sense). It is often welcome if thus constructed model is *continuous* in the sense that when (input) fuzzy sets $A', A'' \in L^X$ are close to each other (in some space) so do output fuzzy sets $A' \circ R$ and $A'' \circ R$ (respectively, $A' \triangleleft R$ and $A'' \triangleleft R$). We proved in [9] that this is possible if and only if R solves the respective system (1) or (2) of fuzzy relation equations. This fact gives additional importance to the problem of solvability of systems of fuzzy relation equations.

In general, solutions of (1) or (2) may not exist. Therefore, investigation of necessary and sufficient conditions for solvability (or at least, sufficient conditions) is needed. This problem has been widely studied in the literature, and some nice theoretical results have been obtained in the cited above papers.

If the universes of discourse X and Y are infinite then the complexity of verification of necessary and sufficient conditions is comparable with the direct checking of solvability. Therefore, the problem of discovering easy-to-check solvability conditions or criteria is still open (see [11, 14] for some results). Besides other, this paper is a contribution to this topic.

We recall basic facts concerning solvability of system (1) of fuzzy relation equations (similar conditions are known for system (2), so that we will not recall theme (see e.g. [2, 10])).

Theorem 1
(a) [13] If system (1) with respect to unknown fuzzy relation R is solvable then relation

$$\hat{R}(x, y) = \bigwedge_{i=1}^{n} (A_i(x) \rightarrow B_i(y)) \tag{4}$$

is the greatest solution to (1).

ISBN: 978-989-95079-6-8

(b) *[5] Let fuzzy sets $A_i \in L^X$ and $B_i \in L^Y$, $1 \leq i \leq n$, be normal. Then the fuzzy relation*

$$\check{R}(x,y) = \bigvee_{i=1}^{n} (A_i(x) * B_i(y)) \tag{5}$$

is a solution to (1) if and only if

$$\bigvee_{x\in X} (A_i(x) * A_j(x)) \leq \bigwedge_{y\in Y} (B_i(y) \leftrightarrow B_j(y)) \tag{6}$$

holds for all $i, j = 1, \ldots, n$.

(c) *[7] Let fuzzy sets $A_1, \ldots, A_n \in L^X$ be normal and $x_1, \ldots, x_n \in X$ be pairwise different elements such that for all $i = 1, \ldots, n$, $A_i(x_i) = 1$. Let moreover, for all $i, j = 1, \ldots, n$,*

$$\bigvee_{x\in X} (A_j(x) * A_i(x)) \leq \bigwedge_{x\in X} (A_j(x) \leftrightarrow A_i(x)) \tag{7}$$

holds true. Then (1) is solvable if and only if

$$\bigvee_{x\in X} (A_j(x) * A_i(x)) \leq \bigwedge_{y\in Y} (B_i(y) \leftrightarrow B_j(y)).$$

In this contribution, we will prove weaker criteria of solvability for systems (1) and (2) than those ones, consider above in cases (b) and (c) (see Section 5 with the Discussion). Moreover, an algorithm of verifying solvability that is based on each proposed here criterion has a polynomial complexity.

2 Fuzzy Preorders and Their Normal Upper Sets

This section is auxiliary with respect to the problem of solvability discussed above. In this section, we will first recall basic facts that relate to spaces with fuzzy preorder [6]. Then we will prove new results that are used in Sections below where we consider the problem of solvability.

We will fix $\mathscr{L} = \langle L, \vee, \wedge, *, \rightarrow, 0, 1\rangle$ and a non-empty universe X. Recall that a binary fuzzy relation is a *$*$-fuzzy preorder* if it is reflexive and $*$-transitive. Fuzzy preorder $Q : X \times X \longrightarrow L$ on X can be generated by an arbitrary family of fuzzy subsets $(A_i)_{i\in I}$ of X:

$$Q(x,y) = \bigwedge_{i\in I} (A_i(x) \rightarrow A_i(y)).$$

If Q is a fuzzy preorder on X then fuzzy set $A \in L^X$ is called [6] an *upper set* of Q if

$$A(x) * Q(x,y) \leq A(y), \quad x, y \in X.$$

A necessary and sufficient condition that a family of fuzzy subsets of X constitutes a family of upper sets of Q has been proved in [6]. A new result will be proved below in Theorem 2 for a family of normal fuzzy subsets of X. Let us remark that our assumptions are different from those in [6].

Theorem 2
Let $(A_i)_{i\in I} \subseteq L^X$ be a family of normal fuzzy subsets of X and $(x_i)_{i\in I} \subseteq X$ be a family of pairwise different elements such that for all $i \in I$, $A_i(x_i) = 1$. Then the following two statements are equivalent:

(i) there exists fuzzy preorder Q on X such that for each $i \in I$, $A_i(x) = Q(x_i, x)$, $x \in X$.

(ii) For all $i, j \in I$,

$$A_i(x_j) \leq \bigwedge_{x\in X} (A_j(x) \rightarrow A_i(x)). \tag{8}$$

PROOF: It is easy to see that (i) $\Rightarrow$ (ii) so that we will prove the reverse implication. Assume that $(A_i)_{i\in I} \subseteq L^X$ is a family of normal fuzzy subsets of X, and (8) holds true. If $Q(x,y) = \bigwedge_{i\in I}(A_i(x) \rightarrow A_i(y))$ then by the assertion above, Q is a fuzzy preorder on X. We will show that statement (i) is valid for Q.

Let us choose and fix i, $i \in I$. For all $x \in X$, and arbitrary $j \in I$, such that $i \neq j$, we have:

$$Q(x_i, x) = \bigwedge_{j\in I} (A_j(x_i) \rightarrow A_j(x)) \leq$$
$$\leq A_i(x_i) \rightarrow A_i(x) = A_i(x).$$

On the other hand, by (8), $A_j(x_i) \leq A_i(x) \rightarrow A_j(x)$ or, equivalently, $A_i(x) \leq A_j(x_i) \rightarrow A_j(x)$. Therefore,

$$A_i(x) \leq \bigwedge_{j\in I} (A_j(x_i) \rightarrow A_j(x)) = Q(x_i, x).$$

Hence, for all $x \in X$, $A_i(x) = Q(x_i, x)$. □

Corollary 1
Let $(A_i)_{i\in I} \subseteq L^X$ be a family of normal fuzzy subsets of X and $(x_i)_{i\in I} \subseteq X$ be a family of pairwise different elements such that for all $i \in I$, $A_i(x_i) = 1$. Then $Q(x,y) = \bigwedge_{i\in I}(A_i(x) \rightarrow A_i(y))$ is the coarsest fuzzy preorder on X such that (8) holds true.

The following lemma gives another necessary and sufficient condition that a family of normal fuzzy subsets of X constitutes a family of upper sets of Q. We will use that condition in our new criteria of solvability.

Lemma 1
Let A_i, $i \in I$, be a family of normal fuzzy subsets of L^X, such that $A_i(x_i) = \mathbf{1}$ for the respective $x_i \in X$, $i \in I$. Moreover, let for all $i, j \in I$, inequality (8) hold true. Then inequality (8) turns to the equality

$$A_i(x_j) = \bigwedge_{x\in X} (A_j(x) \rightarrow A_i(x)). \tag{9}$$

PROOF: Assume that for all $i, j \in I$, (8) holds true, i.e. $A_i(x_j) \leq \bigwedge_{x\in X}(A_j(x) \rightarrow A_i(x))$. On the other hand,

$$\bigwedge_{x\in X} (A_j(x) \rightarrow A_i(x)) \leq A_j(x_j) \rightarrow A_i(x_j) = A_i(x_j)$$

so that (9) follows.

Assume that for all $i, j \in I$, (9) holds true. Then (8) follows immediately. □

ISBN: 978-989-95079-6-8

3 SFRE$*$ in a Space with a Fuzzy Preorder and Their Solvability

Assume that $\mathscr{L}$, X, Y are as above, and we are given fuzzy sets $A_1, \ldots, A_n \in L^X$ and $B_1, \ldots, B_n \in L^Y$. In this section we will investigate the problem of solvability of system (1) and propose a new criterion that is weaker than all criteria, based on the assumption that fuzzy sets $A_1, \ldots, A_n$, establish a fuzzy partition of X, i.e. that they are classes of a respective similarity on X.

In order to simplify denotation we will choose and fix $y \in Y$ and work with the following instance of the system:

$$A_i \circ r = b_i, \quad 1 \leq i \leq n, \tag{10}$$

where $b_i = B_i(y)$, $1 \leq i \leq n$. In this particular case, system (10) is considered with respect to unknown fuzzy set $r \in L^X$.

Theorem 3
Let fuzzy sets $A_1, \ldots, A_n \in L^X$ be normal and $x_1, \ldots, x_n \in X$ be pairwise different elements such that for all $i = 1, \ldots, n$, $A_i(x_i) = 1$. Let moreover, for all $i, j = 1, \ldots, n$,

$$A_i(x_j) = \bigwedge_{x \in X} (A_j(x) \to A_i(x)) \tag{11}$$

holds true. Then system (10) is solvable if and only if

$$A_i(x_j) \leq (b_j \to b_i). \tag{12}$$

PROOF: By Lemma 1, (11) is equivalent to

$$(\forall i, j)(\forall x \in X) \quad (A_j(x) \leq (A_i(x_j) \to A_i(x))). \tag{13}$$

Assume that system (10) is solvable. Then

$$\hat{r}(x) = \bigwedge_{j=1}^{n} (A_j(x) \to b_j)$$

is a solution so that

$$(\forall j) \quad \bigvee_{x \in X} (A_j(x) * \hat{r}(x)) \geq b_j.$$

Let us fix j, $j = 1, \ldots, n$. By (13),

$$(\forall k) \quad \bigvee_{x \in X} ((A_k(x_j) \to A_k(x)) * \hat{r}(x)) \geq b_j.$$

Then

$$\bigvee_{x \in X} ((A_k(x_j) \to A_k(x)) * (A_k(x) \to b_k)) \geq b_j$$

so that

$$A_k(x_j) \to b_k \geq b_j.$$

The last inequality is equivalent to

$$A_k(x_j) \leq b_j \to b_k,$$

and by arbitrariness of k, j, it is equivalent to (12).

On the other hand, assume that (12) holds true. We will prove that $\hat{r}$ is a solution of system (10). Let us fix i, $i = 1, \ldots, n$ and prove that $\hat{r}$ solves the i-th equation of (10). Indeed,

$$\bigvee_{x \in X} (A_i(x) * \hat{r}(x)) \leq \bigvee_{x \in X} (A_i(x) * (A_i(x) \to b_i)) \leq b_i.$$

Let us prove the opposite inequality.

$$\bigvee_{x \in X} (A_i(x) * \hat{r}(x)) \geq A_i(x_i) * \hat{r}(x_i) = \bigwedge_{j=1}^{n} (A_j(x_i) \to b_j).$$

By (12), $A_j(x_i) \to b_j \geq b_i$, so that the opposite inequality easily follows:

$$\bigvee_{x \in X} (A_i(x) * \hat{r}(x)) \geq b_i.$$

Thus, $\hat{r}$ is a solution of (10), and the system is solvable. □

4 SFRE$\to$ in a Space with a Fuzzy Preorder and Their Solvability

Assume that $\mathscr{L}$, X, Y and fuzzy sets $A_1, \ldots, A_n \in L^X$ and $B_1, \ldots, B_n \in L^Y$ are as above. In this section we will investigate the problem of solvability of system (2) and propose a new criterion. As above we will work with an instance of system (2), i.e. with the following equation:

$$A_i \triangleright r = b_i, \quad 1 \leq i \leq n, \tag{14}$$

where $b_i = B_i(y)$, $1 \leq i \leq n$. System (14) will be considered with respect to an unknown fuzzy set $r \in L^X$.

Theorem 4
Let fuzzy sets $A_1, \ldots, A_n \in L^X$ be normal and $x_1, \ldots, x_n \in X$ be pairwise different elements such that for all $i = 1, \ldots, n$, $A_i(x_i) = 1$. Let moreover, for all $i, j = 1, \ldots, n$, (11) holds true. Then system (14) is solvable if and only if

$$(\forall i, j) \quad A_i(x_j) \leq (b_i \to b_j). \tag{15}$$

PROOF: By Lemma 1, (11) is equivalent to

$$(\forall i, j)(\forall x \in X) \quad A_j(x) \leq (A_i(x_j) \to A_i(x)),$$

and can be equivalently transformed to

$$(\forall i, j)(\forall x \in X) \quad b_i * A_j(x_j) \leq (A_j(x) \to b_i * A_i(x)). \tag{16}$$

Assume that system (14) is solvable. Then

$$\check{r}(x) = \bigvee_{j=1}^{n} (A_j(x) * b_j)$$

is a solution so that

$$(\forall j) \quad \bigwedge_{x \in X} (A_j(x) \to \check{r}(x)) \leq b_j,$$

and hence

$$(\forall j) \quad \bigwedge_{x \in X} (A_j(x) \to A_i(x) * b_i) \leq b_j,$$

ISBN: 978-989-95079-6-8

where $i = 1, \ldots, n$. By (16),

$$(\forall i, j) \quad b_i * A_j(x_j) \leq b_j$$

that is equivalent to (15).

On the other hand, assume that (15) holds true. We will prove that $\check{r}$ is a solution of system (14). Let us fix i, $i = 1, \ldots, n$ and prove that $\check{r}$ solves the i-th equation of (14). First we observe that

$$\bigwedge_{x \in X} (A_i(x) \to \check{r}(x)) \geq \bigwedge_{x \in X} (A_i(x) \to A_i(x) * b_i) \geq b_i.$$

On the other hand,

$$\bigwedge_{x \in X} (A_i(x) \to \check{r}(x)) \leq A_i(x_i) \to \check{r}(x) = \bigvee_{j=1}^{n} (A_j(x_i) * b_j).$$

By (15), for all i, j, $A_j(x_i) * b_j \leq b_i$, so that

$$\bigwedge_{x \in X} (A_i(x) \to \check{r}(x)) \leq b_i.$$

Thus, $\check{r}$ is a solution of (14), and the system is solvable. □

5 Discussion

In this section, we will show how the discovered criteria relate to those existed in literature. Moreover, we will justify our claim in the Abstract that the new criteria are weaker than all known ones that are based on the assumption that fuzzy sets on the left-hand side of a system establish a fuzzy partition of a respective universe. The following criterion is among those to which we have referred to as known ones (it has been recalled in Theorem 1). The formulation below is adapted to the instance (10).

Let fuzzy sets $A_1, \ldots, A_n \in L^X$ be normal and $x_1, \ldots, x_n \in X$ be pairwise different elements such that for all $i = 1, \ldots, n$, $A_i(x_i) = 1$. Let moreover, for all $i, j = 1, \ldots, n$,

$$\bigvee_{x \in X} (A_j(x) * A_i(x)) \leq \bigwedge_{x \in X} (A_j(x) \leftrightarrow A_i(x)) \qquad (17)$$

holds true. Then (10) is solvable if and only if

$$\bigwedge_{x \in X} (A_j(x) \to A_i(x)) \leq (b_i \leftrightarrow b_j).$$

Let us examine condition (7) and compare it with (11). For simplicity, assume that $n = 2$ and sets A_1, A_2 are ordinary (not fuzzy). It is easy to see that in this case, membership functions A_1, A_2 are characteristic functions of the respective sets and $A_1(x) * A_2(x)$ is a characteristic function of the intersection $A_1 \cap A_2$. For this particular case, (7) is fulfilled if and only if either $A_1 = A_2$ or $A_1 \cap A_2 = \emptyset$. Assume that $A_1 \subseteq A_2$ and $A_1 \neq A_2$. Then (7) fails while condition (11) is valid provided that $x_2 \notin A_1$.

Therefore, in Theorems 3, 4 we have obtained new criteria of solvability that have weaker assumptions than all those that explicitly or implicitly use the condition that fuzzy sets $A_1, \ldots, A_n \in L^X$ establish a fuzzy partition of X.

Conclusion

The problem of solvability of systems of fuzzy relation equations with two different compositions has been considered. We established new solvability criteria for systems of both types. The new criteria are based on the assumption that fuzzy sets on the left-hand side of a system are upper sets of a respective $*$-fuzzy preorder. This assumption is weaker than the assumption that fuzzy sets on the left-hand side of a system establish a fuzzy partition of a respective universe.

Acknowledgment

The paper has been supported partially by the grant IAA108270902 of GA AV ČR and partially by the project 1M0572 of the MŠMT ČR.

References

[1] W. Bandler, L. Kohout, Semantics of implication operators and fuzzy relational products, Int. J. Man-Machine Studies 12 (1980) 89–116.

[2] B. De Baets, Analytical solution methods for fuzzy relation equations, in: D. Dubois, H. Prade (eds.), The Handbooks of Fuzzy Sets Series, Vol.1, Kluwer, Dordrecht, 2000, pp. 291–340.

[3] A. Di Nola, S. Sessa, W. Pedrycz, E. Sanchez, Fuzzy Relation Equations and Their Applications to Knowledge Engineering, Kluwer, Boston, 1989.

[4] S. Gottwald, On the existence of solutions of systems of fuzzy equations, Fuzzy Sets and Systems 12 (1984) 301–302.

[5] F. Klawonn, Fuzzy points, fuzzy relations and fuzzy functions, in: V. Novák, I. Perfilieva (eds.), Discovering the World with Fuzzy Logic, Springer, Berlin, 2000, pp. 431–453.

[6] H. Lai, D. Zhang, Fuzzy preorder and fuzzy topology, Fuzzy Sets and Systems 157 (2006) 1865–1885.

[7] I. Perfilieva, Fuzzy function as an approximate solution to a system of fuzzy relation equations, Fuzzy Sets and Systems 147 (2004) 363–383.

[8] I. Perfilieva, S. Gottwald, Solvability and approximate solvability of fuzzy relation equations, Int. J. of General Systems 32 (2003) 361–372.

[9] I. Perfilieva, S. Lemhke, Correct models of fuzzy if–then rules are continuous, Fuzzy Sets and Systems 157 (2006) 3188–3197.

[10] I. Perfilieva, L. Nosková, System of fuzzy relation equations with $\inf \to$ composition: complete set of solutions, Fuzzy Sets and Systems 159 (2008) 2256–2271.

[11] I. Perfilieva, V. Novák, System of fuzzy relation equations as a continuous model of if–then rules, Information Sciences 177 (2007) 3218–3227.

[12] I. Perfilieva, A. Tonis, Compatibility of systems of fuzzy relation equations, Int. J. of General Systems 29 (2000) 511–528.

[13] E. Sanchez, Resolution of composite fuzzy relation equations, Information and Control 30 (1976) 38–48.

[14] M. Štěpnička, B. De Baets, L. Nosková, On additive and multiplicative fuzzy models, in: M. Štěpnička, V. Novák, U. Bodenhofer (eds.), New Dimensions in Fuzzy Logic and Related Technologies. Int. Conference EUSFLAT, Ostrava, 2007, pp. 95–102.

ISBN: 978-989-95079-6-8

[15] X. Wang, Infinite fuzzy relational equations on a complete Brouwerian lattice, Fuzzy Sets and Systems 138 (2003) 657–666.

[16] L. A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning I, II, III, Information Sciences 8-9 (1975) 199–257, 301–357, 43–80.

ISBN: 978-989-95079-6-8

On $\mathbb{M}$-Approximative Operators and $\mathbb{M}$-Approximative Systems

Alexander Šostak
Department of Mathematics, University of Latvia
and Institute of Mathematics and Computer Sciences, University of Latvia
Riga, Latvia
Email: sostaks@latnet.lv

Abstract— *The concept of an $\mathbb{M}$-approximative system is introduced. Basic properties of the category of $\mathbb{M}$-approximative systems and in a natural way defined morphisms between them are studied. It is shown that categories related to fuzzy topology as well as categories related to rough sets can be described as special subcategories of the category of $\mathbb{M}$-approximative systems.*

Keywords— Approximative operator, approximative system, *cl*-monoid, fuzzy topology, L-topology, rough set

1 Introduction and motivation

In 1968, that is only 3 years after L. Zadeh has published his famous work *"Fuzzy Sets"*, thus laying down the principles of what can be called *Mathematics of Fuzzy Sets*, his student C.L. Chang [3] introduced the concept of a fuzzy topological space thus marking the beginning of Fuzzy Topology, the conterpart of General Topology in the context of fuzzy sets. Now Fuzzy Topology is one of the most well developed fields of Mathematics of Fuzzy Sets, and there are published dozens of fundamental works on this subject.

In 1982 Z. Pawlak [14] has introduced the concept of a rough set which can be viewed as a certain alternative for the concept of a fuzzy set for the study of mathematical problems of applied nature. Pawlak's work was followed by many other publications where rough sets and mathematical structures on the basis of rough sets were introduced, studied, and applied.

Although at the first glance it may seem that the concepts of a fuzzy set, of a (fuzzy) topological space and of a rough set are of an essentially different nature and "have nothing in common", this is not the case. Probably, the first one to start studying the intermediate relations between topologies, fuzzy sets and rough sets was J. Kortelainen [11], see also [12], etc. Further a detailed analysis of different relations between fuzzy sets, rough sets and some other related concepts was done in a series of papers by Y. Yao (e.g. [20]), and other researchers.

The aim of this work is to present an alternative view on the relations between fuzzy sets, fuzzy topological spaces and rough sets and to develop a framework allowing to generalize these concepts and corresponding theories. In order to realize this aim we introduce the concept of an $\mathbb{M}$-approximative system (cf [19]) and thus come to the category $\mathbf{AS}^{\mathbb{M}}$ of $\mathbb{M}$-approximative systems. Properties of this category are studied and connections between $\mathbf{AS}^{\mathbb{M}}$ and its subcategories related to fuzzy topology, fuzzy sets and rough sets are described.

2 The context

In our work two lattices will play the fundamental role. The first one is a complete infinitely distributive lattice

$$\mathbb{L} = (\mathbb{L}, \leq, \wedge, \vee),$$

whose top and bottom elements are $1_{\mathbb{L}}$ and $0_{\mathbb{L}}$ respectively. Besides sometimes we will assume that the lattice $\mathbb{L}$ is equipped with one of the following operations: a monotone mapping $^c : \mathbb{L} \to \mathbb{L}$ or a binary operation $* : \mathbb{L} \times \mathbb{L} \to \mathbb{L}$.

A lattice $\mathbb{L} = (\mathbb{L}, \leq, \wedge, \vee, ^c)$ will be called *adjunctive* if the pair $(^c, ^c)$ is an adjunction

$$(^c, ^c) : \mathbb{L} \vdash \mathbb{L}^{op},$$

that is $a \leq b^c \iff b \leq a^c \quad \forall\, a, b \in \mathbb{L}$, cf e.g. [4]. A lattice $\mathbb{L} = (\mathbb{L}, \leq, \wedge, \vee, ^c)$ will be called *involutive* if $^c : \mathbb{L} \to \mathbb{L}$ is an involution, that is if $(a^c)^c = a \quad \forall a \in \mathbb{L}$. One can easily see that in an adjunctive involutive lattice involution $^c : \mathbb{L} \to \mathbb{L}$ is order reversing:

$$a \leq b \implies b^c \leq a^c \; \forall a, b \in \mathbb{L},$$

and conversely, if $^c : \mathbb{L} \to \mathbb{L}$ is order reversing involution, then $(^c, ^c) : \mathbb{L} \vdash \mathbb{L}^{op}$ is an adjunction.

Concerning the second, binary operation $* : \mathbb{L} \times \mathbb{L} \to \mathbb{L}$ (conjunction) it will be assumed that $\mathbb{L} = (\mathbb{L}, \leq, \wedge, \vee, *)$ is a commutative cl-monoid (see e.g. [2]), that is
$*$ is commutative: $a * b = b * a$ for all $a, b \in \mathbb{L}$;
$*$ distributes over arbitrary joins:
$a * \left(\bigvee_{i\in\mathcal{I}} b_i\right) = \bigvee_{i\in\mathcal{I}}(a * b_i) \; \forall\, a \in \mathbb{L}, \; \forall \{b_i \mid i \in \mathcal{I}\} \subseteq \mathbb{L}$
and $a * 1_{\mathbb{L}} = a, \quad a * 0_{\mathbb{L}} = 0_{\mathbb{L}}$.

It is well-known (see e.g. [2]) that in a cl-monoid a further binary operation $\mapsto : \mathbb{L} \to \mathbb{L}$ (residuation) is defined related to conjunction $*$ by Galois connection:

$$a * b \leq c \iff a \leq b \mapsto c \; \forall a, b, c \in \mathbb{L}.$$

One can easily see that residuation is nonincreasing by the first argument and nondecreasing by the second argument, and that $b * (b \mapsto a) \leq a \; \forall a, b \in \mathbb{L}$. In particular $b * (b \mapsto 0) \leq 0$, and hence

$$b \leq (b \mapsto 0) \mapsto 0.$$

This allows to conclude, that by setting $a^c = a \mapsto 0$ we obtain an adjunction $(^c, ^c) : \mathbb{L} \vdash \mathbb{L}^{op}$. Indeed, if $a \leq b \mapsto 0$, then

$$b \leq (b \mapsto 0) \mapsto 0 \leq a \mapsto 0.$$

A cl-monoid is called a Girard monoid [10] if $(a \mapsto 0) \mapsto 0 = a \; \forall a \in \mathbb{L}$. Hence in case $\mathbb{L}$ is a Girard monoid, residuation $\mapsto$ induces an order reversing involution $^c : \mathbb{L} \to \mathbb{L}$.

An important situation in our research will be the following. Let $\mathrm{L} = (\mathrm{L}, \leq, \wedge, \vee)$ be a a lattice and X be a set. Then the L-powerset $\mathrm{L}^X =: \mathbb{L}$ becomes a lattice $(\mathbb{L}, \leq, \wedge, \vee)$ by pointwise extending the lattice structure from L to $\mathbb{L}$. Besides $\mathbb{L}$ is

ISBN: 978-989-95079-6-8

infinitely distributive whenever L was infinitely distributive. Moreover, if $\mathrm{L} = (\mathrm{L}, \leq, \wedge, \vee, ^c)$ is an adjunctive (involutive) lattice then by pointwise extending operation ^{c} from L to $\mathbb{L}$, an adjunctive (resp. involutive) lattice $\mathbb{L} = (\mathbb{L}, \leq, \wedge, \vee, ^c)$ is obtained. In case $\mathrm{L} = (\mathrm{L}, \leq, \wedge, \vee, *)$ is a cl-monoid, by pointwise extension of $* : \mathrm{L} \times \mathrm{L} \to \mathrm{L}$ to $* : \mathbb{L} \times \mathbb{L} \to \mathbb{L}$ we obtain a cl-monoid $\mathbb{L} = (\mathbb{L}, \leq, \wedge, \vee, *)$

The second lattice belonging to the context of our work is denoted by $\mathbb{M}$. At the moment we assume only its completeness, however sometimes it will be requested that $\mathbb{M}$ is completely distributive. The bottom and the top elements of $\mathbb{M}$ are $0_{\mathbb{M}}$ and $1_{\mathbb{M}}$ resp. As different from $\mathbb{L}$ we do not exclude the case when $\mathbb{M}$ is a one-point lattice and hence in this case $0_{\mathbb{M}} = 1_{\mathbb{M}}$.

3 Basic definitions

Definition 3.1 *An upper $\mathbb{M}$-approximative operator on* L *is a mapping $u : \mathbb{L} \times \mathbb{M} \to \mathbb{L}$ such that*

1. $u(0, \alpha) = 0 \; \forall \alpha \in \mathbb{M}$;
2. $a \leq u(a, \alpha) \; \forall a \in \mathbb{L}, \; \forall \alpha \in \mathbb{M}$;
3. $u(a \vee b, \alpha) = u(a, \alpha) \vee u(b, \alpha)$
4. $u(u(a, \alpha), \alpha) = u(a, \alpha)$;
5. $\alpha \leq \beta, \alpha, \beta \in \mathbb{M} \Longrightarrow u(a, \alpha) \leq u(a, \beta)$.

Operator u is called (upper) semicontinuous (usc) if

- (usc) $u(a, \bigvee_{i \in \mathcal{I}} \alpha_i) = \bigwedge_{i \in \mathcal{I}} u(a, \alpha_i)$;

u is called (upper) weakly semicontinuous (uwsc) if

- (uwsc) *If $u(a, \alpha_i) = \bar{a} \; \forall \alpha_i, i \in \mathcal{I}$ and $\alpha = \bigwedge_{i \in \mathcal{I}} \alpha_i$, then $u(a, \alpha) = \bar{a}$*

Definition 3.2 *A lower $\mathbb{M}$-approximative operator on $\mathbb{L}$ is a mapping $l : \mathbb{L} \times \mathbb{M} \to \mathbb{L}$ such that*

1. $l(1, \alpha) = 1 \; \forall \alpha \in \mathbb{M}$;
2. $a \geq l(a, \alpha) \; \forall a \in \mathbb{L}, \; \forall \alpha \in \mathbb{M}$;
3. $l(a \wedge b, \alpha) = l(a, \alpha) \wedge l(b, \alpha)$
4. $l(l(a, \alpha), \alpha) = l(a, \alpha)$;
5. $\alpha \leq \beta, \alpha, \beta \in \mathbb{M} \Longrightarrow l(a, \alpha) \geq l(a, \beta)$.

Operator l is called (lower) semicontinuous (lsc) if

- (lsc) $l(a, \bigvee_{i \in \mathcal{I}} \alpha_i) = \bigvee_{i \in \mathcal{I}} l(a, \alpha_i)$;

l is called (lower) weakly semicontinuous (lwsc) if

- (lwsc) *If $l(a, \alpha_i) = a^0 \; \forall \alpha_i, i \in \mathcal{I}$ and $\alpha = \bigvee_{i \in \mathcal{I}} \alpha_i$, then $l(a, \alpha) = a^0$*

Definition 3.3 *A triple $(\mathbb{L}, u, l)$, where $u : \mathbb{L} \times \mathbb{M} \to \mathbb{L}$ and $l : \mathbb{L} \times \mathbb{M} \to \mathbb{L}$ are upper and lower $\mathbb{M}$-approximative operators on $\mathbb{L}$, is called an $\mathbb{M}$-approximative system. In case when X is a set* L *is a lattice, $\mathbb{L} = \mathrm{L}^X$ and $(\mathbb{L}, u, l)$ is an approximative system, the quadruple (X, L, u, l) is called an $\mathbb{M}$-approximative space.*

Definition 3.4 *An $\mathbb{M}$-approximative system $(\mathbb{L}, u, l)$ is called semicontinuous (s.c) if u is u.s.c. and l is l.s.c. An $\mathbb{M}$-approximative system $(\mathbb{L}, u, l)$ is called weakly semicontinuous (w.s.c) if u is u.w.s.c. and l is l.w.s.c.*

Definition 3.5 *In case $\mathbb{L}$ is equipped with unary operation $^c : \mathbb{L} \to \mathbb{L}$, an $\mathbb{M}$-approximative system $(\mathbb{L}, u, l)$ is called self dual if*

$$u(a^c, \alpha) = (l(a, \alpha))^c \text{ and}$$
$$l(a^c, \alpha) = (u(a, \alpha))^c \; \forall a \in \mathbb{L}, \; \forall \alpha \in \mathbb{M}$$

Note that in case when $(\mathbb{L}, \leq, \wedge, \vee, ^c)$ is involutive, the system is self-dual iff $(u(a^c, \alpha))^c = l(a, \alpha)$, and $(l(a^c, \alpha))^c = u(a, \alpha), \; \forall a \in \mathbb{L}, \forall \alpha \in \mathbb{M}$;

Remark 3.6 Sometimes we consider $\mathbb{M}$-approximative systems in case of a one-point lattice $\mathbb{M} = \{\cdot\}$. Obviously, in this case the use of the second argument in the notation of approximative systems is redundant and we write just $u(a)$ and $l(a)$ instead of $u(a, \cdot)$ and $l(a, \cdot)$ respectively. Besides, in this case we use the terms *upper and lower approximative operator*, *approximative system*, etc., omitting the prefix $\mathbb{M}$.

4 Lattice of $\mathbb{M}$-approximative systems on a lattice $\mathbb{L}$

Let $\mathcal{AS}^{\mathbb{M}}(\mathbb{L})$ stand for the family of $\mathbb{M}$-approximative systems $(\mathbb{L}, u, l)$ where $\mathbb{L}$ and $\mathbb{M}$ are fixed. We introduce an order $\preceq$ on $\mathcal{AS}^{\mathbb{M}}(\mathbb{L})$ by setting $(\mathbb{L}, u_1, l_1) \preceq (\mathbb{L}, u_2, l_2)$ iff $u_1 \geq u_2$ and $l_1 \leq l_2$.

Theorem 4.1 *$(\mathcal{AS}^{\mathbb{M}}(\mathbb{L}), \preceq)$ is a complete lattice. Its top and bottom elements are given respectively by*

$$u_\top(a, \alpha) = l_\top(a, \alpha) = a \; \forall a \in \mathbb{L}, \; \forall \alpha \in \mathbb{M};$$

$$u_\bot(a, \alpha) = \begin{cases} 1_{\mathbb{L}} & \text{if } a \neq 0_{\mathbb{L}} \\ 0_{\mathbb{L}}, & \text{if } a = 0_{\mathbb{L}} \end{cases}$$

$$l_\bot(a, \alpha) = \begin{cases} 0_{\mathbb{L}} & \text{if } a \neq 1_{\mathbb{L}} \\ 1_{\mathbb{L}}, & \text{if } a = 1_{\mathbb{L}} \end{cases}$$

The infimum of a family $\{(\mathbb{L}, u_i, l_i) \mid i \in \mathcal{I}\} \subseteq \mathcal{AS}^{\mathbb{M}}(\mathbb{L})$ is given by $(u_0, l_0) = (\bigvee_i u_i, \bigwedge_i l_i)$ and its supremum $\mathbb{L}$ is given by $(u^0, l^0) = (\bigwedge_i u_i, \bigvee_i l_i)$.

Proof Let $\mathcal{S} = \{(\mathbb{L}, u_i, l_i) \mid i \in \mathcal{I}\} \subseteq \mathcal{AS}^{\mathbb{M}}(\mathbb{L})$. Let

$$\bigwedge \mathcal{S} := (\mathbb{L}, u_0, v_0) \text{ where } u_0 = \bigvee_i u_i, \; l_0 = \bigwedge_i l_i.$$

Since $u_0 \geq u_i$ and $l_0 \leq l_i$ for all $i \in \mathcal{I}$, to show the completeness of $\mathcal{AS}^{\mathbb{M}}(\mathbb{L})$, it is sufficient to show that u_0 and l_0 are respectively the upper and lower $\mathbb{M}$-approximative operators on $\mathbb{L}$. However this can be established by the direct verification of the conditions (1)-(5) in definitions 3.1, 3.2.

Further, let

$$\bigvee \mathcal{S} = (\mathbb{L}, u^0, l^0) \text{ where } u^0 = \bigwedge_i u_i, l^0 = \bigvee_i l_i.$$

To show that (u^0, l^0) is the supremum of $\mathcal{S}$ it is sufficient to notice that u^0 and l^0 are resp. the upper and lower $\mathbb{M}$-approximative operators. However the validity of properties

ISBN: 978-989-95079-6-8

(1),(2),(4),(5) in definitions 3.2, 3.1 is obvious while the validity of property (3) can be established referring to the distributivity of $\mathbb{L}$.

One can easily establish also the following

Theorem 4.2 *1. The family $(SC\mathcal{AS}^{\mathbb{M}}(\mathbb{L}), \preceq)$ of semicontinuous $\mathbb{M}$-approximative systems is a complete sublattice of $\mathcal{AS}^{\mathbb{M}}(\mathbb{L})$.*

2. The family $(WSC\mathcal{AS}^{\mathbb{M}}(\mathbb{L}), \preceq)$ of weakly semicontinuous $\mathbb{M}$-approximative systems is a complete sublattice of $\mathcal{AS}^{\mathbb{M}}(\mathbb{L})$.

3. The family $(SD\mathcal{AS}^{\mathbb{M}}(\mathbb{L}), \preceq)$ of self-dual $\mathbb{M}$-approximative systems is a complete sublattice of $\mathcal{AS}^{\mathbb{M}}(\mathbb{L})$.

5 Category $\mathbf{AS}^{\mathbb{M}}$ of $\mathbb{M}$-approximative systems

Let $\mathbb{M}$ be fixed and let $\mathbf{AS}^{\mathbb{M}}$ be the family of all $\mathbb{M}$-approximative systems $(\mathbb{L}, u, v)$. To consider $\mathbf{AS}^{\mathbb{M}}$ as a category whose class of objects are all $\mathbb{M}$-approximative systems we have to specify its morphisms. Given $(\mathbb{L}_1, u_1, l_1), (\mathbb{L}_2, u_2, l_2) \in \mathcal{O}b(\mathbf{AS}^{\mathbb{M}})$ by a morphism

$$f : (\mathbb{L}_1, u_1, l_1) \to (\mathbb{L}_2, u_2, l_2)$$

we call a mapping $f : \mathbb{L}_2 \to \mathbb{L}_1$ such that

1. $f : \mathbb{L}_1 \to \mathbb{L}_2$ is a morphism in the category $\mathbf{LAT^{op}}$ where $\mathbf{LAT}$ is the category of complete infinitely distributive lattices;
2. $u_1(f(b), \alpha) \leq f(u_2(b, \alpha))\ \forall b \in \mathbb{L}_2, \forall \alpha \in \mathbb{M}$;
3. $f(l_2(b, \alpha)) \leq l_1(f(b), \alpha)\ \forall b \in \mathbb{L}_2, \forall \alpha \in \mathbb{M}$

A morphsim $f : (\mathbb{L}_1, u_1, l_1) \to (\mathbb{L}_2, u_2, l_2)$ is aslo referred to as a continuous mapping between the corresponding $\mathbb{M}$-approximative systems

Theorem 5.1 $\mathbf{AS}^{\mathbb{M}}$ *thus obtain is indeed a category.*

Proof Let $f : (\mathbb{L}_1, u_1, l_1) \to (\mathbb{L}_2, u_2, l_2)$ and $g : (\mathbb{L}_2, u_2, l_2) \to (\mathbb{L}_3, u_3, l_3)$ be continuous mappings and let $g \circ f : \mathbb{L}_1 \to \mathbb{L}_3$ be their composition in $\mathbf{LAT^{op}}$. We have to verify that $g \circ f$ satisfies conditions (2) and (3) above. Since it is sufficient to verify these conditions for a fixed $\alpha \in \mathbb{M}$, to simplify the reasonings we omit the second argument in the notation of the approximative operators. Let $c \in \mathbb{L}_3$. Then

$$u_1(f(g(c))) \leq f(u_2(g(c))) \leq f(g(u_3(c))),$$

In a similar way we can show that $f(g(l_3(c))) \leq l_1(g(f(c)))$. Thus the composition $g \circ f : (\mathbb{L}_1, u_1, l_1) \to (\mathbb{L}_3, u_3, l_3)$ is continuous whenever f and g are continuous. We conclude the proof noticing that the identity mapping $f : (\mathbb{L}, u, l) \to (\mathbb{L}, u, l)$ is continuous.
□

In the sequel, when discussing categorical properties of $\mathbf{AS}^{\mathbb{M}}$ and other categories we refer to the monograph [1].

Theorem 5.2 *Every source $f_i : \mathbb{L}_1 \to (\mathbb{L}_i, u_i, l_i), i \in \mathcal{I}$ has a unique initial lift $f_i : (\mathbb{L}_1, u_1, l_1) \to (\mathbb{L}_2, u_2, l_2)$.*

Proof Taking into account Theorem 4.1 it is sufficient to consider the case when the source contains only one morphism $f : \mathbb{L}_1 \to (\mathbb{L}_2, u_2, l_2)$ in $\mathbf{LAT}^{op}$.

Define upper approximative operator $u_1 : \mathbb{L}_1 \times \mathbb{M} \to \mathbb{L}_1$ by

$$u_1(a, \alpha) = \bigwedge \{f(u_2(b, \alpha)) \mid f(b) \geq a\}\ \forall a \in \mathbb{L}_1, \alpha \in \mathbb{M}.$$

Note first that the condition

$$u_1(f(b), \alpha) \leq f(u_2(b, \alpha))\ \forall b \in \mathbb{L}_2\ \forall \alpha \in \mathbb{M}$$

is obviously fulfilled. We verify that u_1 thus defined is indeed an upper approximative operator. As in the previous theorem in our reasoning we fix $\alpha \in \mathbb{M}$ and omit it in notation of approximative operators when verifying the properties (1) - (4).

The first two properties are obvious: $u_1(0_{\mathbb{L}_1}) = 0_{\mathbb{L}_1}$; $u_1(a) \geq a \forall a \in \mathbb{L}_1$.
To verify property (3) let $a_1, a_2 \in \mathbb{L}_1$, then

$$u_1(a_1 \vee a_2) = \bigwedge\{f(u_2(b)) \mid f(b) \geq a_1 \vee a_2\} \leq \bigwedge\{f(u_2(b_1 \vee b_2)) f(b) \geq a_1, f(b_2) \geq a_2\} = \bigwedge\{f(u_2(b_1)) \vee f(u_2(b_2)) \mid f(b_1) \geq a_1, f(b_2) \geq a_2\} = \bigvee_{i=1,2} (\bigwedge\{f(u_2(b_i)) \mid f(b_i) \geq a_i\}) = u_1(a_1) \vee u_1(a_2).$$

The converse inequality is obvious.

To verify the fourth condition notice that $u_1(u_1(a)) = u_1(\bigwedge\{f(u_2(b)) \mid f(b) \geq a\}) \leq \bigwedge\{u_1(f(u_2(b))) \mid f(b) \geq a\} \leq \bigwedge\{f(u_2(u_2(b))) \mid f(b) \geq a\} = u_1(a)$. The converse inequality is obvious

To verify property (5) for u_1 note that

$$\alpha \leq \beta, \alpha, \beta \in \mathbb{M} \Longrightarrow u_1(a, \alpha) \leq u_1(a, \beta)$$

is guaranted by the analogous property of the operator $u_2 : \mathbb{L}_2 \times \mathbb{M} \to \mathbb{L}_2$ and the definition of u_1.

Define lower $\mathbb{M}$-approximative operator $l_1 : \mathbb{L}_1 \times \mathbb{M} \to \mathbb{L}_1$ by

$$l_1(a, \alpha) = \bigvee\{f(l_2(b), \alpha) \mid f(b) \leq a\}\ \forall a \in \mathbb{L}_1 \forall \alpha \in \mathbb{M}.$$

Notice first that

$$f(l_2(b, \alpha)) \leq l_1(f(b), \alpha)\ \forall b \in \mathbb{L}_2, \alpha \in \mathbb{M}.$$

We show that $l_1 : \mathbb{L}_1 \times \mathbb{M} \to \mathbb{L}_1$ thus defined is an lower $\mathbb{M}$-approximative operator. Again, we omit in notation α when it is fixed. The first two conditions from Definition 3.2 are obvious. To verify the third condition let $a_1, a_2 \in \mathbb{L}_1$. Then $l_1(a) \wedge l_1(a_2) = \bigvee\{f(l_1(b_1)) \wedge f(l_1(b_2)) \mid f(l_1(b_i)) \leq a_i, i = 1, 2\} \leq \bigvee\{f(l_1(b_1 \wedge b_2) \mid f(b_1) \wedge f(b_2) \leq a_1 \wedge a_2\} = \bigvee\{f(b) \mid f(b) \leq a_1 \wedge a_2\} = l_1(a_1 \wedge a_2)$,
The converse inequality is obvious, The idempotence of the operator $l_1 : \mathbb{L}_1 \to \mathbb{L}_1$ is establihed as follows:
$l_1(l_1(a)) = l_1(\bigvee\{f(l_2(b)) \mid f(b) \leq a\}) \geq \bigvee\{l_1 f(l_2(b)) \mid f(b) \leq a\} \geq \bigvee\{f(l_2(b)) \mid f(b) \leq a\} = l_1(a)$.
The opposite inequality is obvious.

Finaly, the condition $\alpha \leq \beta, \alpha, \beta \in \mathbb{M} \Longrightarrow l_1(a, \alpha) \geq l_1(a, \beta)$ is guaranted by the analogous property of the operator $l_2 : \mathbb{L}_2 \to \mathbb{L}_2$ and the definition of l_1.

Let $g : (\mathbb{L}_3, u_3, l_3) \to (L_2, u_2, l_2)$ be a morphism in $\mathbf{AS}^{\mathbb{M}}$ and $h : \mathbb{L}_3 \to \mathbb{L}_1$ be a morphism in $\mathbf{LAT}^{op}$ such that $f \circ h = g$.

ISBN: 978-989-95079-6-8

Then from the construction it is clear that $h : (\mathbb{L}_3, u_3, l_3) \to (L_1, u_1, l_1)$ is a morphism in $\mathbf{AS}^{\mathbb{M}}$. Thus $f : (\mathbb{L}_1, u_1, l_1) \to (\mathbb{L}_2, u_2, l_2)$ is indeed the initial lift of $f : \mathbb{L}_1 \to (\mathbb{L}_2, u_2, l_2)$. The uniqueness of the lift is obvious.

Let $A\mathcal{I}\mathbf{LAT}$ denote the category of adjunctive involutive complete infinitely distributive systems.

Theorem 5.3 *Let $\mathbb{L}_1, \mathbb{L}_2$ be adjunctive involutive lattices and $f : \mathbb{L}_1 \to \mathbb{L}_2$ be a morphism in $A\mathcal{I}\mathbf{LAT}^{op}$. If $\mathbb{M}$-approximative operators $l_2 : u_2 : \mathbb{L}_2 \times \mathbb{M} \to \mathbb{L}_2$ are self-dual, then $\mathbb{M}$-approximative operators $l_1, u_1 : \mathbb{L}_1 \times \mathbb{M} \to \mathbb{L}_1$ constructed above are self-dual as well.*

Indeed, let $a \in \mathbb{L}$. Then
$l_1(a^c) = \bigvee\{f(l_2(b)) \mid f(b) \le a^c\} = (\bigwedge\{(f(l_2(b)))^c \mid f(b) \le a^c\})^c = (\bigwedge\{f(u_2(b^c)) \mid f(b) \le a^c\})^c = (\bigwedge\{f(u_2(b^c)) \mid f(b^c) \ge a\})^c = (\bigwedge\{f(u_2(d)) \mid f(d) \ge a\})^c = (u_1(a))^c$.

Theorem 5.4 *If approximative system (u_2, l_2) on $\mathbb{L}_2$ is semi-continuous (weakly semicontinuous), then the approximation system (u_1, l_1) constructed in the previous theorem is semi-continuous (resp. weakly semicontinuous), too.*

Indeed, if (u_2, l_2) is semicontinuous, then $u_1(a, \bigvee_i \alpha_i) = \bigwedge f(u_2(b, \bigvee_i \alpha_i) \mid f(b) \ge a\} = \bigwedge_i \bigwedge\{f(u_2(b, \alpha_i) \mid f(b) \ge a\} = \bigwedge_i u_1(a, \alpha_i)$, and $l_1(a, \bigvee_i \alpha_i) = \bigvee\{f(l_2(b, \bigvee_i \alpha_i) \mid f(b) \le a\} = \bigvee_i \bigvee\{f(l_2(b, \alpha_i) \mid f(b) \le a\} = \bigvee_i l_1(b, \alpha_i)$.
In a similar way one can establish weak semicontinuity of (u_1, l_1) in case (u_2, l_2) was weakly semicontinuous.

Theorem 5.5 *Every sink $f_i : (\mathbb{L}_i, u_i, l_i) \to \mathbb{L}_2,\ i \in \mathcal{I}$ has a unique final lift: $f_i : (\mathbb{L}_i, u_i, l_i) \to (\mathbb{L}_2, u_2, l_2)\ i \in \mathcal{I}$.*

Proof . Taking into account Theorem 4.1 it is sufficient to consider the case of the sink consisting of a single morphism $f : (\mathbb{L}_1, u_1, l_1) \to \mathbb{L}_2$. We define an upper $\mathbb{M}$-approximative operator $u_2 : \mathbb{L}_2 \times \mathbb{M} \to \mathbb{L}_2$ by:

$$u_2(b, \alpha) = \bigwedge\{c \in \mathbb{L}_2 \mid c \ge b, f(c) \ge u_1(f(b), \alpha)\}.$$

It is obvious that $u_1(f(b), \alpha) \le f(u_2(b, \alpha))\ \forall\ b \in \mathbb{L}_2$. We show that $u_2 : \mathbb{L}_2 \times \mathbb{M} \to \mathbb{L}_2$ is an upper $\mathbb{M}$-approximative operator. We omit notation α when it can be fixed.

It is obvious that $u_2(0) = 0$ and $b \le u_2(b)$ for every $b \in \mathbb{L}_2$. Further, let $b_1, b_2 \in \mathbb{L}_2$. Then $u_2(b_1) \vee u_2(b_2) = (\bigwedge\{c_1 \in \mathbb{L}_2 \mid c_1 \ge b_1, f(c_1) \ge u_1(f(b_1))\}) \vee (\bigwedge\{c_2 \in \mathbb{L}_2 \mid c_2 \ge b_2, f(c_2) \ge u_1(f(b_2))\}) = \bigwedge\{c_1 \vee c_2 \mid c_i \ge b_i, f(c_i) \ge u_1(f(b_i)), i = 1, 2\} \ge \bigwedge\{c \mid c \le b_1 \vee b_2, f(c) \ge b_1 \vee b_2, f(c) \ge u_1(f(b_1)) \vee u_1(f(b_2))\} = \bigwedge\{c \mid c \le b_1 \vee b_2, f(c) \ge b_1 \vee b_2, f(c) \ge u_1(f(b_1 \vee b_2))\} = u_2(b_1 \vee b_2)$.
The opposite inequality is obvious.

$u_2(u_2(b)) = \bigwedge\{c \mid c \ge u_2(b), u_1(f(u_2(b))) \le f(c)\}$. Noticing that $u_2(b)$ is among the elements c satisfying the above conditions, we conclude that $u_2(u_2(b)) \le u_2(b)$. The opposite inequality is obvious and hence $u_2(u_2(b)) = u_2(b)$.

Property (5) for u_2 is guaranted by the analogous property of the operator $u_1 : \mathbb{L}_1 \times \mathbb{M} \to \mathbb{L}_1$ and the definition of u_2.

Define lower approximation operator $l_2 : \mathbb{L}_2 \times \mathbb{M} \to \mathbb{L}_2$ by

$$l_2(b, \alpha) = \bigvee\{c \in \mathbb{L}_2 \mid c \le b, l_1(f(c), \alpha) \le f(b)\}.$$

The validity of the first two conditions for $l_2 : \mathbb{L}_2 \times \mathbb{M} \to \mathbb{L}_2$ is obvious. To verify the third property let $b_1, b_2 \in \mathbb{L}_2$. Then $l_2(b_1) \wedge l_2(b_2) = \bigvee\{c_1 \wedge c_2 \mid c_1 \le b_1, c_2 \le b_2, l_1(f(c_1) \le f(b_1), l_1(f(c_2)) \le f(b_2)\} \le \bigvee\{c_1 \wedge c_2 \mid c_1 \wedge c_2 \le b_1 \wedge b_2,\ l_1(f(c_1)) \wedge l_1(f(c_2)) \le f(b_1) \wedge f(b_2)\} \le \bigvee\{c \mid c \le b_1 \wedge b_2, l_1(f(c)) \le f(b_1 \wedge b_2)\} = l_2(b_1 \wedge b_2)$.
The opposite inequality is obvious.

To show the fourth axiom note that

$$l_2(l_2(b)) = \bigvee\{c \mid c \le l_2(b), l_1(f(c)) \le f(l_2(b))\}$$

and since $l_2(b)$ is one of c appearing in the above formula, it holds $l_2(l_2(b)) \ge l_2(b)$. The converse inequality is obvious.

Property (5) for l_2 is guaranted by the analogous property of the operator $l_1 : \mathbb{L}_1 \times \mathbb{M} \to \mathbb{L}_2$ and the definition of l_2.

Theorem 5.6 *If approximation system (u_1, l_1) on $\mathbb{L}_1$ is self-dual, and $f : \mathbb{L}_1 \to \mathbb{L}_2$ is a morphism in the category of adjunkctive involutive lattices, then the approximation system (u_2, l_2) constructed in the previous theorem is self-dual.*

Indeeed, given $b \in \mathbb{L}_2$, we have
$(l_2(b))^c = (\bigvee\{d \mid d \le b, l_1(f(d)) \le f(b)\})^c = \bigwedge\{d^c \mid d \le b,\ l_1(f(d)) \le f(b)\} = \bigwedge\{d^c \mid d^c \ge b^c, (l_1(f(d)))^c \ge f(b^c)\} = \bigwedge\{d^c \mid d^c \ge b^c,\ u_1((f(d))^c) \ge f(b^c)\} = \bigwedge\{d^c \mid d^c \ge b^c,\ u_1((f(d^c))) \ge f(b^c)\} = \bigwedge\{e \mid e \ge b^c,\ u_1(f(e)) \ge f(b^c)\} = u_2(b^c)$.

One can easily establish also the following

Theorem 5.7 *If $\mathbb{M}$-approximation system (u_1, l_1) on $\mathbb{L}_1$ is semicontinuous (weakly semicontinuous), then the approximation system (u_2, l_2) constructed in the previous theorem is semicontinuous (resp. weakly semicontinuous).*

From theorems 5.2, 5.5, we obtain the following important

Corollary 5.8 *Category $\mathbf{AS}^{\mathbb{M}}$ is topological over the category $\mathbf{LAT}^{op}$ lattices with respect to the forgetful functor $\mathfrak{F} : \mathbf{AS}^{\mathbb{M}} \to \mathbf{LAT}^{op}$.*

Besides, taking into account theorems 5.3, 5.6, 5.4, 5.7 we have

Corollary 5.9 *The category $SD\mathbf{AS}^{\mathbb{M}}$ of self-dual $\mathbb{M}$-approximative systems is topological over the category $A\mathcal{I}\mathbf{LAT}^{op}$ with respect to the forgetful functor $\mathfrak{F} : SD\mathbf{AS}^{\mathbb{M}} \to A\mathcal{I}\mathbf{LAT}^{op}$.*

Corollary 5.10 *The categories $SC\mathbf{AS}^{\mathbb{M}}$ and $WSC\mathbf{AS}^{\mathbb{M}}$ of semicontinuous and weakly semicontinuous $\mathbb{M}$-approximative systems are topological over the category $\mathbf{LAT}^{op}$ with respect to the forgetful functor $\mathfrak{F} : (W)SC\mathbf{AS}^{\mathbb{M}} \to \mathbf{LAT}^{op}$.*

6 Categories $\mathbf{AS}^{\mathbb{M}}(\mathrm{L})$ of $\mathbb{M}$-approximative L-spaces

An important subcategory of the category $\mathbf{AS}^{\mathbb{M}}$ is the category whose ojects are L-powersets L^X of arbitrary sets (where L is a fixed lattice) and whose morphisms are induced by mappings of the corresponding sets X. Here are the details:

Let L be a fixed complete infinitely distributive lattice and let the objects of $\mathbf{AS}^{\mathbb{M}}(\mathrm{L})$ be approximation systems of the

ISBN: 978-989-95079-6-8

form (L^X, u, l) where L^X are L-powersets of arbitrary sets X. Sometimes it is more convenient to interpret objects of this category as the corresponding quadruples (X, L, u, l). To define a morphism $F : (\mathrm{L}^{X_1}, u_1, l_1) \to (\mathrm{L}^{X_2}, u_2, l_2)$ consider a mapping $f : X_1 \to X_2$ and let $f^{\leftarrow} : \mathrm{L}^{X_2} \to \mathrm{L}^{X_1}$ be the backward powerset operator induced by f [16]. Now as morphisms in $\mathbf{AS}^{\mathbb{M}}(\mathrm{L})$ we take $F := f^{\leftarrow^{op}} : (\mathrm{L}^{X_1}, u_1, l_1) \to (\mathrm{L}^{X_2}, u_2, l_2)$ in case it is a morphism in the category $\mathbf{AS}^{\mathbb{M}}$.

An important special case is a two point lattice $\mathrm{L} = \mathbf{2}$: in this case we come to the category of $\mathbb{M}$-approximative structures on ordinary sets (of course, for this one has to interpret a subset A of a set X as the characteristic function $\chi_A : X \to \mathbf{2}$). In particular, if $\mathbb{M}$ is a one-point lattice we come to the concept of an approximation system as it was considered by some authors, see e.g. [20].

7 Categories of fuzzy topologies as subcategories of $\mathbf{AS}^{\mathbb{M}}$

7.1 Category of $(\mathrm{L}, \mathbb{M})$*-fuzzy topological spaces*

We start with interpreting the category **FTOP**$(\mathrm{L}, \mathbb{M})$ of $(\mathrm{L}, \mathbb{M})$-fuzzy topological spaces see, e.g. [17], [13], [18], [7] as a subcategory of $\mathbf{AS}^{\mathbb{M}}$. In this section $\mathbb{M}$ is assumed to be completely distributive.

Definition 7.1 *A mapping $\mathcal{T} : \mathrm{L}^X \to \mathbb{M}$ is an $(\mathrm{L}, \mathbb{M})$-fuzzy topology on X if*

1. $\mathcal{T}(0_X) = \mathcal{T}(1_X) = 1$;
2. $\mathcal{T}(U \wedge V) \geq \mathcal{T}(U) \wedge \mathcal{T}(V) \forall U, V \in \mathrm{L}^X$;
3. $\mathcal{T}\left(\bigvee_{i\in\mathcal{I}} U_i\right) \geq \bigwedge_{i\in\mathcal{I}} \mathcal{T}(U_i) \ \ \forall\{U_i \mid i \in \mathcal{I}\} \subseteq \mathrm{L}^X$

A pair $(X, \mathcal{T})$ is called an $(\mathrm{L}, \mathbb{M})$-fuzzy topological space and the value $\mathcal{T}(U), U \in \mathrm{L}^X$ is interpreted as the degree of openess of a fuzzy set U. A mapping $f : (X, \mathcal{T}_X) \to (Y, \mathcal{T}_Y)$ is called continuous if $\mathcal{T}_X\left(f^{-1}(V)\right) \geq \mathcal{T}_Y(V)) \ \ \forall V \in \mathrm{L}^Y$.

Let $(X, \mathcal{T})$ be an $(\mathrm{L}, \mathbb{M})$-fuzzy topological space. By setting

$$\mathrm{int}_{\mathcal{T}}(A, \alpha) = \bigvee\{U \in \mathrm{L}^X \mid U \leq A, \ \mathcal{T}(U) \geq \alpha\},$$

we define the s.c. interior operator $\mathrm{int}_{\mathcal{T}} : \mathrm{L}^X \times \mathbb{M} \to \mathrm{L}^X$.

The relations between (L, M)-fuzzy topologies and lower $\mathbb{M}$-approximative operators are described in the theorem;

Theorem 7.2 *The interior operator* int *is a weakly s.c. lower $\mathbb{M}$-approximative operator on $\mathbb{L} = \mathrm{L}^X$. Conversely, if $l : \mathrm{L}^X \times \mathbb{M} \to \mathrm{L}^X$ is a weakly s.c. lower $\mathbb{M}$-approximative operator, then by setting*

$$T_l(U) = \bigvee\{\alpha \mid l(U, \alpha) \geq U\}$$

we obtain a mapping $\mathcal{T}_l : \mathrm{L}^X \to \mathrm{L}^X$ satisfying conditions (1) and (3) of Definition 7.1. Besides $\mathcal{T}_{\mathrm{int}_{\mathcal{T}}} = \mathcal{T}$ and $l_{\mathcal{T}_l} = l$.

Further, assume that L is an adjunctive involutive lattice and let $^c : \mathrm{L} \to \mathrm{L}$ be the corresponding involution. Then by setting

$$cl_{\mathcal{T}}(A, \alpha) = \bigwedge\{B \mid B \geq A, \mathcal{T}(B^c) \geq \alpha\}$$

a closure operator $cl_{\mathcal{T}} : \mathrm{L}^X \times \mathbb{M} \to \mathrm{L}^X$ is defined. One can easily show that $cl_{\mathcal{T}}$ is a weakly s.c. upper $\mathbb{M}$-approximative operator and prove a theorem establishing relations between weakly s.c. upper $\mathbb{M}$-approximative operators and $(\mathrm{L}, \mathbb{M})$-fuzzy topologies via closure operators, analogous to Theorem 7.2. Besides the $\mathbb{M}$-approximation system $(\mathrm{L}^X, cl_{\mathcal{T}}, int_{\mathcal{T}})$ is self-dual.

Thus in case of an involutive adjunctive lattice L an $(\mathrm{L}, \mathbb{M})$-fuzzy topological space $(X, \mathcal{T})$ can be interpreted as a weakly s.c. $\mathbb{M}$-approximative self-dual system $(\mathbb{L}, \mathrm{cl}, \mathrm{int})$ where $\mathbb{L} = \mathrm{L}^X$.

This allows us to identify the category $\mathbf{FTOP}(\mathrm{L}, \mathbb{M})$ with the subcategory $\mathbf{TopAS}^{\mathbb{M}}(\mathrm{L})$ of the category $\mathbf{AS}^{\mathbb{M}}$ whose objects are self dual weakly s.c. $\mathbb{M}$-approximative systems of the form $(\mathrm{L}^X, \mathrm{int}, \mathrm{cl})$ and the morphisms are $F = f^{\leftarrow^{op}} : (\mathrm{L}^X, \mathrm{int}_X, \mathrm{cl}_X) \to (\mathrm{L}^Y, \mathrm{int}_Y, \mathrm{cl}_Y)$, where $f^{\leftarrow} : \mathrm{L}^Y \to \mathrm{L}^X$ are powerset operators induced by (see e.g. [9]) by coninuous mappings $f : (X, \mathcal{T}_X) \to (Y, \mathcal{T}_Y)$ (cf also the previous section).

7.2 Category of Chang-Goguen L*-topological space*

To obtain characterization of L-topological spaces (see [3], [5]) by means of approximative systems we can restrict the theory developed in 7.1 by taking the two-point lattice $\mathbf{2} = \{0, 1\}$ in the role of $\mathbb{M}$. Then the category of Chang-Goguen L-topological spaces can be identified with the subcategory $\mathbf{TopAS}(\mathrm{L})$ of the category $\mathbf{TopAS^2}(\mathrm{L})$. In particular $\mathbf{TopAS^2}(\mathbf{2})$ can be identified with the classical category **TOP** of ordinary topological spaces and continuous mappings.

7.3 Category of L*-fuzzifying topological space*

To obtain characterization of L-fuzzifying topological spaces, originally defined by U.Hohle [6] and then independently discovered by Mingsheng Ying [21], by means of approximative systems we restrict the theory developed in 7.1 by taking the two-point lattice $\mathbf{2}$ in the role of L (thus $\mathbb{L} = 2^X$) and the lattice L in the role of $\mathbb{M}$. Then the category of L-fuzzifying topological spaces can be identified with the category $\mathbf{TopAS}^{\mathrm{L}}(\mathbf{2})$.

7.4 Category of Hutton fuzzy topological spaces

According to B. Hutton [8], a fuzzy topological space is a pair (L, τ) where L is a completely distributive lattice and $\tau \subseteq \mathrm{L}$ such that $0, 1 \in \tau$; $\ a, b \in \tau \Longrightarrow a \wedge b \in \tau$: $\ a_i \in \tau \forall i \in \mathcal{I} \Longrightarrow \bigvee_{i\in\mathcal{I}} a_i \in \tau$. The morphisms $f : (L_1, \tau_1) \to (L_2, \tau_2)$ in the category **H-TOP** of Hutton fuzzy topological spaces are mappings $f : L_2 \to L_1$ such that $f(\tau_2) \subseteq \tau_1$. One can show that the category **H-TOP** can be identified with the subcategory **HAS** of the category $\mathbf{AS}^{\mathbb{M}}$. whose objects are self dual approximative systems $(\mathbb{L}, l, u)$ where $\mathbb{L}$ is a completely distributive involutive adjunctive lattice and $\mathbb{M} = \{\cdot\}$

7.5 Category of variable basis fuzzy topological spaces

In [15] S.E. Rodabaugh has introduced the notion of a variable-basis fuzzy topological space and defined the corresponding category **R-TOP**. Further the theory of variable basis fuzzy topological spaces and some related categories was developed in a series of papers by S.E. Rodabaugh, P. Eklund and other authors. The category of variable-basis fuzzy topological spaces also can be obtained as a subcategory of the category $\mathbf{AS}^{\mathbb{M}}$. However to describe it in this way and to give an

ISBN: 978-989-95079-6-8

explicite characterization by means of $\mathbb{M}$-approximative systems we need more space than it is allowed here.

8 Categories related to rough sets

8.1 Rough sets

Let $\rho \subseteq X \times X$ be a binary relation on a set X and let $R(x) = \{x' \mid x\rho x'\}$ be the right ρ-class of $x \in X$. Given $A \in 2^X$ let $l(A) = A^{\blacktriangledown} = \{x \mid R(x) \subseteq A\}$, $u(A) = A^{\blacktriangle} = \{x \mid R(x) \cap A \neq \emptyset\}$. In case ρ is reflexive and transitive $u : 2^X \to 2^X$ and $l : 2^X \to 2^X$ are, respectively, upper and lower approximative operators on $2^X = \mathcal{P}(X)$ and $(X, 2, ^{\blacktriangledown}, ^{\blacktriangle})$ is an approximative space. Besides, one can easily see that the system $(\mathbf{2}^X, ^{\blacktriangledown}, ^{\blacktriangle})$ is self dual: $A^{c\blacktriangle} = A^{\blacktriangledown c}$) for any $A \subseteq X$. Such operators and corresponding approximative spaces in case when ρ is an equivalence relation were introduced by Pawlak [14] under the name "rough set". Further approximative operators operators induced by binary relations, either general or satisfying special properties, were studied by different authors, see e.g. [11], [12], [20], etc. Note however, that in case ρ is not reflexive or transitive, this operators may fail to be approximative operators in our sense.

In case when ρ is only reflexive, J. Järvinen and J. Kortelainen [12] along with operators $A^{\blacktriangledown}$ and $A^{\blacktriangle}$ consider also operators $u'(A) = A^{\triangle} = \{x \mid R^{-1}(x) \cap A \neq \emptyset\}$, $l'(A) = A^{\triangledown} = \{x \mid R^{-1}(x) \subseteq A\}$ and show that (u, l') and (u', l) form Galois connection: $u(a) \leq b \iff a \leq l'(b)$; $\quad l(a) \leq b \iff a \leq u'(b)$. Thus in case ρ is also transitive, we obtain "Galoi-connected" approximative systems $(\mathbf{2}^X, ^{\blacktriangledown}, ^{\triangle})$ and $(\mathbf{2}^X, ^{\triangledown}, ^{\blacktriangle})$.

8.2 L*-rough sets*

Generalizing the previous situation let L be a cl-monoid $(\mathrm{L}, \wedge, \vee, *)$, X be a set and $\rho : X \times X \to \mathrm{L}$ be an L-relation on X. Further, assume that ρ is reflexive (that is $\rho(x, x) = 1 \ \forall x \in X$) and transitive (that is $\rho(x, y) * \rho(y, z) \leq \rho(x, z) \ \forall x, y, z \in X$.) For every $x \in X$ we define $\mathcal{R}(x) : X \to \mathrm{L}$ by $\mathcal{R}(x)(x') = \rho(x, x')$ Further, given $A \in \mathrm{L}^X$ let lower and upper approximative operators $l(A) \in \mathrm{L}^X$, and $u(A) \in \mathrm{L}^X$ be defined by $l(A)(x) = \inf_{x' \in X}(\mathcal{R}(x)(x') \mapsto A(x'))$ and $u(A)(x) = \sup_{x' \in X}(\mathcal{R}(x)(x') * A(x'))$ respectively. One can show that (L^X, u, l) is an L-approximative system. We refer to such kind of an approximative system as an L-rough system induced by the L-relation ρ. In case $(\mathrm{L}, \wedge, \vee, *)$ is a Girard monoid, the system (L^X, u, l) is self dual. Further, if $\mathrm{L} = \mathbf{2}$ is a two-point lattice we come to the situation described in the previous subsection. In an natural way we define morphisms for the category $\mathbf{Rgh}(\mathbb{L})$ of L- rough systems and characterize it as a category of approximative systems.

9 Defuzzification approximation operators

Finally we sketch how the concept of an approximative systems can be applied for fuzzy sets themselves.

Let $\mathrm{L} = (\mathrm{L}, \wedge, \vee, \leq)$ be a complete lattice, X be a set and $\mathbb{L} = \mathrm{L}^X$ Define $u : \mathbb{L} \times \mathrm{L} \to \mathbb{L}$ and $l : \mathbb{L} \times \mathrm{L} \to \mathbb{L}$ as follows: Given $A \in \mathrm{L}^X$ let

$$u(A, \alpha) = A \vee 1_{A_\alpha}, \ l(A, \alpha) = \alpha \cdot 1_{A_\alpha} \wedge A,$$

where $A_\alpha = \{x \in X \mid A(x) \geq \alpha\}$ In this way we obtain L-approximation operators on the L-powerset of a set X which can be interpreted as resp. upper and lower level-defuzzification operators .

References

[1] J. Adámek, H. Herrlich and G.E. Strecker, **Abstract and Concrete Categories**, John Wiley & Sons, New York 1990.

[2] G. Birkhoff, **Lattice Theory**, AMS Providence, RI, 1995.

[3] C.L. Chang, *Fuzzy topological spaces*, J. Math. Anal. Appl., **24** (1968), 182-190.

[4] G. Gierz, K.H. Hofmann, K. Keimel, J.D. Lawson, M.W. Mislove, D.S. Scott, **Continuous Lattices and Domains**, Cambridge University Press, Cambridge, 2003.

[5] J.A. Goguen, *The fuzzy Tychonoff theorem*, J. Math. Anal. Appl., **18** (1973), 734-742.

[6] U. Höhle, *Upper semicontinuous fuzzy sets and applications*, J. Math. Anal. Appl., **78** (1980) 659-673.

[7] U. Höhle, A. Šostak, *Axiomatics for fixed-based fuzzy topologies*, Chapter 3 in [8]

[8] B. Hutton, *Normality in fuzzy topologicl spaces*, J. Math. Anal. Appl., **50** (1975), 74-79.

[9] **Mathematics of Fuzzy Sets: Logic, Topology and Measure Theory** , U. Höhle, S.E. Rodabaugh eds. - Handbook Series, vol.3, Kluwer Acad. Publ., - 1999.

[10] S. Jenei, *Structure of Girad Monoids on [0,1]* Chapter 10 In: **Topological and Algebraic Structures in Fuzzy Sets** S.E. Rodabaugh, E.P. Klement eds., Kluwer Acad. Publ. 2003.

[11] J. Korteleinen, On relationship between modified sets, topological spaces and rough sets, Fuzzy Sets and Syst. **61** (1994), 91-95.

[12] J. Järvinen, J. Kortelainen, **A unified study between modal-like operators, Topologies and Fuzzy Sets**, Turku Centre for Computer Sciences, Turku, Finland, Technical Reports **642**, 2004.

[13] T. Kubiak, *On Fuzzy Topologies*, PhD Thesis, Adama Mickiewicz University Poznan, Poland, 1985.

[14] Z. Pawlak, *Rough sets*, International J. of Computer and Information Sciences, **11** (1982), 341-356.

[15] S.E. Rodabaugh, A categorical accomodation of various notions of fuzzy topology, *Fuzzy Sets and Syst.*, **9** (1983), 241-265.

[16] S.E. Rodabaugh, *Powerset Operator Foundations For Poslat Fuzzy Set Theories and Topologies*, Chapter 2 in [8].

[17] A. Šostak, *On a fuzzy topological structure*, Suppl. Rend. Circ. Matem. Palermo, Ser II **11** (1985), 125-186.

[18] A. Šostak, *Basic structures of fuzzy topology*, J. Math. Sci. **78** (1996), 662-701.

[19] A. Šostak, *Approximative fuzzy systems and approximative fuzzy spaces*, In: 9^{th} International Conference on Fuzzy Set Theory and Applications: FSTA2008. Abstracts, p. 118.

[20] Y.Y. Yao, *A comparative study of fuzzy sets and rough sets*, Inf. Sci., **109** (1998), 227 - 242.

[21] Ying Mingsheng, *A new approach to fuzzy topology, Part* I, Fuzzy Sets and Syst., **39** (1991), 303-321.

ISBN: 978-989-95079-6-8

A Fuzzy Weight Representation for Inner Dependence Method AHP

Shin-ichi Ohnishi[12] Takahiro Yamanoi[1] Hideyuki Imai[2]

1 Faculty of Engineering, Hokkai-Gakuen University
Sapporo, 0640926 JAPAN
2 Graduate School of Information Science and Technology, Hokkaido University
Sapporo, 0600814 JAPAN
Email: {ohnishi,yamanoi}@eli.hokkai-s-u.ac.jp, imai@ist.hokudai.ac.jp

Abstract— *AHP (Analytic Hierarchy Process) has been widely used in decision making. Inner dependence method AHP is one technique even in case of criteria have dependency. However using original AHP or Inner dependence method, the results often lose reliability because the comparison matrix does not always have sufficient consistency. In these cases, fuzzy representation for weighting criteria and alternatives using results from a sensitivity analysis is useful. In this paper, we present alternative overall weights by employing some assumptions. Since an idea of less ambiguity is employed, the results show how inner dependence AHP has fuzziness when the comparison matrix is not sufficiently consistent..*

Keywords—Decision making, AHP (Analytic Hierarchy Process), Fuzzy sets, Inner Dependence, Sensitivity analysis.

1 Introduction

AHP (Analytic Hierarchy Process) was proposed by Saaty T.L. in 1977 [1], [2]. The method has been popular and widely used in the domain of decision making, since it can include vagaries such as humans feelings. In addition, it can be developed to ANP (Analytic Network Process) models.

Usually normal AHP must assume independency among criteria, although it is difficult to choose enough independent criteria. Inner dependence method AHP[10] is one technique of solving this kind of problem even in case of criteria have dependency.

However, the comparison matrix often does not have enough consistency when AHP or Inner dependence method is used since, for instance, a problem may contain too many criteria for decision making. In these cases, we consider that answers from decision-makers (i.e. components of the comparison matrix) have ambiguity or fuzziness. For resolving this type of problem, fuzzy reciprocal components have been proposed as components of the data matrix in some research [12]. In this paper, we consider that weights should also have ambiguity or fuzziness. Therefore, it is necessary to represent these weights by use of fuzzy sets.

Sensitivity analysis is applied to Inner dependence AHP to analyze the amount the components of a pairwise comparison matrix influences the weights and consistency of a matrix. This makes it possible to show the magnitude of the fuzziness in the weights.

In previous researches, we proposed a new representation for weights of criteria and alternatives in normal AHP. [7][8][11]. In this paper, a representation of weights of inner dependence method is proposed. It is represented as L-R fuzzy numbers by using the results from the sensitivity analysis. This paper encompasses methodology to represent weights by fuzzy sets. In addition, a representation of fuzziness as a result of inner dependence is presented when a comparison matrix does not have enough consistency.

2 Inner depedence AHP

2.1 Process of Normal AHP

(Process 1) Representation of structure by a hierarchy. The problem under consideration can be represented in a hierarchical structure. The highest level of the hierarchy consists of a unique element that is the overall objective. At the lower levels, there are multiple activities (i.e. elements within a single level) with relationships among elements of the adjacent higher level to be considered. The activities are evaluated using subjective judgments of a decision maker. Elements that lie at the upper level are called parent elements while those that lie at lower level are called child elements. Alternative elements are put at the lowest level of the hierarchy

(Process 2) Paired comparison between elements at each level. A pairwise comparison matrix A is created from a decision maker's answers. Let n be the number of elements at a certain level. The upper

ISBN: 978-989-95079-6-8

triangular components of the comparison matrix a_{ij} ($i< j$ = 1,…,n) are 9, 8, .. , 2, 1, 1/2, …, or 1/9. These denote intensities of importance from activity i to j. The lower triangular components a_{ji} are described with reciprocal numbers as follows

$$a_{ji} = 1/a_{ij}\,, \tag{1}$$

in addition, for diagonal elements, let a_{ii} = 1. The lower triangular components and diagonal elements are occasionally omitted from the written equation as they are evident if upper triangular components are shown. The decision maker should make $n(n\text{-}1)/2$ paired comparisons at a level with n elements.

(Process 3) Calculations of weight at each level. The weights of the elements, which represent grade of importance among each element, are calculated from the pairwise comparison matrix. The eigenvector that corresponds to a positive eigenvalue of the matrix is used in calculations throughout in this paper.

(Process 4) Priority of an alternative by a composition of weights. The composite weight can be calculated from the weights of one level lower. With repetition, the weights of the alternative, which are the priorities of the alternatives with respect to the overall objective, are finally found.

2.2 Consistency

Since components of the comparison matrix are obtained by comparisons between two elements, coherent consistency is not guaranteed. In AHP, the consistency of the comparison matrix A is measured by the following consistency index (C.I.)

$$\text{C.I.} = \frac{\lambda_A - n}{n-1}, \tag{2}$$

where n is the order of matrix A, and λ_A is its maximum eigenvalue.

It should be noted that $\text{C.I.} \geq 0$ holds. And if the value of C.I. becomes smaller, then the degree of consistency becomes higher, and vice versa. The comparison matrix is consistent if the following inequality holds.

$$\text{C.I.} \leq 0.1$$

Also consistency ratio (C.R.) is defined as

$$\text{C.R.} = \frac{\text{C.I.}}{M},$$

Where M is random consistency value. However we only employ C.I., since we mainly use 4 or 5-dimensional data whose random consistency value is not far from 1.

2.3 Inner Dependence Method

Usually normal AHP must assume independency among criteria, although it is difficult to choose enough independent criteria. Inner dependence method AHP[10] is one technique of solving this kind of problem even in case of criteria have dependency.

In the method, using a dependency matrix $F=\{f_{ij}\}$, we can calculate real weights $\boldsymbol{w_n}$ as follows,

$$\boldsymbol{w_n} = F\boldsymbol{w} \tag{3}$$

where $\boldsymbol{w}$ is weights from independent criterion, i.e. normal weights of normal AHP and F is calculated as eigen value of influenced matrix.

3 Sensitivity Analysis

When AHP is used, the comparison matrix is often inconsistent or large differences among the overall weights of the alternatives do not appear. Thus, it is very important to investigate how the components of a pairwise comparison matrix influence the consistency or weights. Sensitivity analysis is used to analyze how results are influenced when certain variables change. Therefore, it is necessary to establish a sensitivity analysis of AHP.

In our research, a previously proposed method [7] is used to evaluate the fluctuation of the consistency index and weights when a comparison matrix is perturbed. This method is useful as it does not change the structure of the data.

Evaluating the consistency index and the weights of a perturbed comparison matrix are performed as follows.

(1) Perturbations $\varepsilon a_{ij}d_{ij}$ are imparted to component a_{ij} of a comparison matrix, and the fluctuation of the consistency index and the weight are expressed by the power series of ε.

(2) Fluctuations of the consistency index and the weights are represented by the linear combination of d_{ij}.

(3) By the coefficient of d_{ij}, it can be shown that how the component of the comparison matrix gives influence on the consistency index and the weight.

Since the pairwise comparison matrix A is a positive square matrix, the following Perron-

ISBN: 978-989-95079-6-8

Frobenius theorem [4] holds.

Theorem 1 (Perron – Frobenius) *For a positive square matrix A, the following holds true.*

1. *Matrix A has a positive eigenvalue. If λ_A is the largest eigenvalue then λ_A is a simple root. The positive eigenvector $\boldsymbol{w}$, corresponding to λ_A, exists. λ_A is called the Frobenius root of A.*
2. *Any positive eigenvectors of A are the constant multiples of $\boldsymbol{w}$.*
3. *The absolute value of the eigenvalues of A, except for λ_A, is smaller than λ_A.*
4. *The Frobenius root of the transposed matrix A' is equivalent to the Frobenius root of A.*

This theorem ensures the existence of a weight vector in a pairwise comparison matrix.

From Theorem 1, the following theorem regarding a perturbed comparison matrix holds true [7].

Theorem2 *Let $A = (a_{ij})$, $i,j = 1,\ldots,n$ be a comparison matrix and let $A(\varepsilon) = A+\varepsilon D_A$, $D_A=(a_{ij}d_{ij})$ be a matrix that has been perturbed. Moreover, let λ_A be the Frobenius root of A with $\boldsymbol{w}_1$ being the corresponding eigenvector. Let $\boldsymbol{w}_2$ be the eigenvector corresponding to the Frobenius root of A', then, the Frobenius root $\lambda(\varepsilon)$ of $A(\varepsilon)$ and the corresponding eigenvector $\boldsymbol{w}_1(\varepsilon)$ can be expressed as follows*

$$\lambda(\varepsilon) = \lambda_A + \varepsilon\lambda^{(1)} + o(\varepsilon), \tag{4}$$

$$\boldsymbol{w}_1(\varepsilon) = \boldsymbol{w}_1 + \varepsilon\boldsymbol{w}^{(1)} + \boldsymbol{o}(\varepsilon), \tag{5}$$

where

$$\lambda^{(1)} = \frac{\boldsymbol{w}_2' D_A \boldsymbol{w}_1}{\boldsymbol{w}_2' \boldsymbol{w}_1}, \tag{6}$$

$\boldsymbol{w}^{(1)}$ is an n-dimension vector that satisfies

$$(A - \lambda_A I)\boldsymbol{w}^{(1)} = -(D_A - \lambda^{(1)} I)\boldsymbol{w}_1, \tag{7}$$

where $\boldsymbol{o}(\varepsilon)$ denotes an n-dimension vector in which all components are $o(\varepsilon)$.

Proof of this theorem can be found in Ohnishi's paper [7].

3.1 Sensitivity analysis of consistency index

Regarding a fluctuation of the consistency index, the following corollary can be obtained from Theorem 2.

Corollary 1 *Using an appropriate p_{ij}, we can represent the consistency index* C.I.(ε) *of the perturbed comparison matrix as follows*

$$\text{C.I.}(\varepsilon) = \text{C.I.} + \varepsilon\sum_i^n\sum_j^n p_{ij}d_{ij} + o(\varepsilon). \tag{8}$$

(Proof)

From the definition of the consistency index (3) and (4),

$$\text{C.I.}(\varepsilon) = \text{C.I.} + \varepsilon\frac{\lambda^{(1)}}{n-1} + o(\varepsilon).$$

Let $\boldsymbol{w}_1=(w_{1i})$ and $\boldsymbol{w}_2=(w_{2i})$ from (6). $\lambda^{(1)}$ is can now be represented as

$$\lambda^{(1)} = \frac{1}{\boldsymbol{w}_2' \boldsymbol{w}_1}\sum_i^n\sum_j^n w_{2i}\, a_{ij}\, w_{1j}\, d_{ij},$$

therefore, the second part of the right side is expressed by a linear combination of d_{ij}. (Q.E.D)

p_{ij} in equation (8) in Corollary 1 shows the influence of comparison matrix components on the consistency.

On the other hand, since the comparison matrix $A(\varepsilon) = (a_{ij}(\varepsilon))$ is reciprocal, then $a_{ji}(\varepsilon) = 1/a_{ij}(\varepsilon)$ and becomes

$$a_{ji} + \varepsilon a_{ji}d_{ji} = \frac{1}{a_{ij}} - \varepsilon\frac{d_{ij}}{a_{ij}} + o(\varepsilon). \tag{9}$$

Here, since $a_{ji} = 1/a_{ij}$,

$$d_{ji} = -d_{ij} \tag{10}$$

is obtained. The impact on the consistency can be easily shown by use of this property.

3.2 Sensitivity analysis of weights

With regards to the fluctuation in weighs, the following corollary can also be obtained from Theorem 2.

Corollary 2 *Using an appropriate $q_{ij}^{(k)}$, we can represent the fluctuation $\boldsymbol{w}^{(1)}=(w_k^{(1)})$ of the weight (i.e. the eigenvector corresponding to the Frobenius root) as follows*

$$w_k^{(1)} = \sum_i^n\sum_j^n h_{ij}^{(k)} d_{ij}. \tag{11}$$

(Proof)

The *k*-th row component of the right side of (7) in

ISBN: 978-989-95079-6-8

Theorem 2 is represented as

$$\sum_i^n \sum_j^n \{\frac{w_{1k}\, w_{2i}\, a_{ij}\, w_{1j}}{\boldsymbol{w}_2'\ \boldsymbol{w}_1} - \delta(i,k) a_{ij}\, w_{1j}\}\, d_{ij},$$

and is expressed by a linear combination of d_{ij}. Here,$\delta(i,k)$ is Kronecker's symbol

$$\delta(i,k) = \begin{cases} 1 & (i = k), \\ 0 & (i \neq k). \end{cases}$$

In contrast, since λ_A is a simple root, Rank(A-$\lambda_A I$) = n-1. Accordingly, the weight vector is normalized as

$$\sum_k^n (w_{1k} + \varepsilon w_k^{(1)}) = \sum_k^n w_{1k} = 1,$$

then the condition is as follows.

$$\sum_k^n w_k^{(1)} = 0. \qquad (12)$$

By using an elementary transformation to formula (7) in the condition above, we also can represent $w_k^{(1)}$ by linear combinations of d_{ij}. (Q.E.D)

As seen in equation (5) in Theorem 2, the component that has a great influence on weight $\boldsymbol{w}_1(\varepsilon)$ is the component which has the greatest influence on $\boldsymbol{w}^{(1)}$. $q_{ij}^{(k)}$ in equation (11) from Corollary 2 shows how the influence by the components of a comparison matrix on the weights can be calculated.

The influence can also be shown easily by use of equation (10).

3.3 Sensitivity for inner dependence method

We can also calculate ith regards to the fluctuation in weighs, the followi

Corollary 3 *Using an appropriate $h_{ij}^{(k)}$,we can represent the fluctuation $\boldsymbol{w}_n^{(1)}$=($w_n^{(1)}$) of the weights of inner dependence AHP as follows*

$$w_n^{(1)} = \sum_i^n \sum_j^n h_{ij}^{(k)} d_{ij}.$$

(Proof)

From the equation (3), weights of inner dependence method $\boldsymbol{w}_n$ is calculated from linear transformation of normal weights $\boldsymbol{w}$ of normal AHP, and from Corollary 2 $\boldsymbol{w}$ is represented as sum of linear combination of d_{ij} .Therefore weights $\boldsymbol{w}_n$ is also represented as sum of linear combination of $d_{ij.}$. (Q.E.D)

4 A Weights Representation

The comparison matrix often has poor consistency (i.e. 0.1<C.I.<0.2) because it encompasses several activities. In these cases, the components of a comparison matrix are considered to have fuzziness since they result from the fuzzy judgment of humans. Therefore, weights should be treated as fuzzy numbers.

To represent fuzziness of weight w_{1k}, an L-R fuzzy number is used.

4.1 L-R fuzzy number

L-R fuzzy number

$$M = (m, \alpha, \beta)_{LR}$$

is defined as fuzzy sets whose membership function is as follows.

$$\mu_M(x) = \begin{cases} R\left(\dfrac{x-m}{\beta}\right) & (x > m), \\ L\left(\dfrac{m-x}{\alpha}\right) & (x \leq m). \end{cases}$$

where $L(x)$ and $R(x)$ are shape function which satisfies

(1) $L(x) = L(-x)$,

(2) $L(0) = 1$,

(3) $L(x)$ is a non increasing function

4.2 Fuzzy weights of criteria

From the fluctuation of the consistency index, the multiple coefficient $g_{ij}h_{ij}^{(k)}$ in Corollary 1 and 3 is considered as the influence on a_{ij}.

Since g_{ij} is always positive, if the coefficient $h_{ij}^{(k)}$ is positive, the real weight of criterion k is considered to be larger than w_{1k}. Conversely, if $h_{ij}^{(k)}$ is negative, the real weight of activity k is considered to be smaller. Therefore, the sign of $h_{ij}^{(k)}$ represents the direction of the fuzzy number spread. The absolute value $g_{ij}|h_{ij}^{(k)}|$ represents the size of the influence.

On the other hand, if C.I. becomes bigger, then the judgment becomes more fuzzy.

Consequently, multiple C.I. $g_{ij}|h_{ij}^{(k)}|$ can be regarded as a spread of a fuzzy weight $\tilde{w}_k$ concerned with

ISBN: 978-989-95079-6-8

a_{ij}.

Definition 1 (fuzzy weight) *Let* w_{nk} *be a crisp weight of criterion k of inner dependence model, and* g_{ij} $|h_{ij}^{(k)}|$ *denote the coefficients found in* Corollary 1 and 3. *If* 0.1<C.I.<0.2*, then a fuzzy weight* $\tilde{w}_k$ *is defined by*

$$\tilde{w}_k = (w_{nk}, \alpha_k, \beta_k)_{LR} \tag{13}$$

where

$$\alpha_k = \text{C.I.} \sum_i^n \sum_j^n s(-, h_{kij}) g_{ij} \mid h_{kij} \mid, \tag{14}$$

$$\beta_k = \text{C.I.} \sum_i^n \sum_j^n s(+, h_{kij}) g_{ij} \mid h_{kij} \mid, \tag{15}$$

$$s(+,h) = \begin{cases} 1, (h \geq 0) \\ 0. (h < 0) \end{cases}, \; s(-,h) = \begin{cases} 1, (h < 0) \\ 0. (h \geq 0) \end{cases}$$

4.2 Fuzzy weights of alternatives

Using the fuzzy weights of criteria defined above and local crisp weights of alternatives with respect to certain criterion, we can calculate overall weights from the viewpoint of the overall objective by extension. However, the results from the operation of fuzzy numbers are frequently too ambiguous to interpret.

Fuzzy weights of activities are normalized thus their sum is 1, therefore we can avoid much ambiguity since this condition has been considered [9].

In general, operating with constraints is difficult but can be accomplished if every fuzzy membership function is linear.

Especially for every normal triangular function with a core u_i, the constraint $\sum_i^n u_i = 1$ holds, and the order of singleton coefficients is assumed. Thus, the upper and lower limit of α -cut sets of linear sum can be easily calculated.

Let $f_t(x_k)$ be a crisp local weight of alternative *t* with respect to activity *k,* and in this paper, assume $0 \leq f_t(x_1) \leq f_t(x_2) \leq \cdots \leq f_t(x_n)$. Then, the overall weight of an alternative *t* is also the L-R fuzzy number and is represented as follows.

$$\tilde{v}_t = (v_t, l_t, r_t)_{LR}$$

where

$$v_t = \sum_k^n w_{1k} f_t(x_k),$$

$$l_t = v_t - \inf \text{supp}(\tilde{v}_t), \quad r_t = \sup \text{supp}(\tilde{v}_t) - v_t$$

In the above equations, inf supp, sup supp are lower and upper limits of support sets and are calculated as follows.

$$\inf \text{supp}(\tilde{v}_t) = \max_j \left[\sum_{i=1}^{j-1} (w_{1i} + \beta_i) f_t(x_i) + \sum_{i=j+1}^{n} (w_{1i} - \alpha_i) f_t(x_i) + \left\{ 1 - \sum_{i=1}^{j-1} (w_{1i} + \beta_i) - \sum_{i=j+1}^{n} (w_{1i} - \alpha_i) \right\} f_t(x_j) \right]$$

$$\sup \text{supp}(\tilde{v}_t) = \min_j \left[\sum_{i=1}^{j-1} (w_{1i} - \alpha_i) f_t(x_i) + \sum_{i=j+1}^{n} (w_{1i} + \beta_i) f_t(x_i) + \left\{ 1 - \sum_{i=1}^{j-1} (w_{1i} - \alpha_i) - \sum_{i=j+1}^{n} (w_{1i} + \beta_i) \right\} f_t(x_j) \right]$$

5. Conclusions

We proposed a representation for the inner dependence overall weights of alternatives by use of fuzzy sets and the result of a sensitivity analysis for cases in which consistency of the comparison matrix is poor. Our approach shows how to represent weights, as well as how the result of AHP has fuzziness, when inconsistency exists. This was due to reduced ambiguity in the representation presented in this work compared to previous normal fuzzy operations.

Acknowledgements

This research is partly supported by a grant from the ministry of education, sports, science, and technology to the national project of 'Advanced improvements of vision, image, speech and language information processing and the application to the technologies for the intelligent instrument and control' in the High-tech Research Center of Hokkai-Gakuen University.

ISBN: 978-989-95079-6-8

References

[1] T. L. Saaty. A scaling method for priorities in hierarchical structures. *J. Math. Psy*., 15(3):234--281, 1977.

[2] T. L. Saaty. The Analytic Hierarchy Process. *McGraw-Hill, New York*, 1980.

[3] T. L. Saaty. Scaling the membership function. European *J. of O.R.*, 25:320--329, 1986.

[4] M. Saito. An Introduction to Linear Algebra. *Tokyo University Press*, 1966.

[5] Y. Tanaka. Recent advance in sensitivity analysis in multivariate statistical methods. *J. Japanese Soc. Comp. Stat*., 7(1):1--25, 1994.

[6] K. Tone. The Game Feeling Decision Making. *Nikka-giren Press, Tokyo*, 1986.

[7] S. Ohnishi, H. Imai, and M. Kawaguchi. Evaluation of a Stability on Weights of Fuzzy Analytic Hierarchy Process using a sensitivity analysis. *J. Japan Soc. for Fuzzy Theory and Sys.,* 9(1):140--147, 1997..

[8] S. Ohnishi, H. Imai, T. Yamanoi. Weights Representation of Analytic Hierarchy Process by use of Sensitivity Analysis, *IPMU 2000 Proceedings*, 2000.

[9] D. Dubois and H. Prade. Possibility Theory An Approach to Computerized Processing of Uncertainty,. *Plenum Press, New York*, 1988.

[10] .T. L. Saaty. Inner and Outer Dependence in AHP. *University of Pittsburgh*, 1991.

[11] S. Ohnishi, T. Yamanoi, H. Imai, A Fuzzy Representation for Weights of Alternatives in AHP, *New Dimensions in Fuzzy Logic and Related Technologies,* Vol. II , pp.311-316 2007

[12] S. Ohnishi, D. Dubois, H.Prade, T.Yamanoi, A Fuzzy Constraint-based Approach to the Analytic Hierarchy Process, *Uncertainty and Intelligent Information Systems,* pp.2171-228 , 2008.

ISBN: 978-989-95079-6-8

On Conceptual Indexing for Data Summarization

Henrik Bulskov[1] Troels Andreasen[1]

1. Department of Computer Science, Roskilde University,
P.O. Box 260, DK-4000 Roskilde, Denmark Email: bulskov@ruc.dk, troels@ruc.dk

Abstract— A summary is a comprehensive description that grasps the essence of a subject. A text, a collection of text documents, a query answer can be summarized by simple means such as an automatically generated list of the most frequent words or "advanced" by a meaningful textual description of the subject. In between these two extremes are summaries by means of selected concepts exploiting background knowledge providing selected key concepts. We address in this paper an approach where conceptual summaries are provided through a conceptualization as given by an ontology. The idea is to restrict a background ontology to the set of concepts that appears in the text to be summarized and thereby provide a structure, a so-called instantiated ontology, that is specific to the domain of the text and can be used to condense to a summary not only quantitatively but also conceptually covers the subject of the text.

Keywords— conceptual clustering, conceptual descriptions, conceptual summaries, ontologies.

1 Introduction

The purpose of a summary is to provide a simplification to highlight the major points from the subject, e.g. a text or a set of texts such as a query answer. The aim is to provide a summary that grasps the essense of the subject.

Most common are summaries as those provided manually by readers or authors as a result of intellectual interpretation. Summaries can however also be provided automatically. One approach, in the Question Answering style, such as this is investigated in for instance the DUC and TREC conferences (see for instance [6],[5],[7]), is to provide a full natural language generation based summary construction while a less ambiguous, in the same tradition, is rather to perform a sentence selection from the text to be summarized.

In the other end the most simple approach is to select a reasonable short list of words among the most frequent and/or the most characteristic words from the set of words found in the text to be summarized. So rather than a coherent text the summary is a simple set of items.

Summaries in the approach presented here are also sets of items, but involves improvements over the simple set of words approach in two respects. First, we go beyond the level of keywords and aim to provide conceptual descriptions from concepts identified and extracted from the text. Second, we involve background knowledge in the form of an ontology. Strictly these two aspects are closely related – to use the conceptualization in the ontology we need means to map from words and phrases in the text to concepts in the ontology.

In this paper we present two different directions to conceptual summaries based on a background ontology. In both cases an ontology plays a key role as reference for the conceptualization. The general idea is from a world knowledge ontology to form a so-called "instantiated ontology" by restricting to a set of instantiated concepts.

First, we consider conceptual clustering over the instantiated concepts based on a semantic similarity measure and second we present an approach based on probabilities.

Below we first introduce to the the ontology notion, then we discuss extraction of conceptual descriptions, and finally we describe the various approaches to conceptual summaries.

2 Representing background knowledge – Ontology

Background knowledge is knowledge that complements the primary target data (the text or text collection / database) that is subject of the summarization with information that is essential to the understanding of this. Background knowledge can take different forms varying from simple lists of words to formal representations. In this context, however, our goal is conceptual summaries provided as sets of words or concepts so background knowledge to support this can range from unstructured lists of words to ontologies.

A simple list of words can be applied as a filter, mapping from a text to the subset of the word list that appears in the text. Such a controlled list of keywords or a vocabulary of topics can by obvious means be improved to capture also morphology by stemming or inflection patterns. Taxonomies, partonomies, semantic networks and ontologies are structures that potentially contribute also to knowledge-based summarization. Our main focus here is on ontologies ordered around taxonomic relationship. Rather than the common description logic based approach we choose here a simpler concept algebraic approach to ontologies.

One important rationale for this is that our goal here is not ontological reasoning in general but rather extraction of sets of mapped concepts and manipulation of such sets (e.g. contraction).

ISBN: 978-989-95079-6-8

2.1 An algebraic approach to ontologies

Given a basis taxonomy that situates a set of atomic term concepts $\mathcal{A}$ in a multiple inheritance hierarchy. Based on this we define a generative ontology by generalization of the hierarchy to a lattice an by introducing a (lattice-algebraic) concept language (description language) that defines an extended set of well-formed concepts, including both atomic and compound term concepts.

The concept language used here, ONTOLOG[9], has as basic elements concepts and binary relations between concepts. The algebra introduces two closed operations *sum* and *product* on concept expressions φ and ψ, where $(\varphi + \psi)$ denotes the concept being either φ or ψ and $(\varphi \times \psi)$ denotes the concept being φ and ψ (also called *join* and *meet* respectively).

Relationships r are introduced algebraically by means of a binary operator (:), known as the Peirce product $(r : \varphi)$, which combines a relation r with an expression φ. The Peirce product is used as a factor in conceptual products, as in $x \times (r\colon y)$, which can be rewritten to form the feature structure $x[r\colon y]$, where $[r\colon y]$ is an *attribution* of the concept x. Thus we can form compound concepts by attribution.

Given atomic concepts $\mathcal{A}$ and semantic relations $\mathcal{R}$, the set of well-formed terms $\mathcal{L}$ is:

$$\mathcal{L} = \{\mathcal{A}\} \cup \{x[r_1 : y_1, ..., r_n : y_n] | x \in \mathcal{A}, r_i \in \mathcal{R}, y_i \in \mathcal{L}\} \quad (1)$$

Compound concepts can thus have multiple as well as nested attributions. For instance with $\mathcal{R}$= {WRT, CHR, CBY, TMP, LOC, ...}[1] and $\mathcal{A}$ ={*entity, physical_entity, abstract_entity, location, town, cathedral, old*} we get:

$$\begin{aligned}&\mathcal{L} = \\ &\{entity, physical_entity, abstract_entity,\\ &location, town, cathedral, old,\\ &\ldots, \textit{cathedral}[\text{LOC}\colon \textit{town}, \text{CHR}\colon \textit{old}],\\ &\textit{cathedral}[\text{LOC}\colon \textit{town}[\text{CHR}\colon \textit{old}]], \ldots\}\end{aligned}$$

2.2 Modelling Ontologies

Obviously modelling ontologies from scratch is the best way to ensure that the result will be correct and consistent. However, for many applications the effort it takes is simply not at disposal and manual modeling have to be restricted to narrow and specific subdomains while the major part have to be derived from relevant sources. Sources that may contribute to modeling of ontologies may have various forms. A taxonomy is an obvious choice and it may be supplemented with, for instance, word and term lists as well as dictionaries for definition of vocabularies and for handling of morphology. Among the obviously useful resources are the semantic network WordNet [11] and the Unified Medical Language System (UMLS) [4] and unifies several other resources in the biomedical science area.

To go from a resource to an ontology is not necessarily straightforward, but if the goal is a generative ontology and the given resource is a taxonomy, one option is to proceed as follows. Given a taxonomy $\mathcal{T}$ over the set of atomic concepts $\mathcal{A}$ and a language $\mathcal{L}$, over $\mathcal{A}$ for a given set of relations $\mathcal{R}$, being derived as indicated in (1) above. Let $\hat{\mathcal{T}}$ be the transitive closure of $\mathcal{T}$. $\hat{\mathcal{T}}$ can be generalized to an inclusion relation "$\leq$" over all well-formed terms of the language $\mathcal{L}$ by the following

$$\begin{aligned}\text{“}\leq\text{“} = \quad &\hat{\mathcal{T}}\\ &\cup\{\langle x[\ldots, r\colon z], y[\ldots]\rangle | \langle x[\ldots], y[\ldots]\rangle \in \hat{\mathcal{T}}\}\\ &\cup\{\langle x[\ldots, r\colon z], y[\ldots, r\colon z]\rangle | \langle x[\ldots], y[\ldots]\rangle \in \hat{\mathcal{T}}\}\\ &\cup\{\langle z[\ldots, r\colon x], z[\ldots, r\colon y]\rangle | \langle x, y\rangle \in \hat{\mathcal{T}}\}\end{aligned}$$

where repeated ... denote zero or more attributes of the form $r_i\colon w_i$.

The general ontology $\mathcal{O} = (\mathcal{L}, \leq, \mathcal{R})$ thus encompasses a set of well-formed expressions $\mathcal{L}$ derived in the concept language from a set of atomic concepts $\mathcal{A}$, an inclusion relation generalized from the taxonomy relation in $\mathcal{T}$, and a supplementary set of semantic relations $\mathcal{R}$. For $r \epsilon \mathcal{R}$, we obviously have $x[r\colon y] \leq x$, and that $x[r\colon y]$ is in relation r to y. Observe that $\mathcal{O}$ is generative and that $\mathcal{L}$ therefore is potentially infinite.

An example is given in figure 1 showing a segment of a generative ontology build with WordNet as resource.

3 Referencing the background knowledge – providing descriptions

As already indicated the approach involves surveying text through the ontology provided and delivering summaries on top of the conceptualization of the ontology. For this purpose we need to provide a description of the text to be summarized in terms of the concepts in the ontology. So words and/or phrases must be extracted from the text and mapped into the ontology. This is a knowledge extraction problem, and obviously such knowledge extraction can span from full deep natural language processing (NLP) to simplified shallow processing methods.

Here we will consider the latter due to the counterbalance between the need for a full interpretation and the computational complexity of getting it. A very simple solution would match words in text with labels of concepts in the ontology, hence make a many-to-many relation between words in text and labels in the ontology that just accepts the ambiguity of natural language. Improvements can easily be obtained through pattern based information extraction / text mining and through methods in natural language processing.

First, a heuristic part of speech tagging can be performed on the text, and provided that word classes are assigned to the concepts given in the ontology this enables a word class based disambiguation.

Second, a stemming or, provided lexical information is available, a transformation to a standardized inflectional form can significantly improve the matching.

Third, given part of speech tagged input, simple syntactic natural language grammars can be used to chunk words together forming utterances or phrases [3], that

[1] for *with respect to, characterized by, caused by, temporal, location*, respectively.

ISBN: 978-989-95079-6-8

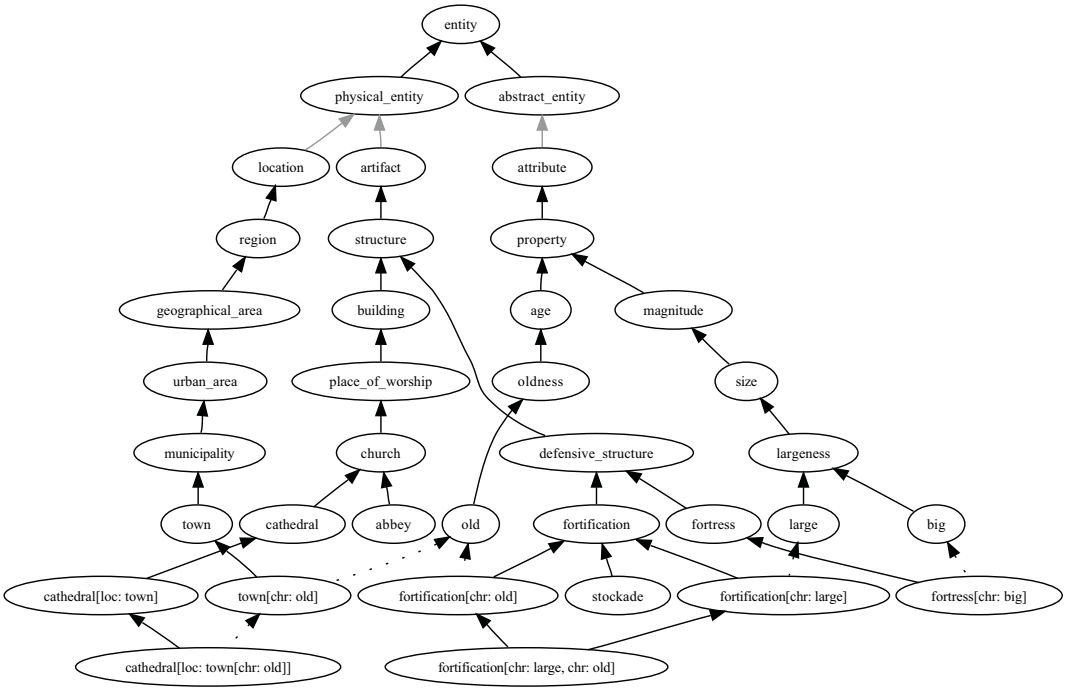


Figure 1: A segment of an ontology based on Wordnet. Does also correspond to an instantiated ontology for the set of instantiated concepts {*cathedral*[LOC : *town*[CHR : *old*]], abbey, *fortification*[CHR : *large*, CHR : *old*], stockade, *fortress*[CHR : *big*]}

.

can be used as the basis for matching against compound concepts in the ontology. Obviously the matching of chunks from the text and concepts in the ontology is in principle the same complex NLP problem over again, but the chunks identified will often correspond to meaningful concepts and therefore lead to a more refined and better result of the matching and, in addition, allow for a simple pattern-based approaches. We refer to [18] and [1] for more refined approaches. Here we will only cover a simple pattern-based approach.

Finally, some kind of word sense disambiguation [20] can be introduced in order to narrow down the possible readings of words, hence ideally mapping words of phrases to exactly one concept in the ontology.

A very simple approach along these lines is the following. Given a part-of-speech tagged and NP-chunked input a grammar for interpretation of the chunks is the following:

$$\begin{aligned} \text{Head} &::= \text{N} \\ \text{NP} &::= \text{A* N* Head} \mid \text{NP P NP} \end{aligned} \tag{2}$$

where A, N and P as placeholders for adjective, noun and preposition respectively. A very course-grained mapping strategy on top of this interpretation can be formed using the following transition rules, where premodifying adjectives relates to the head through *characterized by* (CHR) while premodifying composite nouns and prepositions both relates through *with respect to* (WRT):

$$\begin{aligned} &\text{A}_1, \ldots, \text{A}_n\text{N}_1, \ldots, \text{N}_m\text{Head} \mapsto \\ &\quad \text{Head}[\text{CHR}: \text{A}_1, \ldots, \text{CHR}: \text{A}_n, \text{WRT}: \text{N}_1, \ldots, \text{WRT}: \text{N}_m] \\ &\text{NP (P NP)}_1, \ldots, \text{(P NP)}_n \mapsto \\ &\quad \text{NP}[\text{ WRT}: \text{NP}_1, \ldots, \text{WRT}: \text{NP}_n] \end{aligned} \tag{3}$$

To test this approach we consider the the Metathesaurus in the Unified Medical Language System (UMLS) [13] as resource and build a generative ontology from this. For part of speech tagging and phrase chunking we use the MetaMap application [2].

Consider the following utterance[2] as an example:

> [...] *the plasma patterns of estrogen and progesterone under gonadotropic stimulation simulating early pregnancy* [...]

The first part of the analysis leads to part of speech tagging and phrase recognition as shown in table 1.
By applying the grammar (2) this can be transformed into the following three noun phrases:

plasma/N patterns/N of/P estrogen/N

progesterone/N under/P gonadotropic/A stimulation/N

early/A pregnancy/N

and by using the transition rules (3) we can produce the following compound expressions:

[2]This utterance is from a small 50K abstracts fraction of MEDLINE [14] having both *Homones* and *Reproduction* as major topic keywords.

ISBN: 978-989-95079-6-8

Table 1: Part of speech tagging and phrase recognition

Phrase	Type	Word	POS
Noun Phrase	det	the	det
	mod	plasma	noun
	head	patterns	noun
Preposition	prep	of	prep
	head	estrogen	noun
	conj	and	conj
Noun Phrase	head	progesterone	noun
Preposition	prep	under	prep
	mod	gonadotropic	adj
	head	stimulation	noun
	verb	simulating	verb
Noun Phrase	mod	early	adj
	head	pregnancy	noun

patterns[WRT: plasma, WRT: estrogen]

progesterone[WRT:stimulation[CHR:gonadotropic]]

pregnancy[CHR:early]

3.1 Instantiated Ontology

The description $d_O(T)$ of a text T given the ontology O comprise a set of concepts in O and as indicated the purpose here is to summarize based on relations in the ontology. Now given the set of concepts (the description) $d_O(T)$ an obviously relevant subontology is a subontology that covers all elements of $d_O(T)$. Such a subontology can be consider an instantiation of the text T (or the set of concepts $d_O(T)$).

Given an ontology $\mathcal{O} = (\mathcal{L}, \leq, \mathcal{R})$ and a set of concepts C we define the instantiated ontology $\mathcal{O}_C = (\mathcal{L}_C, \leq_C, \mathcal{R})$ as a restriction of $\mathcal{O}$ to cover only the concepts in C, that is, C and every concept from $\mathcal{L}$ that subsumes concepts in C or attributes for concepts in C. $\mathcal{L}_C$ can be considered an "upper expansion" of C in $\mathcal{O}$. More specifically, with C^+ being C extended with every concept related by attribution from a concept in C:

$$\begin{aligned} \mathcal{L}_C &= C \cup \{x | y \in C^+, x \in \mathcal{L}, y \leq x\} \\ \text{“} \leq_C \text{“} &= \{(x, y) | x, y \in \mathcal{L}_C, x \leq y\} \end{aligned} \quad (4)$$

Thus $\mathcal{O}_C$ is not generative. "$\leq_C$" may be represented by a minimal set "$\leq'_C$"$\subseteq$"$\leq_C$" such that "$\leq_C$" is derivable from "$\leq'_C$" by means of transitivity of "$\leq$" and monotonicity of attribution:

$$\begin{aligned} transitivity : x &\leq y, y \leq z \Rightarrow x \leq z \\ monotonicity : x &\leq y \Rightarrow z[r : x] \leq z[r : y] \end{aligned}$$

Figure 1 shows an example of an instantiated ontology. The general ontology is based on (and includes) WordNet and the ontology shown is "instantiated" wrt. the following set of concepts:

$$\begin{aligned} C = \{&\textit{cathedral}[\text{LOC} : \textit{town}[\text{CHR} : \textit{old}]], \textit{abbey}, \\ &\textit{fortification}[\text{CHR} : \textit{large}, \text{CHR} : \textit{old}], \textit{stockade}, \\ &\textit{fortress}[\text{CHR} : \textit{big}]\} \end{aligned}$$

4 Data summarization through background knowledge

The general idea here is to exploit background knowledge through conceptual summaries, that are to provide a means to survey textual data, for instance a query result. A set of concepts from the background knowledge is first identified in the text and then contracted into a smaller set of, in principle, most representable concepts.

This can be seen as one direction in a more general conceptual querying approach where queries can be posed, and answers be presented, by means of conceptual abstractions. For a general discussion on other means, except from conceptual summaries, of conceptual querying, where also a dedicated language constructs for this purpose is presented we refer to [8]. Here we discuss summaries only.

In the approach to summarization described here we assume an ontology to guide the summarization and, for the text to be summarized, an initial extraction of concepts as described in the previous section. Thus, we can assume an initial set of concepts C and we a facing a challenge to provide a smaller set of representative concepts covering C, that is, an appropriate summary that grasps what´s most characteristic about C. For computation of the summary we restrict to the subontology $\mathcal{O}_C = (\mathcal{L}_C, \leq_C, \mathcal{R})$ corresponding to the instantiated ontology for C.

We introduce two directions for deriving summaries below: one based on clustering of the input concept set C and the other realized as a probability-based restriction of the input ontology $\mathcal{O}_C$. Towards the end we discuss possibilities for combining these approaches.

4.1 Similarity Clustering

Given a similarity measure summaries can be derived from a clustering of concepts applying this measure. Obviously, if the measure is derived from an ontology, and thereby do reflect this, then so will the clustering. We will below assume an ontology-based similarity measure sim. A simple example of such a measure can be derived from the path lenght in the ontology graph (Rada' *Shortest Path Length* [12]). More refined approaches are *Information Content* [16] and *Weighted Shared Nodes* [17].

4.1.1 A hierarchical similarity-based approach

With a given path-length dependent similarity measure derived from the ontology a lub-centered, agglomerative, hierarchical clustering can be performed as follows.

Initially each "cluster" corresponds to an individual element of the set to be summarized. At each particular stage the two clusters which are most similar are joined together. This is the principle of conventional hierarchical clustering. However rather than replacing the two joined clusters with their union as in the conventional approach they are replaced by their lub. Thus given a set of concepts $C = \{c_1, \ldots, c_n\}$ summarizers can be derived as follows.

ALGORITHM – Hierarchical clustering summary

INPUT: Set of concepts $C = \{c_1, \ldots, c_n\}$

ISBN: 978-989-95079-6-8

OUTPUT: Generalizing description $\delta(C)$ for C.

1) Let the instantiated ontology for C be $\mathcal{O}_C = (\mathcal{L}_C, \leq_C \mathcal{R})$
2) Let $T = \{\langle x, y\rangle | sim(x, y) = max_{z,w \in C}(sim(z, w))\}$
3) Let $U = min\left(\{u | u \in \mathcal{L}_C \wedge \exists x, y \in \mathcal{L}_C : x < u \wedge y < u\}\right)$
4) $L = \{x | \langle x, y\rangle \in T \vee \langle y, x\rangle \in T\}$
5) set $\delta(C) = C \cup U/L$

As was also the case with the connectivity clustering, to obtain an appropriate description of C we might have to apply δ several times and at some point m we have that $\delta^m(C) = Top$.

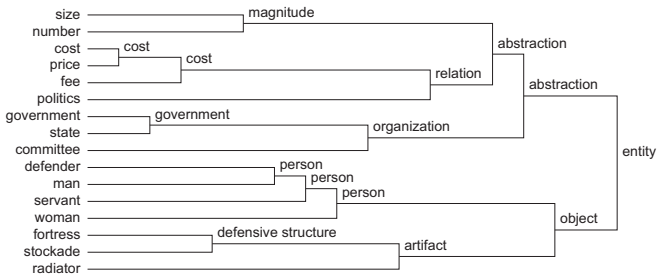


Figure 2: An illustration of the hierarchical clustering summary. The merging of two clusters are shown with their lub.

Figure 2 illustrates the application of δ a total of 15 times to the set of concepts from the previous example, where we might have (depending on the exact similarity values) for instance:

C = *{number, size, committee, government, state, defender, man, servant, woman, bribe, cost, price, fee, fortification, fortress, stockade}*

$\delta^1(C)$ = *{number, size, committee, government, state, defender, man, servant, woman, bribe,* ***cost****, fee, fortification, fortress, stockade}*

$\delta^2(C)$ = *{number, size, committee,* ***government****, defender, man, servant, woman, bribe, cost, price, fee, fortification, fortress, stockade}*

etc.

Thus summaries are generated iteratively and at each step the two closest concepts are clustered and the result is replaced by the corresponding lub.

4.1.2 Simple least upper bound-based approach

The principle of replacing clusters by their least upper bound can by applied on top of in principle any clustering approach. A straightforward similarity based approach is simply to apply a crisp clustering to the set of concepts $C = \{c_1, \ldots, c_n\}$ leading to $\{C_1, \ldots, C_k\}$ and then provide the set of lub's $\{\widehat{c}_1, \ldots, \widehat{c}_k\} = \{lub(C_1), \ldots, lub(C_k)\}$ for the division of C as summary. However to take into account also the importance of clusters in terms of their sizes the summary can be modified by the support of the generalizing concepts, $support(x, C)$, that for a given concept specifies the fraction of elements from the set C covered:

$$support(x, C) = \frac{|\{y | y \in C, y \leq x\}|}{|C|} \quad (5)$$

leading to a fuzzyfied (weighted) summary, based on the division (crisp clustering) of C into $\{C_1, \ldots, C_k\}$:

$$\Sigma_i support\left(lub(C_i), C\right) / lub(C_i) \quad (6)$$

To illustrate this lub-based approach consider table 2. Five groups are given that are derived as clusters of synsets in wordnet [3]

Table 2: A set of clusters and their least upper bounds from WordNet.

cluster	lub
{number, size}	magnitude
{committee, government, state}	organization
{defender, man, servant, woman}	person
{bribe, cost, fee, price}	cost
{fortification, fortress, stockade}	defensive structure

From these clusters the fuzzyfied summary {*.13/magnitude + .19/organization + .25/person + .25/cost + .19/defensive structure*} can be generated.

We may expect a pattern similar to hierarchical clustering in derivation of summaries in an approach based on similarity when the similarity measure reflects simple shortest path in the ontology.

This approach to summarization can be generalized using a fuzzyfied notion in place of the least upper bound as candidate representative. The generalization reduces the sensitivity against noise in the groups resulting from the initial clustering. This approach is described in [8].

4.2 A probability-based approach

We consider above a summary of textual input based on the concepts that appears in the text and how these are related in a background ontology. In the hierarchical approach candidate summarizers are chosen regardless of their coverage of the input, while the lub-based approached is introduced with a support that measures the degree to which all occurring input concepts are summarized. An obvious extension in this direction is to also consider frequencies of terms in the input text and thereby measure the probability of encountering an instance of a concept in the text.

Probabilities provides a means for selecting summarizers without taking other measures into account and thus allows for a straightforward approach as follows.

Given a set of concepts $C = \{c_1, \ldots, c_n\}$ and let $\mathcal{O}_C = (\mathcal{L}_C, \leq_C, \mathcal{R})$ be the instantiated ontology. Let further $child(c)$ denote the set of immediate children and $parent(c)$ the set of immediate parents for any concept $c \in \mathcal{L}_C$. The principle is to accumulate the frequencies to more general concepts but only so that a child c contributes with $\frac{1}{|parent(c)|}$ to each parent. A summary can be derived as follows.

ALGORITHM – Probability-based summary

INPUT: Set of concepts $C = \{c_1, \ldots, c_n\}$, their relative frequencies $C = \{freq(c_1), \ldots, freq(c_n)\}$ and a threshold α

[3] The first four are set of clusters and their least upper bounds in where $C_1, \ldots, C_4$ are from SEMCOR and C_5 is from the example ontology in figure 1.

ISBN: 978-989-95079-6-8

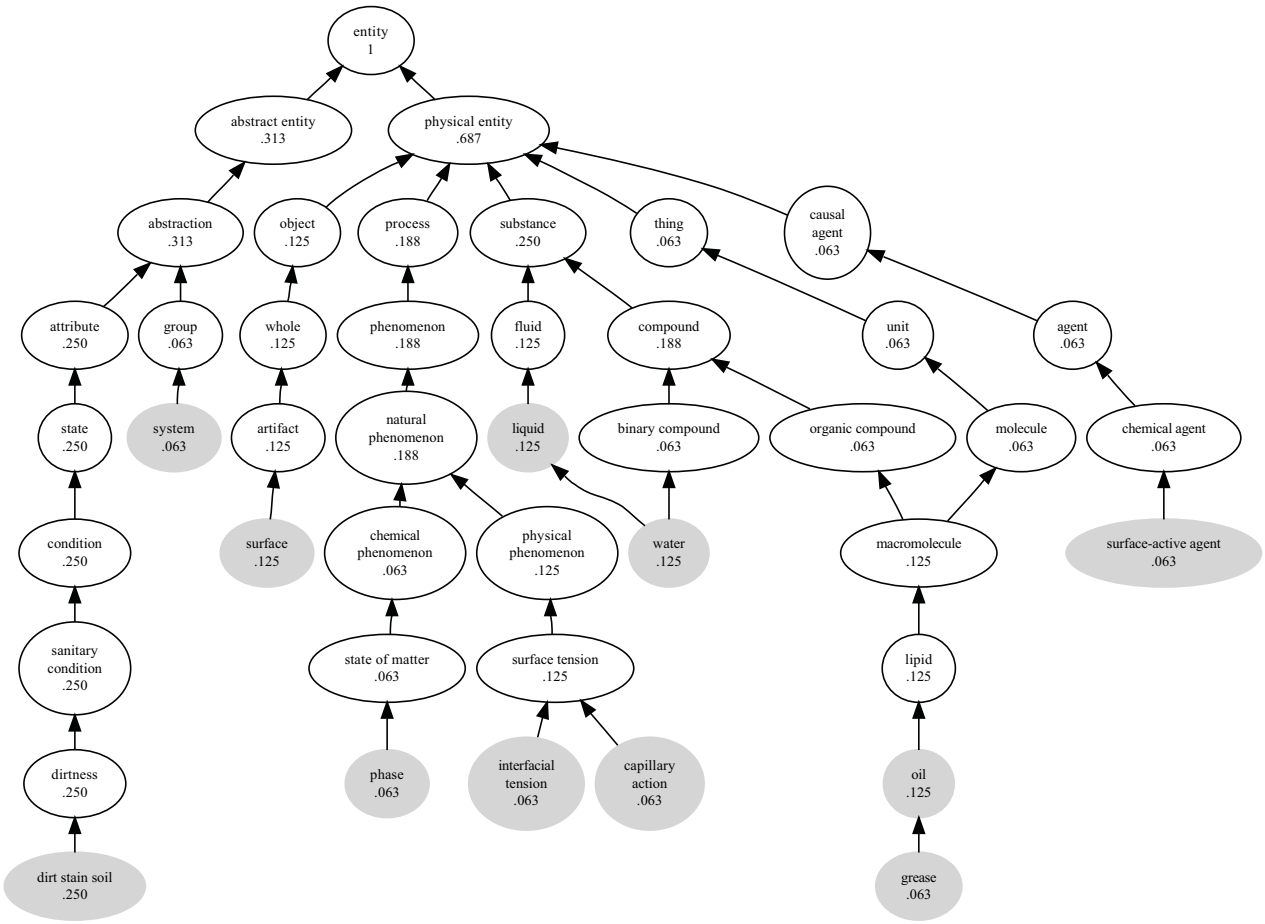


Figure 3: An instantiated ontology based on a paragraph from SEMCOR

OUTPUT: Generalizing description $D(C, \alpha)$ for C.

1) Let the instantiated ontology for C be $\mathcal{O}_C = (\mathcal{L}_C, \leq_C, \mathcal{R})$
2) Accumulate the frequencies in correspondence with the ontology such that $\forall c \in \mathcal{L}_C \backslash C : freq(c) = \sum_{c' \in child(c)} \frac{1}{|parent(c')|} freq(c')$
3) Let $N = |C|$ and $p(c) = freq(c)/N$ be the probability of encountering c.
4) Let $\mathcal{O'}_C = (\mathcal{L'}_C, \leq'_C, \mathcal{R}')$ be the restriction of $\mathcal{O}_C$ to the concepts that appear in $\{c \in \mathcal{L}_C | p(c) \geq \alpha\}$
5) Set the α-level summary of C to the most specific concepts appearing in $\mathcal{O'}_C$, that is $D(C, \alpha) = \{c \in \mathcal{L'}_C | \nexists c' \in \mathcal{L'}_C : c' < c\}$

As an example consider figure 3 that shows an instantiated ontology derived from WordNet for the following paragraph found in SEMCOR[10][4]:

> *Greases*, *stains*, and miscellaneous *soils* are usually sorbed onto the *soiled surface*. In most cases, these *soils* are taken up as *liquids* through *capillary action*. In an essentially static *system*, an *oil* cannot be replaced by *water* on a *surface* unless the *interfacial tensions* of the *water phase* are reduced by a *surface-active agent*.

Words in italics indicate the initial set of concepts[5]. Among the recognized concepts most appear only once, while the frequency of *surface* and *water* is two and the frequency of *soils* is 4 (includes *stains*). We have N =16 and C={*Greases*, *stains*, *soils*, *soiled*, *surface*, *soils*, *liquids*, *capillary action*, *system*, *oil*, *water*, *surface*, *interfacial tensions*, *water*, *phase*, *surface-active agent*} and thus get for instance

> $D(C, 0.1)$={*dirt stain soil*, *surface*, *surface tension*, *water*, *oil*}
>
> $D(C, 0.15)$={*dirt stain*, *soil*, *natural phenomenon*, *compound*}

5 Concluding remarks

In this paper we have considered how to use ontologies to provide data summaries with a special focus on textual data. Such summaries can be used in a querying approach where concepts describing documents, rather than documents directly, are retrieved as query answer. The summaries presented are conceptual due to fact that they exploit concepts from the text to be summarized and ontology-based because these concepts are drawn from a reference ontology.

[4]SEMCOR is a subset of the documents in the Brown corpus which has the advantage of being semantically tagged with senses from WordNet

[5]Notice that due to the use of SEMCOR there is no compound concepts in the initial set here.

ISBN: 978-989-95079-6-8

We have presented three summary principles. Two based on similarity and clustering and the third on probabilities derived from frequencies in the text to be summarized. Obviously a ”meaningful” clustering may lead to good summaries if characterizing subsuming concepts can be found in the ontology, but as indicated also counting occurrences, rather than only recognizing presence, of concepts may contribute to encircling essential concepts. Thus a next step should be to develop a combination that draws on similarity to ensure for instance that close summarizers are joined and on frequencies to ensure that derived summarizers are important.

For a summarization principle to work in practise users clearly need some kind of guidance on how many times to iterate or how to set a threshold. This brings up the important question on evaluation of summaries. Initial considerations on the quality of summaries can be found in [19] but the issue is also an obvious direction for further work in continuation of what has been described here.

References

[1] P. Anker Jensen, J. Fischer Nilsson Ontology-based Semantics for Prepositions, in Syntax and Semantics of Prepositions, P. Paint-Dizier (ed.), Text, Speech and Language Technology, Vol. 29, Springer, 2006.

[2] S. R. Aronson Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program, Proc AMIA Symp, pp 17-21, 2001.

[3] S. Abney Partial parsing via finite-state cascades, in: Proceedings of the ESSLLI´96 Robust Parsing Workshop, 1996.

[4] Olivier Bodenreider The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Research, 32, D267-D270. (2004)

[5] Udo Hahn, Inderjeet Mani The Challenges of Automatic Summarization Computer November 2000

[6] Gabor Melli, YangWang, Yudong Liu, Mehdi M. Kashani, Zhongmin Shi, Baohua Gu, Anoop Sarkar and Fred Popowich: Description of SQUASH, the SFU Question Answering Summary Handler for the DUC-2005 Summarization Task. Proceedings of DUC-2005, Vancouver, Canada pp 103-110 (205).

[7] Zhongmin Shi, Gabor Melli, Yang Wang, Yudong Liu, Baohua Gu, Mehdi M. Kashani, Anoop Sarkar and Fred Popowich Question Answering Summarization of Multiple Biomedical Documents Lecture Notes in Computer Science Springer Berlin / Heidelberg Volume 4509/2007 Advances in Artificial Intelligence pp 284-295 (2007)

[8] Andreasen, T.,Bulskov, H.: Conceptual Querying Through Ontologies. Fuzzy Sets and Systems, Elsevier, to appear (2008).

[9] Nilsson, J.F.: A logico-algebraic framework for ontologies – ONTOLOG. In Jensen, P.A., Skadhauge, P., eds.: First International OntoQuery Workshop, University of Southern Denmark (2001)

[10] Miller, G.A., Chodorow, M., Landes, S., Leacock, C., and Thomas R.G.: Using a semantic concordance for sense identification. In *Proc. of the ARPA Human Language Technology Workshop*, pages 240-243, 1994.

[11] Miller, G.A.: Wordnet: a lexical database for english. Commun. ACM **38**(11) (1995) 39–41

[12] Rada, R., Mili, H., Bicknell, E., Blettner, M.: Development and application of a metric on semantic nets. IEEE Transactions on Systems, Man, and Cybernetics **19**(1) (January 1989) 17–30

[13] Unified Medical Language System U.S. National Library of Medicine http://www.nlm.nih.gov/research/umls/

[14] Medical Literature Analysis and Retrieval System Online U.S. National Library of Medicine http://www.ncbi.nlm.nih.gov/pubmed/

[15] Bulskov, H., Knappe, R., Andreasen, T.: On measuring similarity for conceptual querying. In: FQAS ’02: Proceedings of the 5th International Conference on Flexible Query Answering Systems, Springer-Verlag (2002) 100–111

[16] Resnik, P.: Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language (1999)

[17] Andreasen, T., Knappe, R., Bulskov, H.: Domain-specific similarity and retrieval. Proceedings IFSA 2005, Tsinghua University Press (2005) 496–502

[18] Andreasen, T. and Jensen, P. Anker and Nilsson, J. Fischer and Paggio, P. and Pedersen, B. Sandford and Thomsen, H. Erdman: Content-based text querying with ontological descriptors,Data Knowledge Engineering ,vol 48(2), pp 199–219, Elsevier Science Publishers

[19] Yager, R.R., Petry, F.E.: A Multicriteria Approach to Data Summarization Using Concept Hierarchies IEEE Trans. on Fuzzy Sys. vol 14 no 6 (2006)

[20] Zhou, Xiaohua, Han, Hyoil Survey of word sense disambiguation approaches 18th FLAIRS Conference (2005)

ISBN: 978-989-95079-6-8

Fuzzy analogical model of adaptation for case-based reasoning

Bernadette Bouchon-Meunier [1], Christophe Marsala [1], Maria Rifqi [1]

1 Université Pierre et Marie Curie - Paris 6, CNRS UMR 7606, LIP6,
104 avenue du Président Kennedy, F-75016 Paris, France
Email: {bernadette-bouchon-meunier,christophe.marsala,maria.rifqi}@lip6.fr

Abstract— *Case-based reasoning is a recognized paradigm and has been explored in both applied and methodological directions. In the several phases of CBR, the adaptation phase is certainly the most problematic whereas the most characteristic and interesting phase. We propose to view this task through a fuzzy analogical scheme. The adaptation is realized by focusing on the relation existing between the problem to be solved and the retrieved cases. Two approaches are proposed here: the relation can be captured by a fuzzy linguistic modifier or by a fuzzy interpolation.*

Keywords— Case-based reasoning, analogy, interpolation, adaptation, fuzzy sets.

1 Introduction

Case-based reasoning (CBR) has been explored in both applied and methodological directions. If the retrieval phase of the process has been widely investigated, less modelling has been proposed for the adaptation phase [13]. Derivational analogy [10] has been proposed to solve this problem when general domain knowledge is available, but it remains mainly a domain-dependent approach. More formal analysis of transformational analogy, assuming knowledge about differences in problems and their solutions, has been considered. In particular, a formal model of transformational adaptation has been proposed [2] on the basis of the consideration of a quality function evaluation.

2 Adaptation modeling in CBR

2.1 Presentation of the problem

Let us consider a problem space P and the solution space S and P' the subset of P of already solved problems. For a problem p in P, case-based reasoning can be regarded as the research of a solution s in S which can be associated with p. Let us denote by IAS the application "is a solution of" p and its solution s: $IAS : P \rightarrow S$. A commonly used method is to look for the most resembling elements p' to p in P' already associated with a solution s' in S. Either the best fitting problem is only taken into account and its associated solution is considered, or a group of similar problems is retrieved and various methods can be used to take them into account, by means of an aggregation method or a prototype-based approach for instance.

After the retrieval step, it is generally necessary to incorporate the dissimilarity between p and p' to produce a transformed view s' of the solution s. The transformation is for instance based on a degree of uncertainty attached to the solution s, on the proximity between elements of S, on a modulation of the linguistic expression of s, or on the use of constraints.

Several types of situations occur. In the first type of situation, S is a finite set of crisp solutions and the element s which will be associated with a given p is one of them, associated with the most similar p' in P'.

For *example*, we consider the assignment of a student grant to attend a conference, P is a set of students characterized by several attributes, S the solution space. In this case, S will be reduced to a two-element space "grant, no grant". The student p will be assigned a grant if the most similar element in P' has been assigned a grant.

The only possible adaptation in this case is the assignment of a degree of uncertainty to s'. Obviously, if a decision is to be made, a level of acceptable uncertainty will be chosen and no decision will be possible if this level is not reached.

In the second type of situation, S is infinite. For a given element p in P, the element s which will be associated with p is constructed from the solutions associated with the most similar problems p' in P'. A particular case corresponds to a continuous universe S and a finite number of linguistic values represented by fuzzy sets of S.

For instance, continuing the previous example, we can consider that a grant assignment can be defined as {refused, small, medium, high}. For a given student description p, the grant assignment may be expressed as "rather small" or "very high" if the value of the grant is supposed to depend on his/her merits or characteristics (his/her country is very far away, his/her grades are very good…).

We propose in this paper several adaptation techniques derived from fuzzy set based analogical reasoning and modelling. Our purpose is to propose a general approach based on fuzzy analogy to define adaptation methods in CBR.

2.2 Fuzzy analogy as a model for adaptation in CBR

Problems and solutions are, in many cases, described by means of linguistic terms or approximate values derived from expert knowledge, for instance "if the quality of the paper is very high then the grant assignment is highly recommended". A convenient knowledge representation is thus fuzzy set based. Let us denote by $[0,1]^{\Omega}$ the set of fuzzy sets of any universe Ω.

ISBN: 978-989-95079-6-8

A number of works have already presented various utilizations of fuzzy logic based representations in CBR [1][11][12][14] or analogy [20]. Very little attention has been drawn to the adaptation problem and we focus on this aspect in the present paper.

We consider the universe D of descriptions of problems p present in P and the application $\alpha : P \to [0,1]^D$ such that a problem p is described by a fuzzy set $\alpha(p)$ of D. We consider also the universe E of expressions of solutions in S. Each solution in S is associated by means of a one-to-one mapping γ to a fuzzy set $\gamma(s)$ of E.
Our purpose is to provide a model of transformational adaptation in case-base reasoning, starting from the opportunity to use analogy [18] to describe the assignment of a solution to a problem.

A scheme of analogy has been defined in a fuzzy framework [5] in order to mimic the most classic human means of deriving a solution to a new problem on the basis of past experiences in an automatic approach.
Starting from this scheme, and considering the specific environment of case-based reasoning, we propose the following definition of an analogical scheme in CBR. We consider two similarity relations R_1 on $P \times P$ and R_2 on $S \times S$, such that $R_1(p,p') = 1$ if and only if p and p' are similar, $R_2(s,s') = 1$ if and only if s and s' are similar.

Definition 1. For a given application $IAS : P \to S$ and two similarity relations R_1 on $P \times P$ and R_2 on $S \times S$, an *analogical scheme* on (P,S) is a function (see figure 1):

$$\Re_{IAS,R_1R_2} : P \times S \times P \to S \qquad (1)$$

satisfying $\forall p \in P$ and $\forall s \in S$ such that $s = IAS(p)$
(i) $s = \Re_{IAS,R_1R_2}(p,s,p)$
(ii) $\forall p' \in P$ such that $R_1(p,p') = 1$, $s' = \Re_{IAS,R_1R_2}(p,s,p')$ if and only if $s' = IAS(p')$ and $R_2(s,s') = 1$.

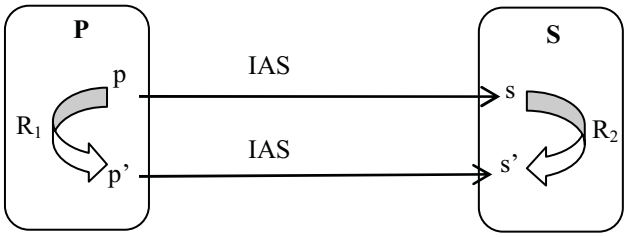


Figure 1. Global analogical scheme for case-based reasoning
x
Definition 1 means that if we want to solve a problem p' and we find a problem p similar to p' with a solution $s = IAS(p)$ in S, then the solution s' of p' will be similar to s. At this point, several problems p can be similar to p' and s' will be similar to all their solutions.
It is now necessary to define an operational way of evaluating the similarity R_1 of problems on the basis of their descriptions and the similarity R_2 of solutions on the basis of their expressions.

2.3 Measures of similitude

In many case-based reasoning applications, gradual evaluations of similarities are used, for instance based on distances between descriptions in D and between solutions in S or based on the compatibility with prototypes [19]. Let S_1 and S_2 be measures of similitude respectively defined on D and E to compare fuzzy sets. Several forms are available, depending of the properties we require from measures of similitude [8].
Following Tversky's seminal work on features of similarity [21], we define a measure of similitude on any universe Ω as follows, for a given *fuzzy set measure* $M : [0,1]^\Omega \to R^+$ such that $M(A) = 0 \Leftrightarrow A = \varnothing$ and M is monotonous with respect to the inclusion $\subseteq$ of fuzzy sets. We suppose also given an operation – of *difference* of fuzzy sets such that $A - A'$ is monotonous with respect to A and $A \subseteq A'$ implies $A - A' = \varnothing$.
With the convention that we use the same symbol for a fuzzy set and its membership function, the most used difference is defined for any y in Ω by:

$$A - A'(y) = \max(0, A'(y) - A(y)) \qquad (2)$$

Definition 2. A *measure of similitude* on Ω is a mapping $S_\Omega : [0,1]^\Omega \times [0,1]^\Omega \to [0,1]$ defined as:

$$S_\Omega(A,A') = F(M(A \cap A'), M(A' - A), M(A - A')) \qquad (3)$$

for a given mapping $F : R^{+3} \to [0,1]$ such that $F(u,v,w)$ is non-decreasing in u, non-increasing in v and w.

A measure of similitude can be regarded as a fuzzy relation on $[0,1]^\Omega \times [0,1]^\Omega$ and a fuzzy version of a similarity relation. Given a function F, we use measures of similitude on D and E to determine the similarities on P and S.
We assume that

$$R_1(p,p') = 1 \Leftrightarrow S_D(\alpha(p), \alpha(p')) \geq \varepsilon \qquad (4)$$

and $$R_2(s,s') = 1 \Leftrightarrow S_E(\gamma(s), \gamma(s')) \geq \varepsilon, \qquad (5)$$

for a chosen threshold ε.

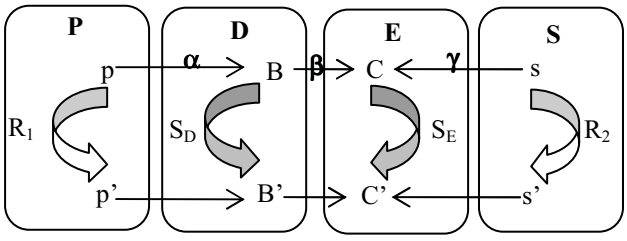


Figure 2. Operational analogical scheme for CBR

We obtain the operational analogical scheme described in Figure 2. The mapping β from D to E is defined in such a way that, for any B in $[0,1]^D$ and C in $[0,1]^E$, $C = \beta(B)$ if

ISBN: 978-989-95079-6-8

and only if there exist p in P and s in S satisfying $B = \alpha(p)$, $C = \gamma(s)$ and $s = IAS(p)$. We have $IAS = \gamma^{-1} \circ \beta \circ \alpha$.

Let us note that, crisp values being particular cases of fuzzy sets, the general scheme remains valid of values of attributes are precise. The measures of similitude must of course be chosen in an appropriate way [15].

2.4 Operational analogical scheme for CBR

It is now necessary to define an operational way of deriving s' from p' on the basis of the two measures of similitude S_D and S_E, and the analogical scheme $\Re_{IAS,R_1,R_2}$. We use a fuzzy version of analogical scheme as follows.

Definition 3. For a given application $\beta : [0,1]^D \to [0,1]^E$ and two measures of similitude S_D on $[0,1]^D \times [0,1]^D$ and S_E on $[0,1]^E \times [0,1]^E$, an $\varepsilon-$*analogical scheme* on (D,E) is a function defined for a given threshold $\varepsilon \in [0,1]$ as:

$$\Re^{\varepsilon}_{\beta,S_D S_E} : [0,1]^D \times [0,1]^E \times [0,1]^D \to [0,1]^E \quad (6)$$

satisfying the following conditions $\forall B \in [0,1]^D$ and $\forall C \in [0,1]^E$ such that $C = \beta(B)$

(iii) $C = \Re^{\varepsilon}_{\beta,S_D S_E}(B, C, B)$

(iv) $\forall B' \in [0,1]^D$ such that $S_D(B,B') \geq \varepsilon$, then

$C' = \Re^{\varepsilon}_{\beta S_D S_E}(B, C, B')$ if and only if $C' = \beta(B')$ and $S_E(C,C') \geq \varepsilon$.

This means that we need to determine a fuzzy $\varepsilon-$analogical scheme $\Re^{\varepsilon}_{\beta,S_D S_E}$ on $\left([0,1]^D, [0,1]^E\right)$ compatible with $\Re_{IAS,R_1R_2}$ (see figure 1). This compatibility is satisfied in the following way.

Property 1: Let us suppose that $IAS = \gamma^{-1} \circ \beta \circ \alpha$. For any p and p' in P, and s in S, the following holds:

$$s' = \Re_{IAS,R_1R_2}(p,s,p') \Leftrightarrow \gamma(s') = \Re^{\varepsilon}_{\beta,S_D S_E}(\alpha(p), \gamma(s), \alpha(p')).$$

Proof. If $p' = p$, then $s' = IAS(p') = IAS(p) = s$ by (i), also $\alpha(p') = \alpha(p)$ and $\gamma(s') = \gamma(s)$. Then (iii) is ensured.

Now, if $s' = \Re_{IAS,R_1R_2}(p,s,p')$, then $R_1(p,p') = 1$, $R_2(s,s') = 1$, and consequently the following inequalities hold:

$S_D(\alpha(p), \alpha(p')) \geq \varepsilon$, $S_E(\gamma(s), \gamma(s')) \geq \varepsilon$,

which entails $\gamma(s') = \Re^{\varepsilon}_{\beta,S_D S_E}(\alpha(p), \gamma(s), \alpha(p'))$ by (iv).

Conversely, if $\alpha(p') = \alpha(p)$, then $\gamma(s) = \gamma(s')$, and $s = s'$. Then (i) is ensured.

Now, if $\gamma(s') = \Re^{\varepsilon}_{\beta,S_D S_E}(\alpha(p), \gamma(s), \alpha(p'))$., then for $S_D(\alpha(p), \alpha(p')) \geq \varepsilon$, we get $\gamma(s') = \beta(\alpha(p'))$ and $S_E(\gamma(s), \gamma(s')) \geq \varepsilon$. For p and p' such that $S_1(p, p') = 1$, $s' = IAS(p')$ since γ is a one-to-one mapping, and $R_2(s, s') = 1$ by (5). Consequently, (ii) holds.

Continuing the previous *example*, we can consider a population P of students p, described by various attributes, such as the quality of their submitted paper, their ages, their countries… Let us consider universes of paper quality Q, of ages A, of countries C, we note D=Q×A×C and we consider fuzzy descriptions of students p, i.e. fuzzy sets of D, for instance (rather good quality, young, far country). Each student p will be assigned an amount of financial support in S which may be expressed by a fuzzy set of the universe of natural numbers E=IN, such as a "small grant".

3 Transformational adaptation in CBR

Since we have established a link between a comparison of problems in P or solutions in S and a comparison of their descriptions in D or expressions in E, we can work on (D,E) instead of (P,S) to determine methods of transformational adaptation in CBR.

3.1 Retrieval of a similar problem

We don't focus in this paper on the retrieval step of CBR. Let us only mention that there exist many ways to use the similarity relations R_1 to retrieve problems in P' which will be used to determine a solution in S to a given problem p.

For instance, the comparison can be restricted to typical problems in P'. The definition of prototypes can easily be solved through methods based on particular measures of similitude and dissimilarity in D [16]. Let P'' be the set of prototypes. They can be particular problems of P' or they can be abstract problems with descriptions in D summarizing the descriptions of a class of problems in P'.

The comparison can also be restricted to clusters of P' determined in agreement with solutions in S.

Let us remark that the threshold ε can be freely chosen by the user. If he/she chooses $\varepsilon = 0$, he/she takes into account all possible problems in P, and the determination of a unique p' associated with a given p lies in this case on the retrieval method. For instance, the retrieval method can be based on an optimization technique and p' will be the most similar to p, corresponding to the greatest value of $S_D(\alpha(p), \alpha(p'))$, under given additional constraints in some cases. Otherwise, the threshold ε can be considered as a level of decidability: if there exists no p' such that $S_D(\alpha(p), \alpha(p')) \geq \varepsilon$, then there is no already solved problem sufficiently similar to p and no solution can be proposed, unless more information is obtained.

If we now suppose that there exists a subset P_0 of such problems p', a method must be picked to determine a unique p'_0 in P_0 which will be used to find a solution to p by adapting $s'_0 = IAS(p'_0)$. Either we use an aggregation

ISBN: 978-989-95079-6-8

method to combine all p'_0 in P_0, or we look for the maximum of $\{S_D(\alpha(p),\alpha(p'))/\, p' \in P_0\}$, under additional constraints if necessary. Another solution is to look for the k-best values among $\{S_D(\alpha(p),\alpha(p'))/\, p' \in P_0\}$, with k chosen by the user, for instance $k = 2$. A refinement of the method considers a restriction of the search to elements of P_0 significantly different, with an additional threshold of decidability ε' such that, if the two best values in $\{S_D(\alpha(p),\alpha(p'))/\, p' \in P_0\}$ correspond to fuzzy subsets b and b' of D such that $|S_D(\alpha(p),b) - S_D(\alpha(p),b')| \le \varepsilon'$ and $\beta(b) \neq \beta(b')$, then no solution can be proposed to solve p, unless additional information is obtained.

We now focus on transformational adaptation techniques to determine the solution s of problem p.

3.2 Transformational adaptation by means of modifiers

We first consider that the retrieval method provides a unique p', such that $R_1(p,p') = 1$, best fitting the considered problem p. The value $S_D(\alpha(p),\alpha(p'))$ of the measure of similitude gives an information on the level of resemblance between p and p'. This value is used to adapt the solution $s = IAS(p)$ in order to define $s' = IAS(p')$. The $\varepsilon-$analogical scheme introduced in Definition 3 indicates that s will be determined in such a way that $S_E(\gamma(s),\gamma(s')) \ge \varepsilon$. It is clear that, $C' = \gamma(s')$ being a fuzzy set of D, there exist many fuzzy sets $C = \gamma(s)$ such that this condition is satisfied.

A convenient way to define C from C' is to use a modifier [4], such that s' will be a linguistically modified form of s. For example, if s is "small", then s' may be "rather small", or "more or less small", depending on the context.
A modifier defined for a universe Ω (Ω being either D or E) is a mapping $m : [0,1]^{\Omega} \to [0,1]^{\Omega}$.
We restrict ourselves to $\varepsilon-$*modifiers* m on Ω such that

$$M(m(X) - X) = 1 - \varepsilon \qquad (7)$$

The modifier is expansive if $X \subseteq m(X)$ for any X in $[0,1]^{\Omega}$. Examples of expansive modifiers are "more or less" or "approximately". Modifiers m are defined by parameterized mathematical transformations such as homotheties [3].
The transformational adaptation process is then the following: given the measure of similitude S_Ω, we define the solution s to problem p in such a way that $S_E(\gamma(s),\gamma(s')) \ge \varepsilon$. The form of m depends on the chosen measure of similitude and the parameters defining m are deduced from ε.

We have proved in [6] that the choice of a particular measure of similitude and the choice of the modifier are linked. Working with expansive modifiers, for instance, leads to use a particular measure of similitude called a measure of *inclusion*, defined as a reflexive measure of similitude (r.m.s.) such that $F(0,v,w) = 0$ for any v and w and $F(u,v,w)$ is independent of v. Classic examples of measures of inclusion are the following:

- $S_\Omega(X,X') = \dfrac{M(X \cap X')}{M(X)}$, (8)

M standing for the fuzzy cardinality of X,

$$M(X) = \int_{y \in \Omega} X(y) \qquad (9)$$

- $S_\Omega(X,X') = 1 - M(X - X')$, (10)

M standing for the height of X:

$$M(X) = \sup_{y \in \Omega} X(y). \qquad (11)$$

Property 2. Let us suppose that the difference of fuzzy sets is defined by (2), the fuzzy set measure by (11) and the similarity by (10). For any given problem p associated with a description B, an $\varepsilon-$modifier m provides the expression of a solution s of p compatible with the $\varepsilon-$analogical scheme i.e. such that $m(C') = \Re^{\varepsilon}_{\beta S_D S_E}(B',C',B)$ for any B,B',C'.

Proof. If m is an $\varepsilon-$modifier, then $M(m(C') - C') = 1 - \varepsilon$, which means that $\sup_{x \in E}(m(C') - C') = 1 - \varepsilon$. In the case where $B = B'$, then $S_D(B,B) = 1$, $\varepsilon = 1$. Then $M(m(C') - C') = 0$ by (7). Then $m(C') = C = C'$ and consequently $C' = \Re^{\varepsilon}_{\beta S_D S_E}(B',C',B')$, satisfying condition (iii). In the case where $\varepsilon \neq 1$, we have $m(C')$ is such that $m(C') = \gamma(s)$ by construction and $S_E(C',m(C')) = \varepsilon$ by (7). Consequently $m(C') = \Re^{\varepsilon}_{\beta S_D S_E}(B',C',B)$.

The properties of a measure of inclusion show that $S_\Omega(X,X')$ is independent of $X' - X$. This means that the choice of a measure of inclusion on E provides $S_E(m(C'),C')$ independent of $C' - m(C')$. This leads to use an expansive modifier m, since it corresponds to $C' - m(C') = \varnothing$ and $C' - m(C')$ has no influence on the value of a measure of inclusion.
An example of such an expansive $\varepsilon-$modifier is the following:

$$C(x) = m(C')(x) = \min\left(1, \frac{C'(x)}{\varepsilon}\right) \text{ for any } x \in E,$$

if we use (10) for M, s expressed as "approximately s'".

Considering again the previous *example*, if a "good" paper's author is assigned a "high grant", a similar paper's author is assigned an "approximately high" grant. The utilization of an expansive modifier reveals a cautious approach.

3.3 Transformational adaptation by means of modifiers in a gradual environment

In a different context, we consider the specific case where there is a gradual link between a variation on D and a

ISBN: 978-989-95079-6-8

variation on E. An example is expressed by "the better the quality of a student's paper, the higher the assigned grant".
Several methods are then possible to obtain a solution s to a problem p. We assume that D and E are included in the universe **R** of real numbers.

We consider again that the retrieval method provides a unique p', such that $R_1(p,p')=1$, best fitting the considered problem p. An appropriate form of modifier corresponds to translations of X to the left or to the right [9], with an amplitude of translation λ (positive or negative), to obtain the modified version $m(X)$. All components of the comparison between X and X' are taken into account, namely $X \cap X'$, $X - X'$ and $X' - X$.
The classic definition of a translatory modifier is the following, if the universe E is $[0,1]$, to simplify:

$m(C')(x) = C'(x+\lambda)$ if $x+\lambda \in [0,1]$

$m(C')(x) = C'(0)$ if $x+\lambda \leq 0$

$m(C')(x) = C'(1)$ if $x+\lambda \geq 1$.

This means that $m(C')$ is generally the translation of C', except at both ends of E. It is easy to check that $M(m(C') - C') = M(C' - m(C')) = |\lambda|$, except at both ends, where $M(m(C') - C') \leq |\lambda|$ or $M(C' - m(C')) \leq |\lambda|$, if we consider the fuzzy set measure (9).
Such a modifier is associated with any form of r.m.s., defined by $F(u,v,w)$, and depending on u, v, w. We can consider the measure of similitude (10) associated with the fuzzy set measure (9).
The amplitude must be defined in our case with respect to ε. Choosing $|\lambda| = 1-\varepsilon$ ensures that $S_E(m(C'),C') \geq \varepsilon$. We then obtain the following result :

Property 3. Let us suppose that the universe is $[0,1]$. If the difference of fuzzy sets is defined by (2), the fuzzy set measure by (9) and the similarity by (10), then a translatory modifier m defined by an amplitude ε such that $|\lambda| = 1-\varepsilon$ satisfies $m(C') = \Re^{\varepsilon}_{\beta S_D S_E}(B',C',B)$ for any B,B',C', according to definition 3.

The solutions we deduce in S from the use of such modifiers in E is more diverse. Depending on the sign and the value of the amplitude λ, depending also on the context, the solution s can take into account a reinforcement or a weakening of $\gamma^{-1}(C')$. In more concrete terms, if a student p' in P' is described by "good quality paper" in D and the solution $s' = IAS(p')$ expressed by "high assigned grant" in E, then a student p similar to p' will correspond to an assignment of "very high grant" (reinforcement) or "rather high grant" (weakening). To determine which of these two possibilities is the right one, more information must be taken into account about the relative position of B and B'.

3.4 Transformational adaptation by means of interpolation

We still consider that there is a gradual link between a variation on D and a variation on E. We now suppose that we take into account two problems in P' similar to p in order to find a solution $s = IAS(p)$ in S in a more comprehensive way.
The problems in P' are associated with descriptions represented by a family $D' = \{\alpha(p_1), \alpha(p_2)...\}$ of fuzzy sets of D and we further suppose that there exists an order $\prec$ on D' such that:

$$\alpha(p_1) \prec \alpha(p_2) \prec ... \prec \alpha(p_i) \prec \alpha(p_{i+1}) \prec ... \prec \alpha(p_n)$$

Let $E' = \{\gamma(s_1), \gamma(s_2)...\}$ be the family of associated expressions of solutions represented by fuzzy sets in E, with $s_i = IAS(p_i)$ for any $i = 1,2...$, supposed to be equipped with the same order $\prec$.
Without any loss of generality, we can then suppose that:

$$\gamma(s_1) \prec \gamma(s_2) \prec ... \prec \gamma(s_i) \prec \gamma(s_{i+1}) \prec ... \prec \gamma(s_n).$$

For a given p in P, we propose to look for the most similar p', namely p_i, and consider the pair (p_i, p_{i+1}) if $\alpha(p_i) \prec \alpha(p)$ or the pair (p_{i-1}, p_i) if $\alpha(p) \prec \alpha(p_i)$. For the sake of simplicity, we consider only the first case, such that $\alpha(p_i) \prec \alpha(p) \prec \alpha(p_{i+1})$.
Because of the assumption of graduality, we will have $\gamma(s_i) \prec \gamma(s) \prec \gamma(s_{i+1})$. The methods to determine s are again various. We propose to use a method based on interpolation [7][18]. The general spirit of the method is based on the basic analogical process "s is to s_i and s_{i+1} as p is to p_i and p_{i+1}".
The method to determine $C = \gamma(s)$ is based on the following steps: first, we compare $\alpha(p)$ to $(\alpha(p_i), \alpha(p_{i+1}))$. Secondly, we construct a family of possible fuzzy sets of E similar to the pair $(\gamma(s_i), \gamma(s_{i+1}))$ in a way analogous to the way $\alpha(p)$ is similar to the pair $(\alpha(p_i), \alpha(p_{i+1}))$. Finally, we deduce one element C from this family.

The detailed process we have proposed takes into account both location in the universe E and shape (understood as their membership function) of fuzzy sets and we proceed as follows.
Step1. We determine the location loc(C) of C, considering that it is to the locations of $\gamma(s_i)$ and $\gamma(s_{i+1})$ as the location of $\alpha(p)$ is to those of $\alpha(p_i)$ and $\alpha(p_{i+1})$.
Step 2. We translate $\alpha(p_i)$ and $\alpha(p_{i+1})$ towards $\alpha(p)$ with respect to locations, to obtain B'_i and B''_i.
Step 3. We compare the shapes of B'_i and B''_i to the shape of $\alpha(p)$.
Step 4. We translate $\gamma(s_i)$ and $\gamma(s_{i+1})$ to loc(C) to obtain respectively C'_i and C''_i.
Step 5. We construct two fuzzy sets C' and C'' of E with location loc(C) such that the shape of C' (resp. C'') can be

ISBN: 978-989-95079-6-8

compared to the shape of C'_i (resp. C''_i) in the same way as the shape of $\alpha(p)$ is compared with the shape of B'_i (resp. B''_i).

Step 6. We aggregate C' and C'' to construct C.

The interest of such a method appears when the family D' is sparse, which means that it is possible to find problems p falling into a "gap" between already solved problems, their similarity being relatively low. This method can also be used in ordinary cases, with a concern of quality of the solution, since taking into account two elements provides a more gradual, and then more robust, treatment of the problem.

4 Conclusion

We have pointed out several methods enabling the user to perform the transformational adaptation of the solution to a similar problem. These methods are diverse and their utilization depends on the context. Their interest is to ensure a gradual passage between cases and the global utilization of the set of already solved problems. Experiments will be available in the future to compare methods.

References

[1] K.S. Aggour, M. Pavese, P.P. Bonissone, W.E. Cheetham. SOFT-CBR: A self-optimizing fuzzy tool for case-based reasoning, *In Proceedings ICCBR 2003*, LNAI 2689, pp. 5-19, 2003.

[2] R. Bergmann, W. Wilke. Towards a New Formal Model of Transformational Adaptation in *Case-Based Reasoning. In Proceedings of the ECAI'98 Conference*, pp.53-57, 1998.

[3] B. Bouchon - Stability of linguistic modifiers compatible with a fuzzy logic ; in *Uncertainty and Intelligent Systems*, Lecture Notes in Computer Sciences n° 313, B. Bouchon, L. Saitta, R.R. Yager (eds.), Springer Verlag, pp. 63-70, 1988

[4] B. Bouchon-Meunier. Fuzzy logic and knowledge representation using linguistic modifier, in *Fuzzy logic for the Management of Uncertainty*, L.A. Zadeh, J. Kacprzyk, (eds.), John Wiley and Sons, 399-414, 1992.

[5] B. Bouchon-Meunier, J. Delechamp, C. Marsala, M. Rifqi. Several Forms of Fuzzy Analogical Reasoning, in *Proceedings 6th IEEE Int. Conf. on Fuzzy Systems, FUZZ-IEEE'97*, Barcelona, Spain, vol. I, pp. 45-50, 1997.

[6] B. Bouchon-Meunier, C. Marsala. Linguistic modifiers and measures of similarity or resemblance. In *Proceedings of the joint 9th IFSA World Congress and 20th NAFIPS International Conference*, Vancouver, Canada, vol. 4, pp. 2195-2199, 2001.

[7] B. Bouchon-Meunier, C. Marsala, M. Rifqi. Interpolative reasoning based on graduality, in *Proceedings of the FUZZ-IEEE'2000 Conference*, San Antonio, USA, pp.483-487, 2000.

[8] B. Bouchon-Meunier, M. Rifqi, S. Bothorel. Towards general measures of comparison of objects, *Fuzzy Sets and Systems*, 84, 2, 143-153, 1996.

[9] B. Bouchon-Meunier, Yao Jia. Linguistic modifiers and imprecise categories, *International Journal of Intelligent Systems* 7, pp.25-36, 1992

[10] J. G. Carbonell. Derivational Analogy: A theory of reconstructive problem solving and expertise acquisition. In *Machine Learning: An Artificial Intelligence Approach*. Michalski, R., Carbonell, J., & Mitchell, T. (Eds.) Morgan Kaufman Publishers: San Mateo, CA. pp. 371-392, 1986.

[11] D. Dubois, F. Esteva, P. Garcia, L. Godo. R. Lopez de Mantaras, H. Prade. Case-based reasoning: a fuzzy approach, *In Fuzzy Logic in Artificial Intelligence*, LNCS vol. 1566, pp. 79-90,1999.

[12] F. Esteva, P. Garcia, L. Godo, E. Ruspini, L. Valverde. On similarity logic and the generalized modus ponens, in *Proceedings of the FUZZ-IEEE'94 Conference*, Orlando, USA, pp. 1423-1427, 1994.

[13] B. Fuchs, A. Mille. A knowledge-level task model of adaptation in case-based reasoning, in *Proceedings of the ICCBR-99: International Conference on Case-Based Reasoning*, Lecture notes in computer science, vol. 1650, pp. 118-131, 1999.

[14] E. Hüllermeier, D. Dubois, H. Prade. Fuzzy rules in case-based reasoning, *In Conf AFIA99, Raisonnement à partir de cas*, pp. 45-54, 1999.

[15] M.-J. Lesot, M. Rifqi et H. Benhadda. Similarity measures for binary and numerical data: a survey. In *International Journal of Knowledge Engineering and Soft Data Paradigms (KESDP),* 1(1):63-84, 2009

[16] M.-J. Lesot, M. Rifqi, B. Bouchon-Meunier. Fuzzy prototypes: From a cognitive view to a machine learning principle, in H. Bustince, P. Herrera, J. Montero (eds.) *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models. Intelligent Systems from Decision Making to Data Mining, Web Intelligence and Computer Vision*, Springer, pp. 431-453, 2007.

[17] C. Marsala, B. Bouchon-Meunier. Interpolative Reasoning with Multi-Variable Rules. In *Proceedings of the joint 9th IFSA World Congress and 20th NAFIPS International Conference*, Vancouver, Canada, vol. 4, pp. 2476-2481, 2001..

[18] R. Schmidt, O. Vorobieva, L. Gierl. Case-Based Adaptation Problems in Medicine. In *Proceedings of WM2003: Professionelles Wissensmanagement Erfahrungen und Visionen*. Kollen-Verlag. 2003.

[19] R. Schmidt, T. Waligora, O. Vorobieva. Prototypes for Medical Case-Based Applications. In *Advances in Data Mining. Medical Applications, E-Commerce, Marketing, and Theoretical Aspects*, Springer, pp. 1-15, 2008.

[20] I. B. Turksen, Zhao Zhong. An approximate analogical reasoning approach based on similarity measures, *IEEE Trans. on Systems, Man and Cybernetics*, 18, 6, pp. 1049-1056, 1988.

[21] A. Tversky, Features of similarity. *Psychological. Review,* 84, pp. 327–352. 1977.

ISBN: 978-989-95079-6-8

Embedded Genetic Learning of Highly Interpretable Fuzzy Partitions

Jorge Casillas
CITIC-UGR (Research Center on Communication and Information Technology),
Dept. Computer Science and Artificial Intelligence, University of Granada, E-18071, Granada, Spain
E-mail: casillas@decsai.ugr.es, URL: http://decsai.ugr.es/∼casillas

Abstract— *A new algorithm is proposed to learn fuzzy partitions with a high interpretability degree. The set of input variables, the number of linguistic terms per variable, and the type (triangular or trapezoidal) and parameters of the membership functions are learnt by means of a meta-algorithm that uses a simple learning method to generate a fuzzy rule set from the derived fuzzy partitions. Interpretability constrains and powerful genetic operators are considered. A multi-objective optimization approach is used to generate different interpretability-accuracy tradeoffs. The algorithm is tested in a set of real-world regression problems with successful results compared to other methods.*

Keywords— Fuzzy rule-based systems, fuzzy partitions, granularity, interpretability, multi-objective genetic algorithms.

1 Introduction

In the last few years the interpretability (i.e. the capability of the fuzzy model to express the behavior of the real system in an understandable way) has received more and more attention in the fuzzy community. Indeed, this property of the fuzzy rule-based systems represents one of the main competitive advantages compared to other system modeling techniques.

In the literature we may find many alternatives to improve the interpretability such as reducing the number of fuzzy rules [1], selecting a subset of input variables [2], or using more compact fuzzy rule expressions [3]. Within these options, one of the more important approaches involves to learn the optimal number of linguistic terms per variable [4, 5, 6]. Indeed, to use a reduced number of linguistic terms is crucial to understand the meaning of the variables and directly influences on the fuzzy rule set size.

In this paper we propose a new algorithm to learn the number of input variables, the number of linguistic terms per variable, and the types of membership functions (triangular or trapezoidal) and their parameters with the aim of generating highly interpretable fuzzy partitions. The learning is performed by a wrapper-based embedded process where a meta-algorithm generates different fuzzy partitions and a simple learning method derives fuzzy rule sets from them.

The proposed algorithm, called EGLFP, has some interesting characteristics that make it very competitive: it uses strong fuzzy partition and includes distinguishability constrains for a better interpretability, and it uses variable-length coding schemes, powerful original crossover and mutation operators, and multi-objective optimization for a better search process.

The paper is organized as follows. Section 2 describes the proposed algorithm. Section 3 shows the results obtained in a set of real-world problems compared with other fuzzy rule learning methods. Finally, Sect. 4 concludes and suggests some further works.

2 EGLFP Algorithm

EGLFP is a multi-objective genetic algorithm with a generational evolutionary approach. A crossover operator that recombines membership function parameters, a first mutation operator that tunes these parameters, and a latter mutation operator that changes the fuzzy partition granularity (i.e., the number of linguistic terms) per variable are used in EGLFP. The multi-objective approach is based on the well-known NSGA-II algorithm [7]. The following sections detail the different components of EGLFP.

2.1 Coding Scheme

For the sake of a good interpretability and in order to reduce the search space tackled by the genetic algorithm, we propose the use of strong fuzzy partitions. Each gene (g) is a 2-tuple that contains the information related to a linguistic term of a specific variable. It consists of two real-valued fields (g_{left} and g_{right}) that encode the left and right extremes of the core of the associated fuzzy set (i.e., the semantic rule of the linguistic term) normalized to the interval $[0, 1]$. Figure 1 illustrates an example of gene's coding.

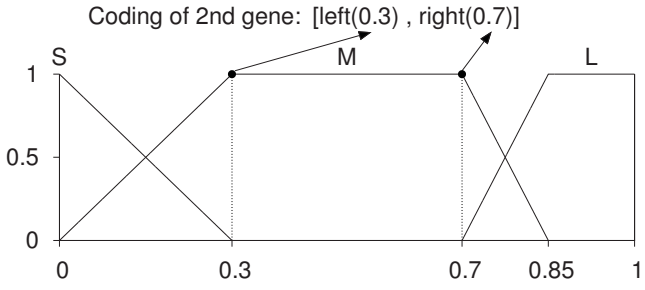


Figure 1: Example of a gene coding. Each gene, which encodes a linguistic term, consists of two fields (*left* and *right*) to encode the extremes of the fuzzy set's core (normalized to $[0, 1]$). Note that, since strong fuzzy partitions are used, the extremes of the core of each fuzzy set coincide with the extremes of the support of the adjacent fuzzy sets

A chromosome is a variable-length string (the length will depend on the number of linguistic terms used in each variable) of these genes that will encode the complete definition of the fuzzy partitions of all the input and output variables:

$$C = \bigcup_{v=1}^{n+m} \bigcup_{i=1}^{l_v} (g^v_{i,left}, g^v_{i,right}) \quad (1)$$

with n being the number of input variables, m the number of output variables, and l_v the number of linguistic terms used in the variable v.

ISBN: 978-989-95079-6-8

It is worth mentioning this coding scheme allows the absence of some variables by using only one linguistic term. Thus, if the variable v is not used, there will be only one gene associated to it with the value $(g^v_{1,left}, g^v_{1,right}) = (0, 1)$.

2.2 Initialization

The initialization process generates the first pool. Among the generated individuals, one of them will encode uniformly distributed triangular-shaped strong fuzzy partitions with the maximum number of linguistic terms allowed for each variable ($maxLV_v$). The remaining pool is filled up at random with individuals of two types.

One the one hand, some individuals (approximately a half of the pool) will preserve in some degree the above mentioned uniformly distributed fuzzy partitions but with a lesser number of linguistic terms. To do so, a fusion operation (which is described below in Sect. 2.5.1) on neighboring fuzzy sets is applied over the original fuzzy partitions. On the other hand, other individuals (approximately the another half of the pool) will be generated completely at random. To do so, firstly the number of linguistic terms is randomly assigned to each variable, then each fuzzy set adopts a triangular or trapezoidal shape at random and, finally, random values are assigned to the extremes of the fuzzy sets' cores.

2.3 Crossover

The crossover generates two individuals that inherit the membership function parameter definition given by two parents, even when these parents hold fuzzy partitions with different number of linguistic terms. This feature, the fact of crossing the parameters of fuzzy partitions with different granularities, provides the proposed algorithm with a powerful search process and is one of the original contributions of the paper.

The proposed crossover is a kind of parent-centric crossover operator [8] where each son is generated from a parent that plays the main role (*dominant parent*) and the another parent playing the secondary role (*recessive parent*). Thus, the son S_1 is generated focused on the parent C_1, but using the parent C_2 to add diversity; analogously with the son S_2.

Since the considered chromosomes have a variable length, a process is followed to select a gene of the recessive parent to be crossed with each gene of the dominant parent. This match is based on the distance (see eq. 5 below) where, given a fuzzy set of the dominant parent to be crossed, the closest fuzzy set in the recessive parent is chosen. Furthermore, to maintain the interpretability and well definition of the fuzzy partitions, an interval variation is fixed every time a gene is going to be crossed. An example is depicted in Fig. 2(a).

Once the dominant and recessive genes are fixed and the interval variation is determined, an original real-coding crossover operator is used. We have named this operator as the Constrained Parent-Centric Crossover (CPCX). Contrary to other previously proposed parent-centric crossover operators [8], CPCX is designed to generate the son's gene constrained to a given interval, which is crucial in our EGLFP algorithm to ensure well-defined and distinguishable fuzzy partitions. The operator is described in Algorithm 1 and an example is given in Figure 2(b).

Algorithm 1: Constrained Parent-Centric Crossover (CPCX) operator

Input: (g, h, min, max, m, M) with g being the dominant value and h the recessive value to be crossed, $[min, max]$ the variation interval of the dominant value, and $[m, M]$ the definition interval of the variable
Output: New value g'
begin
  if $g = h$ **then**
    $g' \leftarrow g$
  else if $h < g$ **then**
    $\alpha \leftarrow (g - h)/(g - m)$
    $l \leftarrow g - \alpha(g - min)$
    $r \leftarrow g$
    $g' \leftarrow U[l, r]$ */* Random number */*
  else
    $\alpha \leftarrow (h - g)/(M - g)$
    $l \leftarrow g$
    $r \leftarrow g + \alpha(max - g)$
    $g' \leftarrow U[l, r]$ */* Random number */*
end

2.4 Parameter Mutation

The parameter mutation changes the real-valued membership function parameter values of the input and output variables encoded in the chromosome. To do that, an original real-coding mutation operator is proposed in this paper. Firstly, a random process is followed to select the field of the gene to be mutated. When the gene to be mutated encodes a triangular fuzzy set, both *left* and *right* fields are mutated with the same value to preserve the original type.

Then, given a gene's field $g^v_{i,e}$ to be mutated, a variation interval around its value is defined with the aim of avoiding lack of distinguishability—i.e., two fuzzy sets very close—and other kinds of deformities that would decrease the interpretability degree. In this way, the following equations are used—Fig. 3(a) shows an example:

$$min_{g^v_{i,e}} = \begin{cases} min_v & \text{if } i = 1 \\ g^v_{i-1,right} + \delta & \text{if } e = \textit{left} \text{ or } g^v_{i,left} = g^v_{i,right} \\ g^v_{i,left} & \text{if } e = \textit{right} \text{ and } g^v_{i,left} \neq g^v_{i,right} \end{cases} \quad (2)$$

$$max_{g^v_{i,e}} = \begin{cases} max_v & \text{if } i = l_v \\ g^v_{i+1,left} - \delta & \text{if } e = \textit{right} \text{ or } g^v_{i,left} = g^v_{i,right} \\ g^v_{i,right} & \text{if } e = \textit{left} \text{ and } g^v_{i,left} \neq g^v_{i,right} \end{cases} \quad (3)$$

with $\delta = 1/(2 \cdot l_v)$ being the allowed minimum distance between the extremes of the cores of the fuzzy sets for the sake of a good distinguishability degree and l_v the number of linguistic terms used in the variable v.

Finally, an original real-coding mutation operator is applied on $g^v_{i,e}$ constrained to the interval $[min_{g^v_{i,e}}, max_{g^v_{i,e}}]$. We have named this operator as the Constrained Asymmetric Mutation (CAM), which is described in Algorithm 2. Figure 3(b) shows an example of the resulting asymmetric probability density function used to mutate a value.

ISBN: 978-989-95079-6-8

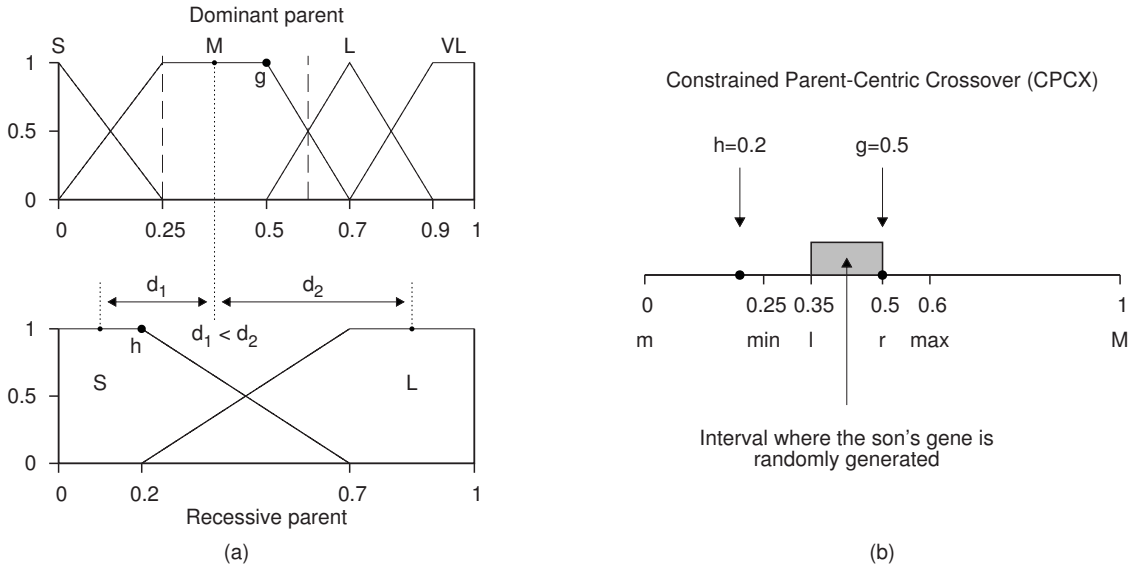


Figure 2: Example of crossover operation: (a) Given the gene g (which encodes the right extreme of the core of the fuzzy set labeled as S_M) of the dominant parent to be crossed, the allowed interval variation $[0.25, 0.6]$ is calculated (dashed lines) and the gene h that encodes the right extreme of the closest fuzzy set in the recessive parent (VS_S_M) is selected; (b) CPCX operator applied on the dominant gene $g = 0.5$, the recessive gene $h = 0.2$, the allowed interval variation $[min, max] = [0.25, 0.6]$, and the definition interval $[m, M] = [0, 1]$; a random value is generated in the interval $[l, r] = [0.35, 0.5]$ ($\alpha = 0.6$)

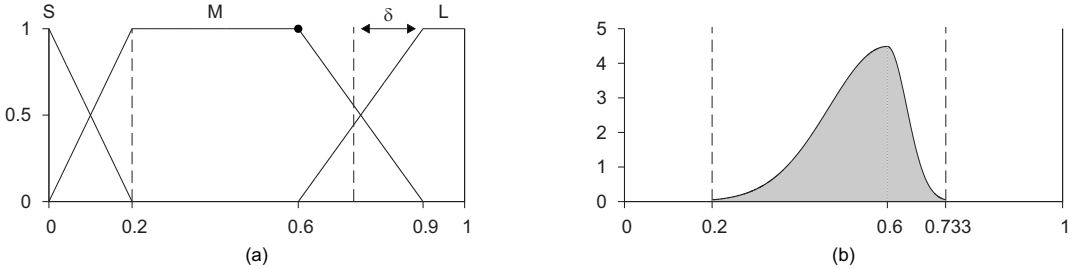


Figure 3: Example of parameter mutation operation: (a) the right extreme of the core of the fuzzy set M is selected to be mutated; then, the interval variation is defined by eq. 2 and 3 with $\delta = 0.167$ since there are three linguistic terms ($l_v = 3$); (b) asymmetric probability density function used in the CAM operator to mutate the right extreme of M's core constrained to the given interval $[0.2, 0.733]$

2.5 Granularity Mutation

The granularity mutation changes the number of linguistic terms in both input and output variables. To do that, it can *fuse* (merge) two neighboring fuzzy sets (thus decreasing in one the number of linguistic terms of the corresponding variable) or *fission* (split) a fuzzy set in two parts (thus increasing in one the number of linguistic terms of the corresponding variable). When fusion is applied on a variable with two linguistic terms, it involves removing the variable from the fuzzy model.

There are some constraints to apply fusion and fission. On the one hand, if fusion is applied on an input variable, this variable must have at least two linguistic terms and there must be more than one input variable with at least two linguistic terms (in order to avoid a fuzzy system without input variables after fusion). When fusion is applied on an output variable, this variable must have at least three linguistic terms since output variable removal is not allowed.

On the other hand, fission on variables with the maximum number of linguistic terms is not allowed. A second constraint is considered to avoid lack of distinguishability after fission. Indeed, if the core width of the fuzzy set associated to the gene to be fissioned is lower than a value inversely proportional to the number of linguistic terms of the corresponding variable, the core of the two fuzzy sets resulting from fission will be too close and, therefore, fission is not allowed. It involves that triangular fuzzy sets can not be fissioned since its core width is zero.

The two following subsections explain in detail how fusion and fission operate.

2.5.1 Fusion (merge) operation

Given a gene g_i^v to be fused, let h be the closest gene to g_i^v:

$$h = \begin{cases} g_{i-1}^v & \text{if } i = l_v \\ g_{i+1}^v & \text{if } i = 1 \\ g_{i-1}^v & \text{if } d(g_i^v, g_{i-1}^v) \leq d(g_i^v, g_{i+1}^v) \\ g_{i+1}^v & \text{otherwise} \end{cases} \quad (4)$$

with

$$d(g, h) = |(g_{left} + g_{right})/2 - (h_{left} + h_{right})/2| \quad (5)$$

To fuse g_i^v and h, the following steps are done:

- Firstly, if $g_{i,left}^v < h_{left}$ then $g_{i,right}^v \leftarrow h_{right}$; otherwise, $g_{i,left}^v \leftarrow h_{left}$.

ISBN: 978-989-95079-6-8

Algorithm 2: Constrained Asymmetric Mutation (CAM) operator

Input: (x, min, max) with x being the value to be mutated and $[min, max]$ its variation interval
Output: Mutated value x'
begin
  $p_{left} \leftarrow (x - min)/(max - min)$
  if $U[0,1] < p_{left}$ **then**
    $\sigma \leftarrow (x - min)/3$
    $x' \leftarrow N[x, \sigma]$ */* Normal random */*
    if $x' > (2x - min)$ **then** $x' \leftarrow 2x - min$
    else if $x' < min$ **then** $x' \leftarrow min$
    if $x' > x$ **then** $x' \leftarrow x - (x' - x)$
  else
    $\sigma \leftarrow (max - x)/3$
    $x' \leftarrow N[x, \sigma]$ */* Normal random */*
    if $x' > max$ **then** $x' \leftarrow max$
    else if $x' < (2x - max)$ **then** $x' \leftarrow 2x - max$
    if $x' < x$ **then** $x' \leftarrow x + (x - x')$
end

- Secondly, gene h is removed from the chromosome and the number of linguistic terms is decreased by one ($l_v \leftarrow l_v - 1$).

2.5.2 Fission (split) operation

The fission operation splits a trapezoidal fuzzy set into two triangular ones. Thus, given a gene g_i^v to be fissioned, the following steps are done:

- A new gene, h, is inserted at the right of g_i^v (h will be the new g_{i+1}^v) with $h_{left} \leftarrow g_{i,right}^v$ and $h_{right} \leftarrow g_{i,right}^v$. Set $g_{i,right}^v \leftarrow g_{i,left}^v$.
- The number of linguistic terms is increased by one ($l_v \leftarrow l_v + 1$).

2.6 Embedded Genetic Learning Approach

Every time a new individual (which encodes the fuzzy partition of each variable) is evaluated, a fuzzy rule set is firstly derived and then the complete fuzzy model (the fuzzy partitions plus the fuzzy rule set) is analyzed. To design the fuzzy rule set an efficient *ad hoc* data-driven method, the well-known Wang-Mendel (WM) algorithm [9] in this paper, is used. Therefore, a wrapper-based embedded genetic learning process is followed [4].

2.7 Inference Mechanism

We consider FITA (first inference, then aggregate) approach, the Max-Min inference scheme (i.e., T-conorm of maximum as aggregation and T-norm of minimum as relational operator), the T-norm of minimum as conjunction, and center-of-gravity as defuzzification. Moreover, the mean of the output domain is returned when no rules are fired for the given input data (this fact may only occur with test data since the algorithm ensures complete fuzzy rule set regarding training data).

2.8 Objective Functions

The multi-objective optimization performed in EGLFP is based on three objective functions to be minimized: O_1 assesses the error of the system, O_2 the complexity of the derived fuzzy rule set, and O_3 the complexity of the learnt fuzzy partitions. Thus, O_1 is focused on improving the accuracy of the fuzzy model while O_2 and O_3 improve its interpretability.

2.8.1 O_1, accuracy objective

The root mean squared error (RMSE) is used to compute the accuracy of the learnt fuzzy model S:

$$O_1(S) = RMSE(S) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(F_S(\mathbf{x}_k) - \mathbf{y}_k)^2} \quad (6)$$

2.8.2 O_2, rule set complexity objective

This second objective simply involves the final number of fuzzy rules obtained after applying the WM algorithm on the decoded fuzzy partitions as explained in Sect. 2.6:

$$O_2(S) = r(S) \quad (7)$$

2.8.3 O_3, fuzzy partition complexity objective

The third objective is the sum of the linguistic terms used in input and output variables:

$$O_3(S) = \sum_{v=1}^{n+m} L_v \quad \text{with} \quad L_v = \begin{cases} l_v & \text{if } l_v > 1 \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

The use of L_v is to not computing the removed input variables, where the number of linguistic terms is $l_v = 1$.

Since the fuzzy partition complexity has a direct influence on the fuzzy rule set complexity, it is expected that a low value of O_3 will imply a low value of O_2. Therefore, O_2 and O_3 are correlated in some degree. To improve the interpretability, O_2 seems to be more influent than O_3. However, we have decided to also use the objective O_3 with the aim of polishing good solutions by reducing the number of linguistic terms as much as possible.

This correlation will involve that, in the practice, the Pareto front shape with three objectives will be similar to the one with the two former ones and therefore, the inclusion of the third objective will not significantly degrade the multi-objective optimization quality.

2.9 Multi-objective Approach

A generational approach with the multi-objective elitist replacement strategy of NSGA-II [7] is used. Crowding distance in the objective function space is considered. Binary tournament selection based on the nondomination rank (or the crowding distance when both solutions belong to the same front) is applied. Instead fixing initial intervals for the objectives (which is not easy with the considered objectives) to compute the crowding distance, it is normalized for each objective according to the extreme values of the solutions contained in the analyzed front.

3 Experimental Analysis

This section includes the obtained results of the proposed EGLFP algorithm in six real-world regression problems where all the input and output variables are real-valued, and compares them with other fuzzy model learning methods.

ISBN: 978-989-95079-6-8

3.1 Problems and Learning Methods

We have considered the following regression problems: *Ele1* (the data set and partitions used in this paper are available at the author's website [1]), *Laser* (available at KEEL website [2]), *Ele2* [1], *DEE* [2], *Concrete* (obtained from the UCI Repository [3]), and *Comp-activ* (obtained from L. Torgo's website[4]).

Table 1 collects the main features of each problem, where *#InputVar* stands for the number of input variables, *#Exam* for the total number of examples, and $maxLV_v$ for the initial number of uniformly distributed triangular-shaped linguistic terms considered for each input variable in the experimental analysis. The output variable is always initially provided with seven uniformly distributed triangular-shaped linguistic terms.

Table 1: Data sets considered in the experimental analysis

Problem	#InputVar	#Exam	maxLV$_v$
Ele1	2	495	7
Laser	4	993	5
Ele2	4	1066	5
DEE	6	365	5
Concrete	8	1030	5
Comp-activ	21	8192	3

The experiments shown in this paper have been performed with a 5-fold cross validation. Thus, the data set is divided into five subsets of (approximately) equal size. The algorithm is then applied five times to each problem, each time leaving out one of the subsets from training, but using only the omitted subset to compute the test error.

We have considered two learning methods for comparison (both of them use the same inference engine described in Sect. 2.7 for our proposal). The *Wang and Mendel (WM)* [9] algorithm is a simple learning method that, in spite of not obtaining accurate results, is a traditional reference in the research community; the algorithm has been implemented by us. The *Cordón, Herrera, and Villar (CHV)* [4] algorithm is a competitive genetic fuzzy system that provides a learning flexibility similar to our EGLFP algorithm since it learns both the number of linguistic terms per variable and the membership function parameters. As in our case, this genetic learning process is also performed by wrapping the WM algorithm. We have used the original algorithm implementation provided by the authors.

Our algorithm has been run with the following parameter values: 1000 iterations, $p = 30$ as population size, $p_c = 0.7$ as crossover probability, and $p_{pm} = p_{gm} = 0.2$ as parameter and granularity mutation probability per chromosome, respectively. We have not performed any previous analysis to fix these values, so better results may probably be obtained by tuning them though we have informally noticed our algorithm is not specially sensitive to any parameter. The same parameter values are also used in the CHV algorithm in addition to $a = 0.35$ (for MMA crossover), $b = 5$ (for non-uniform mutation), and $\alpha = 0.2$ (weight of the number of rules in the fitness function).

3.2 Analysis

Table 2 collects the obtained results according to different quality measures such as approximation error, number of rules, number of linguistic terms, number of input variables, and fitness values. Average values of the five data partitions for each problem are reported.

We have included two versions of our algorithm in the comparative. The first one (EGLFP-1) is guided by a single fitness function in the same way as done in CHV [4] for the sake of a fair comparison. This function is a normalized aggregation of the error and the number of rules as follows:

$$f(S) = 0.8 \cdot MSE(S) + 0.2 \cdot \frac{MSE(S_{WM})}{r(S_{WM})} \cdot r(S) \quad (9)$$

with S_{WM} being the fuzzy model generated by the WM algorithm using uniformly-distributed triangular-shaped strong fuzzy partitions with the maximum number of linguistic terms allowed per variable ($maxLV_v$) and $MSE(S) = RMSE(S)^2$.

The second version (EGLFP-3) is guided by the above three-objectives multicriteria approach. Since in this case several solutions are returned in each run, we show five representative solutions from the final Pareto-optimal set. They are computed by sorting in ascending order the solutions according to the training RMSE and getting the 1 (i.e. best error), 25, 50 (i.e. mean), 75, and 100 (i.e. worst error) percentiles.

From the obtained results we can observe that our algorithm EGLFP-1 clearly overcomes, according to the fitness values (column Fit.), the CHV algorithm in all the considered problems. Indeed, our method generates more accurate fuzzy models in five cases. As regards the interpretability, it is not clear between EGLFP-1 and CHV which method obtains simpler models in terms of number of rules and number of linguistic terms since it depends on the problem. However, our method generates fuzzy models with more easily distinguishable fuzzy partitions thanks to the constraints tackled by the proposed genetic operators. Furthermore, CHV needs, in order to achieve the obtained accuracy values, to use badly formed fuzzy partitions where the extreme fuzzy sets of every variable may not hold normality within the corresponding domain.

If we analyze the results obtained by EGLFP-3 we can see that the process is able to generate a wide range of solutions with different accuracy-interpretability tradeoffs. It worths noticing that the most accurate solution of EGLFP-3 overcomes EGLFP-1 in both accuracy and interpretability in several problems. This fact is due to the niche-based search process caused by the use of the multi-objective approach, which leads the algorithm to a better exploration of the search space. It is also interesting to highlight the significant improvements in interpretability compared with the solutions provided by the WM algorithm. Very simple fuzzy models with a low number of variables, linguistic terms, and fuzzy rules are obtained while preserving a good accuracy.

4 Conclusion and Further Work

We have proposed a competitive genetic algorithm to simultaneously learn many components of a fuzzy model such as the

[1] J. Casillas. FMLib: fuzzy modeling library. http://decsai.ugr.es/~casillas/FMLib/

[2] KEEL: Knowledge extraction based on evolutionary learning. http://www.keel.es

[3] UC Irvine Machine Learning Repository. http://archive.ics.uci.edu/ml/

[4] L. Torgo. Collection of regression datasets. http://www.liacc.up.pt/~ltorgo/Regression/DataSets.html

ISBN: 978-989-95079-6-8

Table 2: Obtained results with $\mathbf{E}_{tra/tst}$ standing for the RMSE on the training/test data set (O_1, eq. 6), **#R** the number of fuzzy rules (O_2, eq. 7), **#L** the sum of the number of linguistic terms of each variable (O_3, eq. 8), **#I** the number of input variables, and **Fit.** the fitness (eq. 9)

Method	$\mathbf{E}_{tra}$	$\mathbf{E}_{tst}$	#R	#L	#I	Fit.
Ele1						
WM	650.743	674.910	22.0	21.0	2.0	
CHV	591.268	627.132	6.8	15.8	2.0	305924
EGLFP-1	558.793	618.857	12.8	15.0	2.0	298660
	560.346	638.527	17.0	16.0	2.0	315981
	577.568	632.652	11.1	12.9	2.0	
EGLFP-3	588.422	618.726	7.7	10.5	2.0	
	643.637	651.702	4.2	6.5	2.0	
	1017.088	988.744	1.8	4.0	1.0	
Laser						
WM	15.994	16.521	58.4	27.0	4.0	
CHV	8.791	10.602	58.8	25.0	4.0	97.694
EGLFP-1	8.509	9.750	33.6	18.6	4.0	87.611
	7.955	9.050	30.2	18.8	4.0	77.168
	9.570	11.099	17.8	14.8	3.8	
EGLFP-3	12.322	13.273	10.7	11.2	3.2	
	22.477	24.230	5.2	8.1	2.4	
	29.411	29.536	2.0	4.0	1.0	
Ele2						
WM	312.446	314.944	65.0	27.0	4.0	
CHV	235.053	238.263	22.6	20.4	4.0	50745
EGLFP-1	185.383	187.804	24.2	18.4	3.2	35371
	183.883	193.290	17.6	17.4	3.0	32462
	211.290	212.999	12.7	14.4	3.0	
EGLFP-3	314.150	314.470	6.9	10.6	3.0	
	553.541	566.163	3.7	8.1	2.0	
	903.109	917.182	2.0	4.0	1.0	
DEE						
WM	0.36302	0.47091	178.4	37.0	6.0	
CHV	0.34580	0.42109	112.0	32.8	0.0	0.11226
EGLFP-1	0.31916	0.41802	142.2	33.0	6.0	0.10254
	0.31795	0.43921	154.8	33.8	6.0	0.10376
	0.34773	0.43448	105.5	28.1	6.0	
EGLFP-3	0.37248	0.43070	50.5	21.4	5.2	
	0.41404	0.43448	13.3	13.5	3.5	
	0.79297	0.79075	2.0	4.0	1.0	
Concrete						
WM	8.2581	9.6227	309.8	47.0	8.0	
CHV	5.5686	7.4688	335.8	43.6	8.0	39.058
EGLFP-1	5.8652	7.1887	235.0	36.0	8.0	37.614
	5.6360	7.3113	324.8	41.2	8.0	39.286
	6.3962	7.1763	210.0	30.8	7.1	
EGLFP-3	7.2349	7.8924	91.4	22.2	5.8	
	8.9515	8.8944	16.8	12.3	3.9	
	14.9830	15.0084	2.0	4.0	1.0	
Comp-activ						
WM	11.9387	11.9762	425.6	70.0	21.0	
CHV	3.6004	3.6022	103.6	51.0	21.0	17.2945
EGLFP-1	3.3237	3.3848	14.0	19.4	6.4	9.7949
	3.3506	3.4277	12.6	17.6	6.6	9.8424
	3.4751	3.5417	8.6	14.1	5.3	
EGLFP-3	4.4530	4.5192	6.3	10.5	3.7	
	6.6188	6.6736	4.0	8.4	2.8	
	18.3997	18.3876	1.8	4.0	1.0	

subset of input variables, the number of linguistic terms per variable, the type of membership functions (trapezoidal or triangular), the membership function parameter values, and the fuzzy rule set. The algorithm is designed to generate highly legible and compact fuzzy partitions thanks to the use of interpretability constraints and powerful genetic operators.

The originality and good performance of the proposal mainly lies in the definition of new real-coding crossover and mutation operators that properly deals with the constraints for distinguishability, orderliness, and non-deformity imposed to the learnt fuzzy partitions; the design of a variable-length coding scheme and a parent-centric crossover capable of handling fuzzy partitions with different number of linguistic terms; and the use of several criteria and a multi-objective optimization scheme that provide with a range of different interpretability-accuracy tradeoffs and endow the algorithm with an effective niche-based search process. Successful results are obtained in six real-world regression problems with up to 21 continuous input variables.

As further work we will investigate the scalability of the algorithm to large-scale problems and the use of more criteria to assess the interpretability quality of the obtained fuzzy models.

Acknowledgment

This work was supported in part by the Spanish Ministry of Science and Innovation (grant no. TIN2008-06681-C06-01) and the Andalusian Government (grant no. P07-TIC-3185).

References

[1] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka. Selecting fuzzy if-then rules for classification problems using genetic algorithms. *IEEE Transactions on Fuzzy Systems*, 3(3):260–270, 1995.

[2] T.-P. Hong and J.-B. Chen. Finding relevant attributes and membership functions. *Fuzzy Sets and Systems*, 103(3):389–404, 1999.

[3] A. González and R. Pérez. Completeness and consistency conditions for learning fuzzy rules. *Fuzzy Sets and Systems*, 96(1):37–51, 1998.

[4] O. Cordón, F. Herrera, and P. Villar. Generating the knowledge base of a fuzzy rule-based system by the genetic learning of the data base. *IEEE Transactions on Fuzzy Systems*, 9(4):667–674, 2001.

[5] J. Espinosa and J. Vandewalle. Constructing fuzzy models with linguistic integrity from numerical data-AFRELI algorithm. *IEEE Transactions on Fuzzy Systems*, 8(5):591–600, 2000.

[6] J.-N. Choi, S.-K. Oh, and W. Pedrycz. Identification of fuzzy models using a successive tuning method with a variant identification ratio. *Fuzzy Sets and Systems*, 159(21):2873–2889, 2008.

[7] K. Deb, A. Pratap, S. Agarwal, and T. Meyarevian. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002.

[8] C. García-Martínez, M. Lozano, F. Herrera, D. Molina, and A. M. Sánchez. Global and local real-coded genetic algorithms based on parent-centric crossover operators. *European Journal of Operational Research*, 185:1088–1113, 2008.

[9] L.-X. Wang and J. M. Mendel. Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(6):1414–1427, 1992.

ISBN: 978-989-95079-6-8

Non Additive Measures for Group Multi Attribute Decision Models

Marta Cardin, Silvio Giove

Dept. of Applied Mathematics, University of Venice
Venice, Italy
Email: mcardin@unive.it, sgiove@unive.it

Abstract— *This paper extends the Choquet integral, widely used in multi-attribute decision problems, to the non monotone case in the context of Group Decision Theory. Even if not so often, preference structures which violate the monotonicity axiom can be observed in real applications. Our aim is twofold. First, we propose the Choquet integral with non monotone non additive measure. Then, we apply the Choquet integral in the context of multi person decision problem, a typical framework of many real world applications, for which the Choquet integral was rarely proposed. Thus in our model this aggregation function is applied twice, both in the cases with possible negative interactions. For this reason, our proposal can be defined as two-step signed Choquet integral.*

Keywords— Non additive measures, Choquet integral, Group Decision Theory.

1 Introduction

In Multi Attribute decision models, non additive measures and the Choquet integral have been intensively applied in many real world decision problems. Despite simpler approaches, like simple additive weighting (SAW), non additive measures can model *interactions* between the criteria. This method is a very general aggregation tool, including as special cases many aggregation algorithms, as SAW, min and max operators, k-statistics, OWA, and many others. Nevertheless, monotonicity is usually considered a strict requirement. Only few papers explicitly considered the non monotone case. Even if rarer than for *preferential independence* axiom, some violations of the monotonicity axiom can be observed in the Decision Maker preference structure. Moreover, even if many real world applications exist so far, little results were obtained in the context of Group Decision Theory. This family of problem deals with complex decision about some alternatives that need to be scored by a tool of Experts, see [4] and the references therein. To simplify, suppose that an Experts Committee is demanded to evaluate a finite set of development strategies on the basis of some criteria. Non additive measures can help to solve this type of *multi person - multi attribute* decision problems.
Negative(non monotone) interactions among the criteria can be modeled by means of *signed* measures.
The aim of the paper is twofold; first we consider the possibility of *negative* interactions among the criteria, and then we apply a non additive aggregation algorithm in the context of Group Decision Theory. In this sense, we apply twice the Choquet integral, to aggregate the Expert's opinion for each alternative, and subsequently to aggregate the Expert alternative scores into a global numerical evaluation (*Group alternative score*). The latter is usually performed by a Decision Maker, which, on the basis of his confidence about the Committee members experience and/or capacity, averages their judgements not necessarily in a linear way. Thus we can define our approach as a two-step Choquet integral with *signed* measures, i.e. *two-step signed Choquet integral*. The paper is organized as follows. The next Section briefly resumes some definitions and properties of non additive measures and of the Choquet integral. Section 3 describes our proposed model, which is deeply analyzed in the following Section 4. Section 5 reports the application of the two-step signed Choquet integral to multi person - multi attribute decision problems, and finally in the last section a numerical example is reported.

2 The discrete Choquet integral

Aggregation has for purpose the simultaneous use of different pieces of information provided by several sources, in order to come to a conclusion or a decision. So aggregation functions transform a finite number of inputs, called arguments, into a single output. They are applied in many different domains and, in particular, aggregation functions play an important role in different approaches to decision making, where values to be aggregated are typically preferences or satisfaction degrees. Many functions of different type have been considered in connection with different situations and various properties of these functionals can be imposed by the nature of the considered aggregation problem. A class of aggregation operators can also be introduced axiomatically by means of a set of properties.
We denote by E a non empty real interval. If the integer n represents the number of values to be aggregated an aggregation operator is a function $A : E^n \to E$. To motivate the use of the Choquet integral as an aggregation operator, we present some basic mathematical properties of the aggregation functions.

- **Monotonicity** For all $\mathbf{x}, \mathbf{y} \in E^n$ if $x_i \leq y_i (i = 1, \ldots, n)$ then $A(\mathbf{x}) \leq A(\mathbf{y})$
- **Positive Homogeneity** If $\mathbf{x} \in E^n$ and $a \in \mathbb{R}, a > 0$ then $A(a\mathbf{x}) = aA\mathbf{x}$

Moreover we define x_{-i} the element of $\mathbb{R}^{n-1}$ that is obtained from x by eliminating component i, and we denote (x_{-i}, y_i) as obtained from x by replacing x_i with y_i.
Now we present the concept of comonotonicity.

Definition 2.1 *If x, y are elements of $\mathbb{R}^n$ then x, y are said* comonotonic *if $x_i < x_j$ implies that $y_i \leq y_j$.*

Two vectors x, y are comonotonic if they have the same ranking of their components or there exists a permutation $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ such that $x_{\sigma(i)}$ and the corresponding $y_{\sigma(i)}$ are arranged in nondecreasing order.

ISBN: 978-989-95079-6-8

Our representation result depends on the following axioms related to the concept of comonotonicity.

- **Comonotonic Monotonicity** If $\mathbf{x}, \mathbf{y} \in E^n$ are comonotonic and $x_i \leq y_i$ $(i = 1, \ldots, n)$ then $A(\mathbf{x}) \leq A(\mathbf{y})$
- **Comonotonic Separability** If $\mathbf{x}, \mathbf{y} \in E^n$ are comonotonic then for every i, $A(x_{-i}, x_i) \geq A(y_{-i}, x_i)$ iff $A(x_{-i}, y_i) \geq A(y_{-i}, y_i)$

The comonotonic separability axiom is obviously a variation of an additive separability axiom and it has been applied successfully in decision making under risk and uncertainty, see [5]and the references therein. It states that preferences between alternatives depend only on the components that differ between the vectors under cons ideration, as long as these alternatives maintain the attributes' ordering.
In order to introduce a non-additive approach to aggregation operators we propose a non-additive integral operator and so we consider the integral as a particular averaging operator. The use of variants of the Choquet integral allows some flexibility in the way criteria are combined. Non additive measures are in the current literature a commonly used method to aggregate numerical information. This is particularly due to the fact that in cooperation with integral aggregation functions, they are a well-founded framework able to capture interactions among the involved variables. Usually the monotonicity property is required in most of practical applications. Monotone measures can model both *synergic* and *redundance* interactions among the criteria. Conversely, they are unable to represent the *neutralization* effect. Such a situation appears, for instance, where increasing one criterion by alone has a positive effect, but the contemporary increase of two criteria has a negative effect. Of course, such a situation cannot be represented by monotone measures.
In particular in this paper we consider a non-monotone Choquet integral as in [5], [8] and [12] and we define a non-monotone Choquet measure and a Choquet integral for a n-dimensional vector.
If N is a finite index set $N = \{1, \ldots, n\}$, a real valued set function $\upsilon : 2^N \to \mathbb{R}$, with $\upsilon(\emptyset) = 0$, is called a *non additive signed measure*. If $A \subseteq B \subseteq N$ implies that $\upsilon(A) \leq \upsilon(B)$, then the function is said to be *monotone* and υ is called a *non-additive measure*. If υ is a non-additive measure then $\upsilon(A) \geq 0$ for all $A \in 2^N$. We say that a measure is *additive* if $\upsilon(A \cup B) = \upsilon(A) + \upsilon(B)$.
We note that if $S \subseteq N$, $\upsilon(S)$ can be viewed as the *importance* of the set of elements S. We introduce now the discrete Choquet integral on N viewed as an aggregation function that generalizes the weighted arithmetic mean.

Definition 2.2 *The* Choquet integral C_υ *of a vector* $\boldsymbol{x} \in \mathbb{R}^n$ *with respect to a non-additive signed measure* υ *is*

$$C_\upsilon(\boldsymbol{x}) := \int_0^{+\infty} \upsilon(\boldsymbol{x} \geq \alpha)\, d\alpha + \int_{-\infty}^{0} (\upsilon(\boldsymbol{x} \geq \alpha) - \upsilon(N))\, d\alpha$$

where $\upsilon(\boldsymbol{x} \geq \alpha) = \upsilon(\{i \in N : x(i) \geq \alpha\})$.

This formula can be interpreted as an expectation operator with respect to a generalized measure. Then, the Choquet integral of a vector $\mathbf{x} \in \mathbb{R}^n$ with respect to a non-additive measure υ can be represented as the following weighted sum:

$$C_\upsilon(\mathbf{x}) = \sum_{1=1}^{n} (x_{(i)} - x_{(i-1)}) \upsilon(A_{(i)}) \tag{1}$$

and $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ is such that $x_{\sigma(i)}$ are arranged in nondecreasing order and $A_{(i)} = \{i, \ldots n\}$ and $A_{(n+1)} = \emptyset$. The Choquet integral can be also computed using the Möbius representation of υ, see [8].

Proposition 2.1 *The Choquet integral* $C_\upsilon : \mathbb{R}^n \to \mathbb{R}$ *can be written as*

$$C_\upsilon(\boldsymbol{x}) = \sum_{T \subseteq N} a(T) \bigwedge_{i \in T} x_i, \qquad \boldsymbol{x} \in \mathbb{R}^n$$

where a *is the Möbius representation of* υ.

The Choquet integral has important properties for aggregation.

Proposition 2.2 *Let* A *be an aggregation function defined on* E^n.

i) If A *is a non-monotone Choquet integral then it is positive homogeneous, comonotonic monotone and comonotonic additive.*

ii) A *is a non-monotone Choquet integral if it satisfies positive homogeneity and comonotonic separability.*

iii) A *is a non-monotone Choquet integral if it is continuous and comonotonic additive.*

Proof The proof of part i) is immediate by the definition of Choquet integral.
Now it is important to note that a comonotonic separable aggregation function A is comonotonic additive that is if $\mathbf{x}, \mathbf{y} \in E^n$ are comonotonic then $A(\mathbf{x} + \mathbf{y}) = A(\mathbf{x}) + A(\mathbf{y})$. In fact if $\mathbf{x}, \mathbf{y}, \mathbf{z} \in E^n$ are comonotonic and $A(\mathbf{x}) \geq A(\mathbf{y})$ then by comonotonic separability $A(\mathbf{x} + \mathbf{z}) \geq A(\mathbf{y} + \mathbf{z})$. Then if we consider two comonotonic vectors $\mathbf{x}, \mathbf{y} \in E^n$ there exist $c, d \in E$ such that $A(\mathbf{x}) = \mathbf{c} = (c, \ldots, c)$ and $A(\mathbf{y}) = \mathbf{d} = (d, \ldots, d)$. By the fact that each comonotonic vector is comonotonic with each other vector with equal component we can prove $A(\mathbf{x} + \mathbf{y}) = A(\mathbf{c} + \mathbf{y}) = A(\mathbf{c} + \mathbf{d}) = A(\mathbf{c}) + A(\mathbf{d}) = A(\mathbf{x}) + A(\mathbf{y})$. Now we can conclude since a homogeneous function that satisfies comonotonic additivity is a non-monotone Choquet integral by theorem 1 of [8] and a continuous function that satisfies comonotonic additivity is a non-monotone Choquet integral by corollary 2.2 of [14].

It is elementary verified that if A satisfies the hypothesis of proposition 1.1 and it is also monotone then it is a monotone Choquet integral.

3 The model

We consider a decision problem where $K = \{1, \ldots, k\}$ Experts have to score M alternatives on the basis of some criteria, each of them normalized in the common scale [0,1]. The preference structure of the k-th Expert, $(1 \leq k \leq K)$, is represented by a set of non additive measures m_k, defined on the space of criteria. Despite the majority of Multi Criteria Group decision models, we admit that each Expert is characterized

ISBN: 978-989-95079-6-8

by his own criteria set, even if the sets can partially overlap or coincide as a particular case. Moreover, the preference structure of each Expert will be represented by a *signed* measure, including the possibility of *neutralization* effects among the criteria themselves. Finally, at higher level, an other measure u is used to aggregate the individual opinions into a global score. Assuming that a Decision Maker (DM) has to score the alternatives on the basis of the judgements expressed by the Experts, the measure u can be interpreted as the degree of confidence that the DM feels towards the Experts themselves, or towards each coalition of Experts. Since the Choquet integral is twice applied, we deal with a *two-step* Choquet integral [11][1]. For a better problem formulation, let us consider the following definitions.
i) The criteria set for the k-th Expert E_k is the set $N_i = \{c_{i_1}, \ldots, c_{i_n}\}$, with $n_i = |N_i|$ is the cardinality of N_i. It follows that $\mathbf{x}^i = (x_{i_1}, \ldots, x_{i_n})$ is the *profile* of criteria associated with an alternative by the Expert E_k and an alternative is characterized by a vector $\mathbf{x}$ in $\mathbb{R}^n$ where $n = \sum_{i=1}^{K} n_i$.

ii) The preference structure of the k-th Expert is represented by a non additive signed measure m_k, defined on 2^{N_k}.

iii) There are M alternatives, each of them characterized by the vector $\mathbf{x}_j$ in $\mathbb{R}^n$, $j = 1, .., M$.

Let $v_k(\mathbf{x}) = \int \mathbf{x}^k \, d\, m_k$ for $1 \leq k \leq K$ and $v(\mathbf{x}) = (v_1(\mathbf{x}), \ldots, v_K(\mathbf{x}))$.
Then we define the two-step Choquet integral $C(\mathbf{x})$,

$$C(\mathbf{x}) = \int v(\mathbf{x}) \, d\, u.$$

4 Multi-step non monotone Choquet integral

In this section we give some basic definitions and we present some results on multi-step non monotone Choquet integral. The two-step (monotone) Choquet integral has been investigated mainly in [13] and [11]. Let us now give a formal definition of a multi-step non-monotone Choquet integral.

Definition 4.1 *Let $\Gamma \subseteq \mathbb{R}^n$. For any $i \in N$, the projection $\mathbf{x} \mapsto x_i$ is a* 0-step non-monotone Choquet integral. *Let us consider $F_i : \Gamma \to \mathbb{R}$, $i \in M := \{1, \ldots, m\}$, being k_i-step non-monotone Choquet integrals, and a non-additive signed measure v on M. Then*

$$F(\mathbf{x}) := C_v(F_1(\mathbf{x}), \ldots, F_m(\mathbf{x}))$$

is a k-step non-monotone Choquet integral*, with $k := \max\{k_1, \ldots, k_m\} + 1$. A* multi-step non-monotone Choquet integral *is a k-step non-monotone Choquet integral for some integer $k > 1$.*

Let us first recall the concept of piecewise linear function.

Definition 4.2 *A real-valued function F on a convex closed subset $\Gamma \subseteq \mathbb{R}^n$ is* piecewise linear *if Γ can be written as a union of closed subspaces $\Gamma_1, \ldots, \Gamma_q$ of the same dimension as Γ, such that $F|_{\Gamma_i}$ is linear, $i = 1, \ldots, q$. A linear function G on $\mathbb{R}^n$ which coincides with F on some Γ_i is a* component *of F.*

[1] Anywise, in the quoted reference the two-step Choquet integral was limited to the monotone case only.

The following proposition gives the basic properties of non-monotone multi-step Choquet integrals.

Proposition 4.1 *The non-monotone multi-step Choquet integral is a continuous, positively homogeneous and piecewise linear function.*

Proof The non monotone Choquet integral is a piecewise linear and positively homogeneous function, then also the multi-step non monotone Choquet integral. Clearly, any piecewise linear function is continuous.

Recall that the multi-step Choquet integral is not comonotonic additive, in general, and hence it cannot be described by a 1-step Choquet integral.

Proposition 4.2 *If F is a 2-step non-monotone Choquet integral and the measure of the second level is additive then F coincides with a 1-step non-monotone Choquet integral.*

Proof If we consider two comonotonic vectors $\mathbf{x}, \mathbf{y}$ then for every $i = 1, \ldots, m$ $F_i(\mathbf{x} + \mathbf{y}) = F_i(\mathbf{x}) + F_i(\mathbf{y})$ since F_i is comonotonic additive. Then it is easy to prove that $F(\mathbf{x}+\mathbf{y}) = F(\mathbf{x}) + F(\mathbf{y})$ where $F(\mathbf{x}) = C_v(F_1(\mathbf{x}), \ldots, F_m(\mathbf{x}))$ and C_v is a linear functional. Then F is a positively homogeneous and comonotonic additive functional. Now the positively homogeneous function F that satisfies comonotonic additivity is a non-monotone Choquet integral by theorem 1 of [8].

5 A Multi Criteria Group Decision problem

In the proposed model, the two-step Choquet integral will be used to compute the score of any alternative, proceeding in two sequential phases. Firstly, the individual measures m_k defined on 2^{N_k} are used to compute the individual score. Subsequently all the individual scores are aggregated with the measure u on the space 2^K. Both in the cases signed measure can be admitted, modeling negative interactions.
The measure u represents the preference structure of the DM about the Expert's coalitions. For instance, consider a rule like the following one

If the first and the second Experts score equally an alternative, then score this alternative with their common value, despite any other opinion of the remaining Experts of the group

This rule can be implemented assigning one to the coalition formed by the Experts n. $1, 2$, that is $m\{X\} = 1$ if $A \cap X \neq \emptyset$ with $A = \{1, 2\}$.
The steps of the aggregation algorithms are then the following ones

i) $\forall k$, $1 \leq i \leq k$, aggregate each of the individual measure using the Choquet integral: $v_k(\mathbf{x}) = \int \mathbf{x}^k \, d\, m_k$, obtaining the individual score for the alternative ($\mathbf{x}$)

ii) at the top level (the DM level), aggregate the individual ranking, obtained at the previous step, using the measure m on the space 2^K applying again the Choquet integral: $C(\mathbf{x}) = \int v(\mathbf{x}) \, d\, u$.

ISBN: 978-989-95079-6-8

Let us recall that the *relative importance* for monotone measure of a single criterion is usually measured by a suitable index like the Shapley index, [6]. The Shapley index for the i-th criterion and for the k-th Expert is given by

$$p_k(i) = \sum_{T \subseteq N \setminus \{i\}} \frac{(n_i - t - 1)!t!}{n_i!}[m_k(T \cup i) - m_k(T)] \quad (2)$$

with $t = card(T)$. The Shapley index measures the *average relative importance* of a criterion and varies between zero and one. In the case of non monotone measure, it can happen that the Shapley index is close to zero even if the criterion is *important*. This is due to possible conflicting interactions that compensate positive marginal gains with negative ones. To this purpose, in [1] the *extended* Shapley index was introduced:

$$q_k(i) = \sum_{T \subseteq N \setminus \{i\}} \frac{(n_i - t - 1)!t!}{n_i!}|m_k(T \cup i) - m_k(T)| \quad (3)$$

for the i-th criterion and for the k-th Expert. Both the *extended* Shapley indices measures the *relative* importance of the criterion. If the measure is monotone, the two indices coincide. Otherwise, it can happen that the Shapley index is close to zero, while the *extended* index is high. This means that the criterion is *important*, but, on average, it is *neutral*, it is neither a benefit, nor a cost (for some coalitions it is a cost, while for others it is a benefit). Then, in the non monotone case, both the two indices should be considered.

For a better comprehension of the proposed methodology, consider the following multi person - multi attribute decision problem. Suppose that an Expert Committee is required to evaluate among some different investment projects that are the finite set of the available alternatives. Each project is characterized by a finite set of criteria. Every Expert scores independently any coalition of the criteria. Anywise, the Experts are not forced to consider *all* the criteria. Conversely, each Expert can consider an his *own* subset of criteria. This preference is formalized by a non additive signed measure, which can model *strict conflict* between the criteria. For instance, an Expert can consider three criteria valid for the judgements of the alternatives, and even if all of them, considered by alone, can be seen as *benefit*, the subset of the first and the second one induce a conflict, so that such a coalition receive an inferior score than the minimum between the first and the second criterion scores, which is usually implied in the monotone case.

6 A numerical example

Consider a decision problem where two alternatives have to be scored, on the basis of the four criteria values (alternative *profile*) reported in Table 1.

Suppose that three Experts are involved in the decision process. Then the DM gives a final evaluation of the alternatives on the basis of the Expert's individual scores. As above described, the decision process splits into two subsequent phases: the Expert's scoring, and the aggregation of each Expert's score into an aggregated score. If signed non additive measures are used for the aggregation in both the activities, the problem can be approached with a two-step non monotone Choquet integral.

Table 1: Alternative profile

x_1	x_2	x_3	x_4
0.4	0.2	0.8	0.6
0.1	0.7	0.4	0.5

Table 2: Expert n. 1

$m^1(1)$	$m^1(2)$	$m^1(1,2)$
0.4	0.2	1

In the first phase, suppose that the first Expert takes only the first two criteria into account, the second discharges the first criterion, while the third Expert considers the first and the fourth criteria only. The values assigned by the three Experts directly assigned or implicitly obtained using a suitable questionnaire as proposed by [3] to every coalition are reported in Tables from 2 up to 4. The first Expert considers more important the first criterion with respect to the third one, but the coalition formed by both the two Experts receives a weight greater than the sum of the two criteria weights. Thus the first Expert exhibits a *synergic*, or *disjunctive*, effect. In MCDA literature, this is sometimes knew as *andness*-type effect, while the opposite redundant behavior is named *orness*-type effect, see [7]. The second Expert is characterized by a *linear* behavior, as it can be easily checked. Both the first and the second Expert evaluate monotonically. But this is not true for the last Expert, which considers the first criterion more important than the fourth, but assigns to the coalition formed by the two criteria an inferior weight than the minimum of them. While the first Expert exhibits a *disjunctive* behavior (*orness*-type), and the second is *linear*, we can say that the last Expert is characterized by an *exclusive-orness* behavior. In fact, his preference structure follows a rule like:

The alternative is highly scored if the first or if the fourth criterion is high, but NOT both of them

The scoring of each alternatives can be obtained aggregating using the (single step) Choquet integral for each Expert. The Table 5 reports the results of the aggregation for the three Experts. For instance, the score of the first alternative calculated for the first Expert, 0.28, is obtained as follows, see (1):

$$\mathbf{x}_1^1 = 0.2 \times 1 + (0.4 - 0.2) \times 0.4 = 0.28 \quad (4)$$

In the subsequent phase, the individual Expert's scores need to be aggregated into a final one. To this purpose, suppose that, after a preliminary briefing, the implicit preference struc-

Table 3: Expert n. 2

$m^2(2)$	$m^2(3)$	$m^2(4)$
0.5	0.2	0.3

ISBN: 978-989-95079-6-8

$m^2(2,3)$	$m^2(2,4)$	$m^2(3,4)$	$m^2(2,3,4)$
0.7	0.8	0.5	1

Table 4: Expert n. 3

$m^3(1)$	$m^3(4)$	$m^3(1,4)$
0.7	0.6	0.2

ture of the DM about the Expert's confidence/experience be heuristically expressed by the following rules:

Rule # 1: *If the Experts 1 and 3 agree, I have complete confidence about their choice (in this case, the opinion of the second Expert is inessential).*

Rule # 2: *The coalition formed by Experts 1 and 2 is more reliable than the coalition formed by Experts 2 and 3.*

Rule # 3: *Considering each Experts by alone, the first one is the most important, while the second one is the least.*

These rules can be translated in the coalition weights reported in Table 6.

The individual measures can now be aggregated, obtaining the following aggregated alternative scores for the two considered alternatives:

$$\sigma_1 = 0.648, \sigma_2 = 0.344$$

The score σ_1 is computed as follows:

$$\sigma_1 = 0.2\times 1+(0.28-0.2)\times 0.5+(0.44-0.28)\times 0.3 = 0.648 \quad (5)$$

While for the first and for the second Experts the Shapley and the *extended* Shapley indices coincide (they are monotone), for the third Expert the Shapley index and the *extended* Shapley index are respectively, for the first criterion:

$$p_3 = 0.5 \times \{0.7 + (0.2 - 0.6)\} = 0.15 \quad (6)$$

$$q_3 = 0.5 \times \{0.7 + |0.2 - 0.6|\} = 0.55 \quad (7)$$

Observe that the Shapley index differs from the *extended* Shapley indices. In particular, the extended Shapley index is significatively high (0.55), while the Shapley index is low (0.15). We can conclude that this criterion of the third Expert is *important* even if it is, on average, neither a *benefit*, nor a *cost*.

In a future work, we intend to analyze the properties and the relationships between the Shapley and the *extended* Shapley for the two-step Choquet integral.

Table 5: Alternative scoring of each Expert

Alternatives	Expert n. 1	Expert n. 2	Expert n. 3
1	0.28	0.44	0.2
2	0.22	0.54	0.26

Table 6: Measure of Expert coalitions

u_1	u_2	u_3	$u_{1,2}$	$u_{1,3}$	$u_{2,3}$
0.2	0.3	0.7	0.5	0.7	1

References

[1] Cardin M., Giove S., Aggregation functions with non-monotonic measures, *Fuzzy Economic review*, 13, 2, 3-15, 2008.

[2] Ceberio M., Modave F., An interval valued, 2-additive Choquet Integral for Multi-criteria Decision Making, *10th Conf. on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU)*, Perugia, Italy, 2004.

[3] Despic O., Simonovic S.P., Aggregation operators for soft decision making in water resources, *Fuzzy Sets and Systems*, 115, 11-33, 2000.

[4] Bordogna, G., Fedrizzi, M., Pasi, G., A Linguistic Modeling of Consensus for a Fuzzy Majority in Group Decision Making, *IEEE Transactions on Systems, Man and Cybernetics* 27, 1, 126-132, 1997.

[5] De Waegenaere A., Wakker P.P., Nonmonotonic Choquet integrals, *Journal of Mathematical Economics*, 36, 45-60, 2001.

[6] Grabish M., Marichal J.L., Roubens M., Equivalent representations of a set function with applications to game theory and multicriteria decision making. *Mathematics of Operations Research*, vol. 25, 2, 157-178, 2000.

[7] Grabisch M., Roubens M., *Application of the Choquet integral in multicriteria decision making*, in Fuzzy Measures and Integrals - Theory and Applications, M. Grabisch, T. Murofushi, and M. Sugeno (eds), Physica Verlag, 415-434, 2001.

[8] Kajii A., Kojima H., Takashi U., Cominimum additive operators, *Journal of Mathematical Economics*, 43, 218-230, 2007.

[9] Marichal J.L., An axiomatic approachof the discrete Choquet integral as a tool to aggregate interacting criteria *IEEE Transactions on Fuzzy Syssyems*, vol. 8, 6, 800-807, 2000.

[10] Meyer P., Roubens M., On the use of the Choquet integral with fuzzy numbers in multiple criteria decision support, *Fuzzy Sets and Systems*, 157, 927-938, 2006.

[11] Mesiar R., Vivona D., (1999), Two-step integral with respect to fuzzy measure, *Tatra Mountains Mathematical Pubblications*, 16, 359-368, 2006.

ISBN: 978-989-95079-6-8

[12] Murofushi T., Sugeno M., Machida M., Non-monotonic fuzzy measures and Choquet integral, *Fuzzy Sets and Systems*, 64, 73-86, 1994.

[13] Narukawa Y., Murofushi T., The n-step Choquet integral on finite spaces. In *Proc. of the 9th Int. Conf. on Information Proc. and Management of Uncertainty in Knowledge-Based Systems*, Annecy, France, 2002.

[14] Rébillé Y., Sequentially continuous non-monotonic Choquet integrals, *Fuzzy Sets and Systems*, 153, 79-94, 2005.

ISBN: 978-989-95079-6-8

A desktop calculator for parametric fuzzy arithmetic

Laerte Sorini

DEMQ (Department of Economics and Quantitative Methods)
University of Urbino 'Carlo Bo', Italy.
Email: laerte@uniurb.it

Abstract— *In this paper we describe a desktop calculator for the fuzzy arithmetic and the fuzzy extension of basic univariate functions. The fuzzy numbers are modelled indifferently in the parametric LR or LU representations. The operations are performed in the setting of Zadeh's extension principle.*

Keywords— LR-Fuzzy Numbers, LU-Fuzzy Parametrization, Fuzzy Arithmetic.

1 Introduction

About three decades ago, Dubois and Prade developed the arithmetical structure of fuzzy numbers and they introduced the well known LR model and the corresponding formulas for the fuzzy operations (see the recent publication [1] and the references therein).

The arithmetic calculations with fuzzy numbers, according to Zadeh's Extension Principle) can be performed using two general settings:

a. the well known LR fuzzy numbers, for which the operations are performed by calculating the membership function $x \rightarrow \mu(x)$ of the result from the membership functions of the operands (as in [1]).

b. the LU representation of fuzzy numbers, developed in [2], [4] and extensively described in [5], where the operations are performed using the $\alpha-cuts$ representation $\alpha \rightarrow u_\alpha^-, \alpha \rightarrow u_\alpha^+$ of the operands and on the basis of interval analysis for each cut $[u_\alpha^-, u_\alpha^+]$.

It is well known that, for the case of fuzzy numbers (but not for general fuzzy sets), the two representations are equivalent as it is possible to go from LR to LU and from LU to LR by inverting the (x, μ) or the (α, u) axes.

Following the theoretical results on LR and LU fuzzy arithmetic operations, we have developed a desktop calculator which performs the basic operations and the extension of elementary (unidimensional) functions using indifferently one of the two settings.

In this paper we describe the structure of the calculator and how it works.

2 LU and LR fuzzy numbers and arithmetic

Fuzzy numbers are fuzzy sets defined over the real numbers $\mathbb{R}$, having membership functions $(x, \mu_u(x))$ for each $x \in \mathbb{R}$, in the form

$$\mu_u(x) = \begin{cases} L(\frac{b-x}{b-a}) & \text{if } x \in [a,b] \\ 1 & \text{if } x \in [b,c] \\ R(\frac{x-c}{c-d}) & \text{if } x \in [c,d] \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where $L, R : [0,1] \longrightarrow [0,1]$ are non-increasing with $R(0) = L(0) = 1$ and $R(1) = L(1) = 0$. The $\alpha - cuts$ are defined by

$$[u]_\alpha = \{x | x \in \mathbb{X},\ \mu_u(x) \geq \alpha\} \quad (2)$$

with

$$[u]_0 = cl\{x | x \in \mathbb{X},\ \mu_u(x) > 0\}.$$

We will denote them by

$$\begin{aligned} [u]_\alpha &= [u_\alpha^-, u_\alpha^+] \quad (3) \\ &\text{where} \\ u_\alpha^- &= b - (b-a)L^{-1}(\alpha) \\ u_\alpha^+ &= c + (d-c)R^{-1}(\alpha) \end{aligned}$$

It is well known that the two functions $\alpha \rightarrow u_\alpha^-$ (the Lower branch) and $\alpha \rightarrow u_\alpha^+$ (the Upper branch) are monotonic (respectively increasing and decreasing) functions for all $\alpha \in [0,1]$.

The LR or LU parametric representations use monotonic interpolation by shape functions $p : [0,1] \longrightarrow [0,1]$ such that $p(0) = 0$ and $p(1) = 1$ with $p(t)$ differentiable and increasing on $[0,1]$; with parameters $\beta_i \geq 0,\ i = 0,1$ we satisfy conditions

$$\begin{aligned} p(0) &= 0,\ p(1) = 1 \\ p'(0) &= \beta_0,\ p'(1) = \beta_1. \end{aligned}$$

An example is the following [4]:

$$p(t; \beta_0, \beta_1) = \frac{t^2 + \beta_0 t(1-t)}{1 + (\beta_0 + \beta_1 - 2)t(1-t)}; \quad (4)$$

Function p in (4) is increasing on [0,1] and is used as model for functions L and R; in fact, if $a \leq b \leq c \leq d$ and $\beta_{0,L}, \beta_{1,L} \geq 0$, $\beta_{0,R}, \beta_{1,R} \geq 0$ are given, an LR-fuzzy number has membership

$$\mu_u(x) = \begin{cases} p(\frac{b-x}{b-a}; \beta_{0,L}, \beta_{1,L}) & \text{if } x \in [a,b] \\ 1 & \text{if } x \in [b,c] \\ 1 - p(\frac{c-x}{c-d}; \beta_{0,R}, \beta_{1,R}) & \text{if } x \in [c,d] \\ 0 & \text{otherwise} \end{cases}.$$

We denote a, b, c, d as $a = u_{0,L}$, $b = u_{1,L}$, $c = u_{1,R}$, $d = u_{0,R}$ so that eight parameters define u:

$$u_{LR} = (u_{0,L}, \beta_{0,L}, u_{0,R}, \beta_{0,R}; u_{1,L}, \beta_{1,L}, u_{1,R}, \beta_{1,R}) \quad (5)$$

provided that $u_{0,L} \leq u_{1,L} \leq u_{1,R} \leq u_{0,R}$ and $\beta_{0,L}, \beta_{1,L} \geq 0$, $\beta_{0,R}, \beta_{1,R} \geq 0$.

ISBN: 978-989-95079-6-8

The LU representation is obtained if the model functions $p(t;\beta_0,\beta_1)$ are used to model the Lower and the Upper branches of the $\alpha-cuts$.

We can also switch the two representations: for example, for a given LR-fuzzy number $u\in\mathbb{F}^{LR}$ given by (5), its approximated LU representation $u\in\mathbb{F}^{LU}$ is

$$\left\{\begin{array}{c} u_{LU}=(u_0^-,\delta u_0^-,u_0^+,\delta u_0^+;\ u_1^-,\delta u_1^-,u_1^+,\delta u_1^+) \\ \text{with} \\ \begin{array}{l} u_0^-=u_{0,L}\,,\delta u_0^-=\frac{1}{\delta u_{0,L}} \\ u_1^-=u_{1,L}\,,\delta u_1^-=\frac{1}{\delta u_{1,L}} \\ u_0^+=u_{0,R}\,,\delta u_0^+=\frac{1}{\delta u_{0,R}} \\ u_1^+=u_{1,R}\,,\delta u_1^+=\frac{1}{\delta u_{1,R}} \end{array} \end{array}\right. \tag{6}$$

(if some $\delta u_{i,L},\delta u_{i,R}$ is zero, the corresponding infinite δu_i^-, δu_i^+ slope can be assigned a BIG number).

If the fuzzy numbers are given in the LR form, then the (LU)-(LR) fuzzy relationship (6) can be used as an intermediate step for LR-fuzzy operations.

Consider two LR-fuzzy numbers u and v ($N=1$ for simplicity)

$$\begin{array}{rcl} u_{LR} & = & (u_{0,L},\delta u_{0,L},u_{0,R},\delta u_{0,R};u_{1,L},\delta u_{1,L},u_{1,R},\delta u_{1,R}) \\ v_{LR} & = & (v_{0,L},\delta v_{0,L},v_{0,R},\delta v_{0,R};v_{1,L},\delta v_{1,L},v_{1,R},\delta v_{1,R}) \end{array} \tag{7}$$

having the LU representations

$$\begin{array}{rcl} u_{LU} & = & (u_0^-,\delta u_0^-,u_0^+,\delta u_0^+;u_1^-,\delta u_1^-,u_1^+,\delta u_1^+) \\ v_{LU} & = & (v_0^-,\delta v_0^-,v_0^+,\delta v_0^+;v_1^-,\delta v_1^-,v_1^+,\delta v_1^+) \end{array} \tag{8}$$

with $u_i^\pm$, $v_i^\pm$, $\delta u_i^\pm$ and $\delta v_i^\pm$ $(i=0,1)$ calculated according to (6).

The model functions above can be adopted not only to define globally the shapes, but also to represent the functions "piecewise", on a decomposition of the interval $[0,1]$ into N subintervals

$$0=\alpha_0<\alpha_1<.....<\alpha_{i-1}<\alpha_i<....<\alpha_N=1.$$

By the transformation $t_\alpha=\frac{\alpha-\alpha_{i-1}}{\alpha_i-\alpha_{i-1}}$, $\alpha\in I_i$, each subinterval I_i is mapped into the standard $[0,1]$ interval to determine each piece independently and obtain general left-continuous LU-fuzzy numbers. Globally continuous or more regular $C^{(1)}$ fuzzy numbers can be obtained directly from the data (for example, $u_{1,i}^-=u_{0,i+1}^-$, $u_{1,i}^+=u_{0,i+1}^+$ for continuity and $d_{1,i}^-=d_{0,i+1}^-$, $d_{1,i}^+=d_{0,i+1}^+$ for differentiability at $\alpha=\alpha_i$).

Let $p_i^\pm(t)$ denote the model function on I_i; we obtain easily

$$p_i^-(t)=p(t;\beta_{0,i}^-,\beta_{1,i}^-)\,,\ p_i^+(t)=1-p(t;\beta_{0,i}^+,\beta_{1,i}^+) \tag{9}$$

with

$$\begin{array}{rcl} \beta_{j,i}^- & = & \dfrac{\alpha_i-\alpha_{i-1}}{u_{1,i}^--u_{0,i}^-}d_{j,i}^- \\ \beta_{j,i}^+ & = & -\dfrac{\alpha_i-\alpha_{i-1}}{u_{1,i}^+-u_{0,i}^+}d_{j,i}^+ \text{ for } j=0,1 \end{array}$$

so that, for $\alpha\in[\alpha_{i-1},\alpha_i]$ and $i=1,2.,,,N$:

$$u_\alpha^- = u_{0,i}^-+(u_{1,i}^--u_{0,i}^-)p_i^-(t_\alpha) \tag{10}$$

$$t_\alpha = \frac{\alpha-\alpha_{i-1}}{\alpha_i-\alpha_{i-1}} \tag{11}$$

$$u_\alpha^+ = u_{0,i}^++(u_{1,i}^+-u_{0,i}^+)p_i^+(t_\alpha) \tag{12}$$

$$t_\alpha = \frac{\alpha-\alpha_{i-1}}{\alpha_i-\alpha_{i-1}}. \tag{13}$$

So, a differentiable shape function requires $4(N+1)$ parameters

$$\begin{array}{rcl} u & = & (\alpha_i;u_i^-,\delta u_i^-,u_i^+,\delta u_i^+)_{i=0,1,...,N} \\ u_0^- & \le & u_1^-\le...\le u_N^-\le u_N^+\le u_{N-1}^+\le...\le u_0^+ \\ \delta u_i^- & \ge & 0,\delta u_i^+\le 0. \end{array} \tag{14}$$

and the branches are computed according to (10) and (12). An example with $N=4$ is in **Table 1.**

Table 1. LU parametrization of a fuzzy number

α_i	u_i^-	δu_i^-	u_i^+	δu_i^+
0.0	-2.0	5.0	2.0	-0.5
0.5	-1.0	1.5	1.2	-2.0
1.0	0.0	2.5	0.0	-0.1

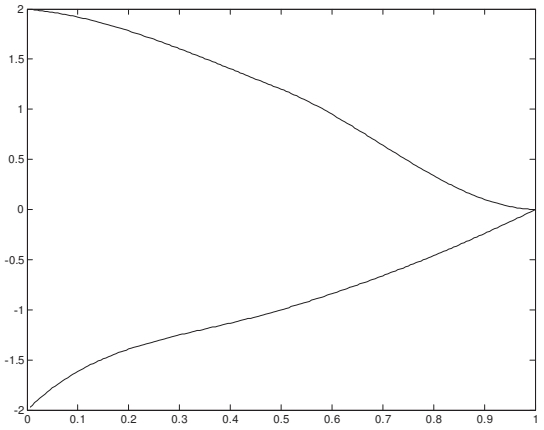


Figure 1. A fuzzy number in LU representation; the parameters are reported in Table 1 and the construction is obtained by the mixed spline with $N=2$.

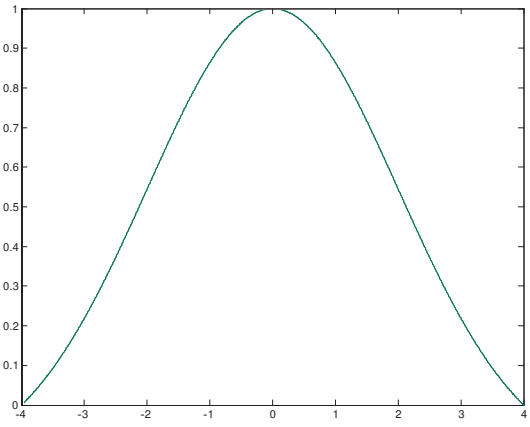


Figure 2. A quasi Gaussian fuzzy number; the parameters are reported in Table 2 and the membership function is obtained by the mixed spline with $N=4$.

For the LR parametrization on a decomposition with N subintervals, we proceed in a similar way; for example, a fuzzy quasi Gaussian number with the following membership

ISBN: 978-989-95079-6-8

function

$$\mu(x)=\begin{cases}\frac{\exp(-\frac{(x-m)^2}{2\sigma^2})-\exp(-\frac{k^2}{2})}{1-\exp(-\frac{k^2}{2})} & \text{if } m-k\sigma\le x\le m+k\sigma\\ 0 & \text{otherwise.}\end{cases} \tag{15}$$

for $m=0,\sigma=2,k=2$ and approximated with $N=4$ (five points) is in **Table 2**:

Table 2. LR parametrization of fuzzy number (15)

α_i	$u_{i,L}$	$\delta u_{i,L}$	$u_{i,R}$	$\delta u_{i,R}$
0.0	-4.0	0.156518	4.0	-0.156518
0.25	-2.8921	0.293924	2.8921	-0.293924
0.5	-2.1283	0.349320	2.1283	-0.349320
0.75	-1.3959	0.316346	1.3959	-0.316346
1.0	0.0	0.0	0.0	0.0

and a hyperbolic tangent fuzzy number with the following membership function:

$$\mu(x)=\begin{cases}\frac{\tanh(-k^2)-\tanh(-\frac{(x-m)^2}{\sigma^2})}{\tanh(-k^2)}\\ \text{if } m-k\sigma\le x\le m+k\sigma\\ 0,\ \text{otherwise}\end{cases}. \tag{16}$$

for $m=0,\sigma=3,k=1$ has the form indicated in **Table 3**:

Table 3.LR parametrization of fuzzy number (16)

α_i	$u_{i,L}$	$\delta u_{i,L}$	$u_{i,R}$	$\delta u_{i,R}$
0.0	-3.0	0.367627	3.0	-0.367627
0.25	-2.4174	0.475221	2.4174	-0.475221
0.5	-1.8997	0.473932	1.8997	-0.473932
0.75	-1.3171	0.370379	1.3171	-0.370379
1.0	0.0	0.0	0.0	0.0

The arithmetic operations associated to the LU representation are performed according to the extension principle and by representing the results by the same model used to represent the operands. The *addition* is defined by:

$$u+v=(u_i^-+v_i^-,\delta u_i^-+\delta v_i^-,u_i^++v_i^+,\delta u_i^++\delta v_i^+)_{i=0,1,\dots,N}.$$

The *scalar multiplication* is defined as follows:
if $k\ge 0$ then

$$ku=(ku_i^-,k\delta u_i^-,ku_i^+,k\delta u_i^+)_{i=0,1,\dots,N};$$

if $k<0$ then

$$ku=(ku_i^+,k\delta u_i^+,ku_i^-,k\delta u_i^-)_{i=0,1,\dots,N}.$$

In particular, if $k=-1$, we have

$$-u=(-u_i^+,-\delta u_i^+,-u_i^-,-\delta u_i^-)_{i=0,1,\dots,N}$$

and the *subtraction* is defined by

$$u-v=u+(-v).$$

We note explicitly that the scalar multiplication is always reproduced exactly in all the models for all $\alpha\in[0,1]$ but, in general, this is not true for the addition as the sum of rational or mixed functions is not always a rational or a mixed function of the same orders.

A particular situation arises for addition (or subtraction) if the mixed model is used. Suppose that the two branches to be added are given by the data $(u_0^{(1)},u_1^{(1)},\delta u_0^{(1)},\delta u_1^{(1)})$ and $(u_0^{(2)},u_1^{(2)},\delta u_0^{(2)},\delta u_1^{(2)})$; the mixed model is characterized by values of w (or a) for each data set

$$w^{(1)}=\frac{\delta u_0^{(1)}+\delta u_1^{(1)}}{u_1^{(1)}-u_0^{(1)}},\ w^{(2)}=\frac{\delta u_0^{(2)}+\delta u_1^{(2)}}{u_1^{(2)}-u_0^{(2)}}.$$

If $w^{(1)+(2)}$ is the w parameter for the addition, then

$$w^{(1)+(2)}=\frac{\delta u_0^{(1)}+\delta u_1^{(1)}+\delta u_0^{(2)}+\delta u_1^{(2)}}{u_1^{(1)}-u_0^{(1)}+u_1^{(2)}-u_0^{(2)}}$$

and it is easy to see that $w^{(1)+(2)}$ is a weighted average of $w^{(1)}$ and $w^{(2)}$ as

$$w^{(1)+(2)}\in\left[\min\left\{w^{(1)},w^{(2)}\right\},\max\left\{w^{(1)},w^{(2)}\right\}\right];$$

if $w^{(1)}=w^{(2)}$ then it follows that $w^{(1)+(2)}=w^{(1)}=w^{(2)}$. So, if the two fuzzy numbers to be added are modelled by a spline of the same *degree,* then the mixed model produces exact addition for all $\alpha\in[0,1]$.

This is true, in particular, for the fuzzy numbers having the slopes d_0 and d_1 not available, then we choose $d_0=n(1-\beta)(u_1-u_0)$ and $d_1=n\beta(u_1-u_0)$ for a fixed integer n and for a parameter $\beta\in[0,1]$. In this case, in fact, if

$$\begin{aligned}\delta u_0^{(k)} &= n(1-\beta^{(k)})(u_1^{(k)}-u_0^{(k)}),\ \delta u_1^{(k)}=\\ &= n\beta^{(k)}(u_1^{(k)}-u_0^{(k)}),\ k=1,2\end{aligned}$$

then

$$\begin{aligned}\delta u_0^{(1)}+\delta u_0^{(2)} &= n(1-\frac{\beta^{(1)}+\beta^{(2)}}{2})(u_1^{(1)}+u_1^{(2)}-u_0^{(1)}-u_0^{(2)})\\ \delta u_1^{(1)}+\delta u_1^{(2)} &= n\frac{\beta^{(1)}+\beta^{(2)}}{2}(u_1^{(1)}+u_1^{(2)}-u_0^{(1)}-u_0^{(2)})\end{aligned}$$

so that $\beta^{(1)+(2)}=\frac{\beta^{(1)}+\beta^{(2)}}{2}$.

For fuzzy *multiplication* we have an easy to implement algorithm, based on the applications of exact fuzzy multiplication at the nodes of the $\alpha-$subdivision; define

$$(uv)_i^-=\min\{u_i^-v_i^-,u_i^-v_i^+,u_i^+v_i^-,u_i^+v_i^+\} \tag{17}$$

$$(uv)_i^+=\max\{u_i^-v_i^-,u_i^-v_i^+,u_i^+v_i^-,u_i^+v_i^+\} \tag{18}$$

and set the following:

$$y=uv=\left(y_i^-,\delta y_i^-,y_i^+,\delta y_i^+\right)_{i=0,1,\dots,N}$$

To implement the multiplication we can proceed as follows: let (p_i^-,q_i^-) be the pair associated to the combination of superscripts $+$ and $-$ giving the minimum $(uv)_i^-$ in (17), and similarly let (p_i^+,q_i^+) the pair associated to the combination of $+$ and $-$ giving the maximum $(uv)_i^+$ in (18), then we obtain:

$$y_i^-=u_i^{p_i^-}v_i^{q_i^-}\ \text{ and }\ y_i^+=u_i^{p_i^+}v_i^{q_i^+}$$

$$\delta y_i^-=\delta u_i^{p_i^-}v_i^{q_i^-}+u_i^{p_i^-}\delta v_i^{q_i^-}\ \text{ and }\ \delta y_i^+=\delta u_i^{p_i^+}v_i^{q_i^+}+u_i^{p_i^+}\delta v_i^{q_i^+}$$

ISBN: 978-989-95079-6-8

where we use the product derivative rule to obtain the new slopes.

Analogous formulas can be deduced for *division:*

$$z = u/v = \left(z_i^-, \delta z_i^-, z_i^+, \delta z_i^+\right)_{i=0,1,...,N}$$

$$\begin{aligned}(u/v)_i^- &= \min\{u_i^-/v_i^-, u_i^-/v_i^+, u_i^+/v_i^-, u_i^+/v_i^+\} \text{ and}\\ (u/v)_i^+ &= \max\{u_i^-/v_i^-, u_i^-/v_i^+, u_i^+/v_i^-, u_i^+/v_i^+\}.\end{aligned}$$

Let (r_i^-, s_i^-) be the pair associated to the combination of $+$ and $-$ giving the minimum in $(u/v)_i^-$ and similarly let (r_i^+, s_i^+) be the pair associated to the combination of $+$ and $-$ giving the maximum in $(u/v)_i^+$, then it follows:

$$z_i^- = u_i^{r_i^-}/v_i^{s_i^-} \qquad z_i^+ = u_i^{r_i^+}/v_i^{s_i^+}$$

$$\begin{aligned}\delta z_i^- &= (\delta u_i^{r_i^-} v_i^{s_i^-} - u_i^{r_i^-}\delta v_i^{s_i^-})/(v_i^{s_i^-})^2\\ \delta z_i^+ &= (\delta u_i^{r_i^+} v_i^{s_i^+} - u_i^{r_i^+}\delta v_i^{s_i^+})/(v_i^{s_i^+})^2.\end{aligned}$$

As pointed out by the results of experimentation reported in [2] and [4] , the operations above are exact at the nodes α_i and have very small global errors on $[0,1]$. Further, it is easy to control the error by using a sufficiently high number of nodes with $\max\{\alpha_i - \alpha_{i-1}\}$ sufficiently small.

The general algorithms for the four arithmetical operations are now detailed.

Let $u = (u_i^-, \delta u_i^-, u_i^+, \delta u_i^+)_{i=0,1,...,N}$ and $v = (v_i^-, \delta v_i^-, v_i^+, \delta v_i^+)_{i=0,1,...,N}$ be given; in order to calculate the LU *addition* $w = u + v$ and the LU *subtraction,* $z = u - v$ with $w = (w_i^-, \delta w_i^-, w_i^+, \delta w_i^+)_{i=0,1,...,N}$ and $z = (z_i^-, \delta z_i^-, z_i^+, \delta z_i^+)_{i=0,1,...,N}$, the following sequence of iterations has to be improved:

for $i = 0, 1, ..., N$
$w_i^- = u_i^- + v_i^-,\ z_i^- = u_i^- - v_i^+$
$\delta w_i^- = \delta u_i^- + \delta v_i^-,\ \delta z_i^- = \delta u_i^- - \delta v_i^+$
$w_i^+ = u_i^+ + v_i^+,\ z_i^+ = u_i^+ - v_i^-$
$\delta w_i^+ = \delta u_i^+ + \delta v_i^+,\ \delta z_i^+ = \delta u_i^+ - \delta v_i^-$
end
test if conditions (14) are satisfied
for $(y_i^-, \delta y_i^-, y_i^+, \delta y_i^+)_{i=0,1,...,N}$

Let $k \in \mathbb{R}$ and $u = (u_i^-, \delta u_i^-, u_i^+, \delta u_i^+)_{i=0,1,...,N}$ be given; the computation of the LU *scalar multiplication* $w = ku$ with $w = (w_i^-, \delta w_i^-, w_i^+, \delta w_i^+)_{i=0,1,...,N}$ is obtained with the following iterations:

for $i = 0, 1, ..., N$
if $k \geq 0$ **then** $w_i^- = ku_i^-,\ \delta w_i^- = k\delta u_i^-$
$w_i^+ = ku_i^+\ ,\ \delta w_i^+ = k\delta u_i^+$
else $w_i^- = ku_i^+,\ \delta w_i^- = k\delta u_i^+$
$w_i^+ = ku_i^-\ ,\ \delta w_i^+ = k\delta u_i^-$
end

Finally, if $u = (u_i^-, \delta u_i^-, u_i^+, \delta u_i^+)_{i=0,1,...,N}$ and $v = (v_i^-, \delta v_i^-, v_i^+, \delta v_i^+)_{i=0,1,...,N}$ are given then the LU multiplication $w = uv$, with $w = (w_i^-, \delta w_i^-, w_i^+, \delta w_i^+)_{i=0,1,...,N}$, is deduced from the following algorithm:

for $i = 0, 1, ..., N$
$m_i = \min\{u_i^- v_i^-, u_i^- v_i^+, u_i^+ v_i^-, u_i^+ v_i^+\}$
$M_i = \max\{u_i^- v_i^-, u_i^- v_i^+, u_i^+ v_i^-, u_i^+ v_i^+\}$
$w_i^- = m_i\ ,\ w_i^+ = M_i$
if $u_i^- v_i^- = m_i$ **then** $\delta w_i^- = \delta u_i^- v_i^- + u_i^- \delta v_i^-$
elseif $u_i^- v_i^+ = m_i$ **then** $\delta w_i^- = \delta u_i^- v_i^+ + u_i^- \delta v_i^+$
elseif $u_i^+ v_i^- = m_i$ **then** $\delta w_i^- = \delta u_i^+ v_i^- + u_i^+ \delta v_i^-$
elseif $u_i^+ v_i^+ = m_i$ **then** $\delta w_i^- = \delta u_i^+ v_i^+ + u_i^+ \delta v_i^+$
endif
if $u_i^- v_i^- = M_i$ **then** $\delta w_i^+ = \delta u_i^- v_i^- + u_i^- \delta v_i^-$
elseif $u_i^- v_i^+ = M_i$ **then** $\delta w_i^+ = \delta u_i^- v_i^+ + u_i^- \delta v_i^+$
elseif $u_i^+ v_i^- = M_i$ **then** $\delta w_i^+ = \delta u_i^+ v_i^- + u_i^+ \delta v_i^-$
elseif $u_i^+ v_i^+ = M_i$ **then** $\delta w_i^+ = \delta u_i^+ v_i^+ + u_i^+ \delta v_i^+$
endif
end

A similar algorithm can be deduced for the division.

3 Computation of unidimensional fuzzy-valued functions with the calculator

We consider first a single variable differentiable function $f : \mathbb{R} \to \mathbb{R}$; its *(EP)*-extension $v = f(u)$ to a fuzzy argument $u = (u^-, u^+)$ has $\alpha - cuts$

$$[v]_\alpha = \left[\min\left\{f(x) \mid x \in [u]_\alpha\right\}, \max\left\{f(x) \mid x \in [u]_\alpha\right\}\right]. \tag{19}$$

If f is monotonic increasing we obtain $[v]_\alpha = [f(u_\alpha^-), f(u_\alpha^+)]$ while, if f is monotonic decreasing, $[v]_\alpha = [f(u_\alpha^+), f(u_\alpha^-)]$; the LU representation of $v = (v_i^-, \delta v_i^-, v_i^+, \delta v_i^+)_{i=0,1,...,N}$ can be obtained.

In the non monotonic (differentiable) case, we have to solve the optimization problems in (19) for each $\alpha = \alpha_i$, $i = 0, 1, ..., N$, i.e.

$$(EP_i)\text{:} \begin{cases} v_i^- = \min\left\{f(x) | x \in [u_i^-, u_i^+]\right\} \\ v_i^+ = \max\left\{f(x) | x \in [u_i^-, u_i^+]\right\}. \end{cases}$$

The $\min$ (or the $\max$) can occur either at a point which is coincident with one of the extremal values of $[u_i^-, u_i^+]$ or at a point which is internal; in the last case, the derivative of f is null and $\delta v_i^- = 0$ (or $\delta v_i^+ = 0$).

To implement the LU-fuzzy calculator, we have written a windows-based frame similar to a standard hand-calculator.

Figure 3 shows a complete view of the calculator; from left to right we can see the grids of the fuzzy numbers X, Y and Z. Z is the result of the operations while X and/or Y are the operands. For each element $u \in \{X, Y, Z\}$the grid contains the values α_i, u_i^-, δu_i^-, u_i^+ and δu_i^+ respectively in the LU View Mode or x, $\mu(x)$, $\delta\mu(x)$ in the LR Mode. To start the calculations, we have implemented a set of predefined types, including triangular, trapezoidal, exponential, gamma, etc. For a given type, it is possible to define the number N of subintervals ($N+1$ points) in the uniform $\alpha-$decomposition.

ISBN: 978-989-95079-6-8

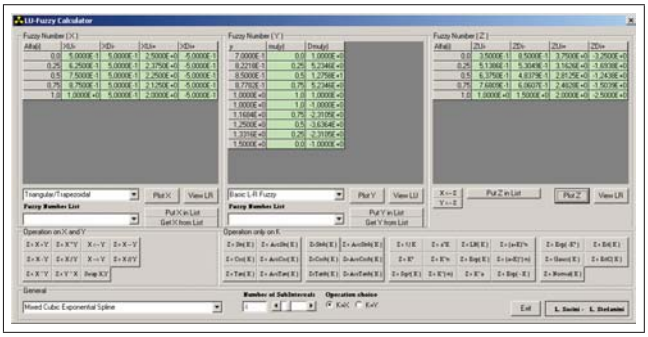


Figure 3. General window of the LU-fuzzy calculator.

The calculations are performed by clicking the button of the corresponding operation. The left group of buttons involves the binary operations (see Figure 4)

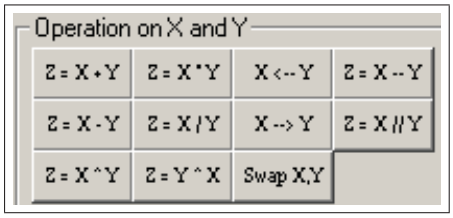


Figure 4. Binary operations and assignments.

The second group of operators (see Figure 5,6,7 and 8.) require the assignment of either X or Y to the temporary K and operate on K itself putting the result into Z.



Figure 5. Extension of univariate functions.

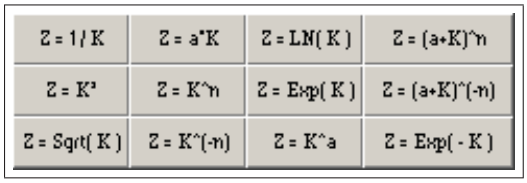


Figure 6. Extension of other univariate functions.

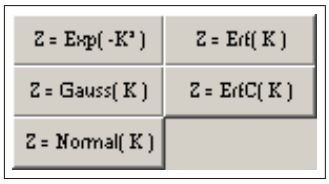


Figure 7. Extension of some other univariate functions.

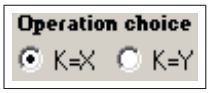


Figure 8. Selection of the argument for fuzzy extension functions.

It is possible to save a given (X, Y or Z) temporary result into a stored list (Put in List button), by assigning a name to it; a saved fuzzy number can be reloaded either in X or Y for further use (Get from List button). The Plot button (see Figure 9.) opens a popup window with the graph of the membership function of the corresponding fuzzy number and is possible select a fuzzy number from a list of predefined types; (View LR,View LU button) allow to switch between LR or LU fuzzy representation.



Figure 9. Fuzzy Number Control Panel

To obtain the graphs or other representations, one of the models can be selected (Figure 10).

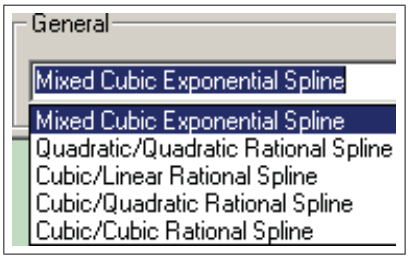


Figure 10. Choosing the monotonic spline model

3.1 *Fuzzy extension of* $(X, Y) \to X^Y$

We illustrate an example to show how the calculator works. We use the definition

$$X^Y = \exp\left(Y \ln\left((X)\right)\right)$$

and we compute

$$Z = X^Y$$

(where the input are the positive fuzzy X and the fuzzy Y), by the sequence of operations:

$$\begin{array}{lll} (i)\text{ (natural logarithm)} & Z & \longleftarrow \quad \ln(X) \\ (ii)\text{ (standard multiplication)} & Z & \longleftarrow \quad YZ \\ (iii)\text{ (exponential)} & Z & \longleftarrow \quad \exp(Z) \end{array}$$

The steps to follow in the calculator are the following.

1. First (see Figure 11) select a trapezoidal fuzzy number here we have 4 sub intervals and 5 $\alpha - cut$.

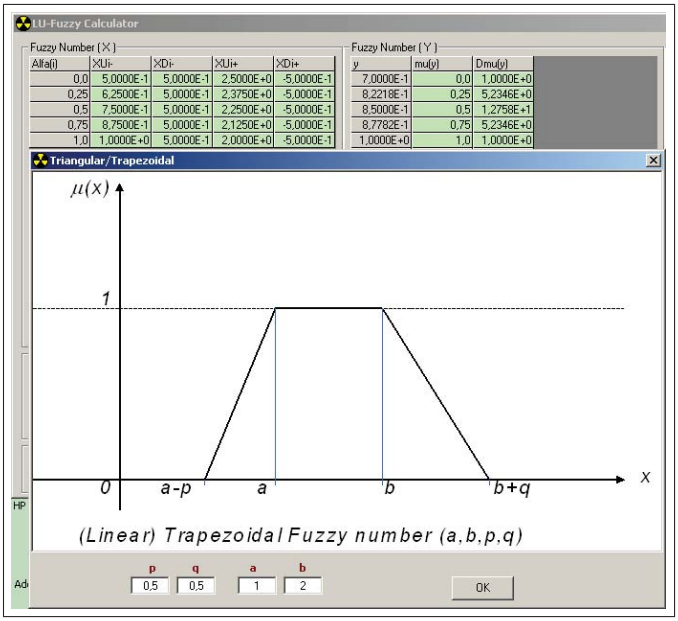


Figure 11. Example for trapezoidal fuzzy number loaded in X

If the selection is loaded into the X-area, the corresponding grid appears as in Figure 12 below and it is possible plot and switch in LR View mode immediately.

ISBN: 978-989-95079-6-8

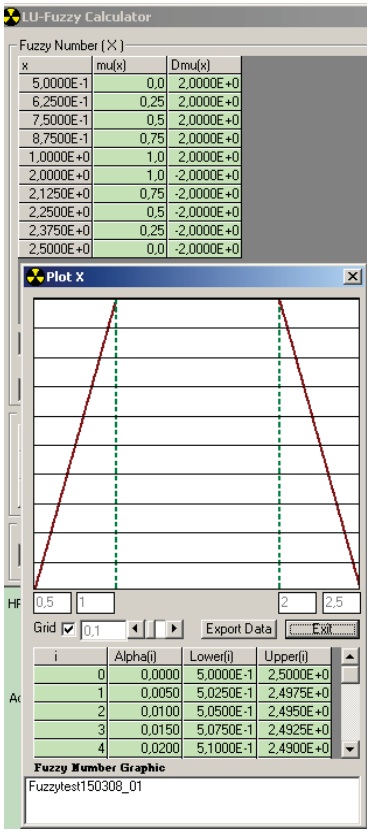


Figure 12. Assigne and plot the trapezoidal fuzzy number to X.

2. A second fuzzy number is loaded into Y and the button corresponding to the operation Z=X^Y is activated. The Z-grid is calculated by the rules of the LU-fuzzy calculus (see Figure 13).

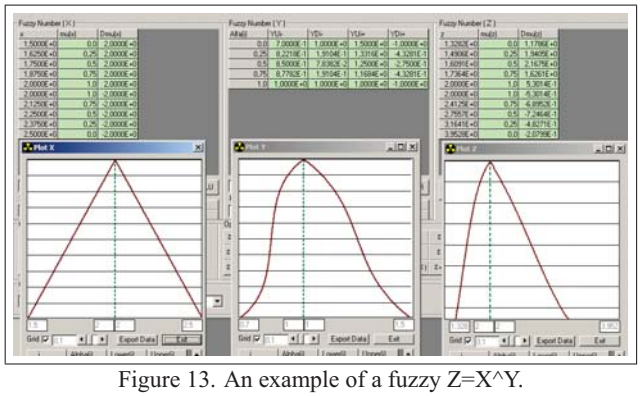


Figure 13. An example of a fuzzy Z=X^Y.

3. To see the graphical representation of X, Y and/or Z, click the corresponding Plot button and the popup windows appear; note that the fuzzy numbers X and Z are represented in LR mode instead Y is in LU mode.

4. Now, we save the result Z of the previous operation and we call it eusflat_Z (figure 14.).

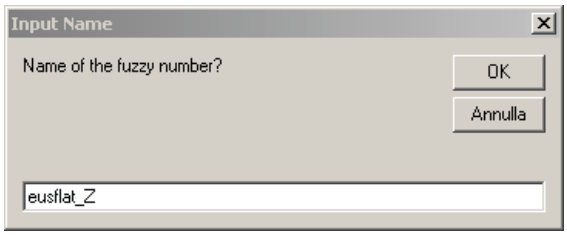


Figure 14. Choose the name for the result.

5. Now we load the saved Fuzzy_Z into the X-area, by getting it from the list of saved elements and with the possibility of reiterating.

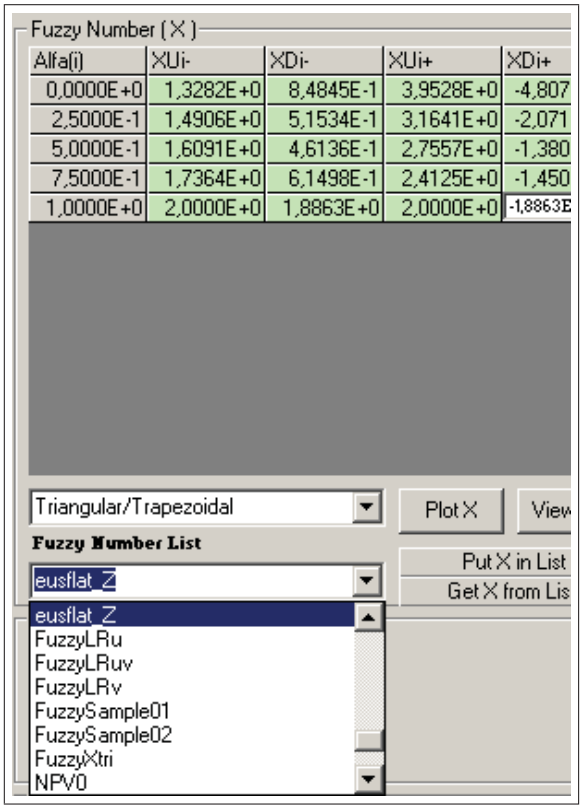


Figure 15. Loading of a saved intermediate fuzzy number into X area.

We believe that the suggested desktop calculator might be of interest in many areas where fuzzy numbers are applied, especially in computer decision support systems. In fact, it enables the interpretation of shapes as inputs and as outputs and the use of LR or LU parametric representations.

References

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

[2] M. L. Guerra, L. Stefanini, Approximate Fuzzy Arithmetic Operations Using Monotonic Interpolations, *Fuzzy Sets and Systems*, 150, 2005, 5-33.

[3] L. Stefanini, L. Sorini, An LU-fuzzy calculator for the basic fuzzy calculus, Working Papers 0701, University of Urbino Carlo Bo, Department of Economics, 2007.

[4] L. Stefanini, L. Sorini, M. L. Guerra, Parametric representation of fuzzy numbers and application to fuzzy calculus, *Fuzzy Sets and Systems*, 157, 2006, 2423-2455.

[5] L. Stefanini, L. Sorini, M.L. Guerra, Fuzzy Numbers and Fuzzy Arithmetic, in W. Pedrycz et al. (Eds) "Handbook of Granular Computing" , Chapter 12, J. Wiley & Sons, 2008.

ISBN: 978-989-95079-6-8

Discussions on Interpretability of Fuzzy Systems using Simple Examples

Hisao Ishibuchi, Yusuke Nojima

Department of Computer Science and Intelligent Systems, Osaka Prefecture University
1-1 Gakuen-cho, Naka-ku, Sakai, Osaka, 599-8531 Japan
Email: hisaoi@cs.osakafu-u.ac.jp, nojima@cs.osakafu-u.ac.jp

Abstract—*Two conflicting goals are often involved in the design of fuzzy rule-based systems: Accuracy maximization and interpretability maximization. A number of approaches have been proposed for finding a fuzzy rule-based system with a good accuracy-interpretability tradeoff. Formulation of the accuracy maximization is usually straightforward in each application area of fuzzy rule-based systems such as classification, regression and forecasting. Formulation of the interpretability maximization, however, is not so easy. This is because various aspects of fuzzy rule-based systems are related to their interpretability. Moreover, user's preference should be taken into account when a single fuzzy rule-based system is to be chosen from several alternatives with different accuracy-interpretability tradeoffs. In this paper, we discuss the difficulty in measuring the interpretability of fuzzy rule-based systems using very simple examples. We do not intend to propose any new interpretability measure. Our intention is to help to activate discussions on how to measure the interpretability of fuzzy rule-based systems.*

Keywords—Fuzzy systems, fuzzy rules, accuracy-interpretability tradeoff, multiobjective design of fuzzy systems.

1 Introduction

Handling of the tradeoff between the accuracy maximization and the interpretability maximization has been a hot issue in the design of fuzzy rule-based systems since the mid-1990s [1]-[3]. A number of approaches have already been proposed for improving the accuracy of fuzzy rule-based systems while maintaining their interpretability [1], [2], [4]-[21]. Genetic algorithms have been frequently used in those approaches to search for an accurate and interpretable fuzzy rule-based system. This is because genetic algorithms can perform not only continuous optimization for parameter tuning but also discrete optimization for structure determination. Studies on fuzzy genetics-based machine learning are called genetic fuzzy systems [22]-[24]. In some recent studies [3], [25]-[36], multi-objective genetic algorithms have been used to search for multiple Pareto-optimal fuzzy rule-based systems along the accuracy-interpretability tradeoff surface. Those studies are often referred to as multi-objective genetic fuzzy systems [37]. Recently multi-objective genetic algorithms have also been used for machine learning [38] and data mining [39].

Let us denote a fuzzy rule-based system by S. We can also view S as a set of fuzzy if-then rules. In each application area of fuzzy rule-based systems such as classification, regression and forecasting, the specification of the accuracy of S for the given training data is not difficult (e.g., the number of correctly classified training patterns by S). Let us denote the accuracy measure of S as *Accuracy*(S). A design problem of fuzzy rule-based systems can be formulated as follows:

$$\text{Maximize } \mathit{Accuracy}(S). \quad (1)$$

Due to the accuracy-interpretability tradeoff relation, the accuracy maximization in (1) often leads to the deterioration in the interpretability of fuzzy rule-based systems. This means that we often obtain from (1) an accurate and complicated fuzzy rule-based system with poor interpretability.

In some application areas, not only the accuracy but also the interpretability is very important. Thus we often want to maximize the accuracy of fuzzy rule-based systems without degrading their interpretability. This maximization problem can be formulated as follows:

$$\text{Maximize } \mathit{Accuracy}(S) \text{ subject to } \mathit{Interpretability}(S) \geq \alpha, \quad (2)$$

where *Interpretability*(S) is the interpretability measure of the fuzzy rule-based system S and α is the required minimum level of the interpretability.

Of course, we can formulate the maximization problem of the interpretability under the given minimum accuracy level β as follows:

$$\text{Maximize } \mathit{Interpretability}(S) \text{ subject to } \mathit{Accuracy}(S) \geq \beta. \quad (3)$$

One may want to maximize both the accuracy and the interpretability. In this case, a simple approach is to use a scalarizing function $f(\cdot)$ which combines the accuracy and interpretability measures into a single objective function:

$$\text{Maximize } f(\mathit{Accuracy}(S), \mathit{Interpretability}(S)). \quad (4)$$

A well-known scalarizing function is the weighted sum:

$$\text{Maximize } w_1 \mathit{Accuracy}(S) + w_2 \mathit{Interpretability}(S), \quad (5)$$

where w_1 and w_2 are non-negative weight values. In addition to the weighted sum in (5), we can use various scalarizing functions developed in the field of multiple criteria decision making (MCDM [40]-[42]).

In general, it is not easy for human users to specify an appropriate scalarizing function for multi-objective problems. Users may want to examine some fuzzy rule-based systems with different accuracy-interpretability tradeoffs (instead of a single best solution with respect to a specific scalarizing function). In this case, the design of fuzzy rule-based systems can be formulated as the following multi-objective problem:

$$\text{Maximize } \{\mathit{Accuracy}(S), \mathit{Interpretability}(S)\}. \quad (6)$$

A large number of Pareto-optimal fuzzy rule-based systems can be obtained by multi-objective genetic algorithms such as NSGA-II [43], SPEA [44] and SPEA2 [45].

ISBN: 978-989-95079-6-8

In many cases, the interpretability maximization is handled as the complexity minimization. Thus the above-mentioned formulations in (2)-(6) can be reformulated accordingly. For example, the multi-objective formulation in (6) is rewritten as

Maximize *Accuracy*(*S*) and minimize *Complexity*(*S*), (7)

where *Complexity*(*S*) is a complexity measure.

The main difficulty in the above-mentioned formulations in (2)-(7) for the design of accurate and interpretable fuzzy rule-based systems is the formulation of their interpretability. Whereas the formulation of the accuracy of fuzzy rule-based systems is usually straightforward from their application task such as classification and regression, it is not easy for human users to appropriately formulate the interpretability. This is because various aspects of fuzzy rule-based systems are related to their interpretability [46]-[52]. Moreover, it is not easy for human users to mathematically formulate each of those aspects even when the close relation of each aspect to the interpretability of fuzzy rule-based systems is clear.

In this paper, we explain the difficulty in formulating the interpretability of fuzzy rule-based systems using simple numerical examples. More specifically, we demonstrate the difficulty in comparing different fuzzy rule-based systems with respect to their interpretability even in very simple situations.

2 Interpretability of Fuzzy Partitions

When we use the same type of fuzzy partitions with different granularities (e.g., uniform fuzzy partitions with symmetric triangular membership functions (MFs) in Fig. 1), we can say that the increase in the number of membership functions degrades the interpretability of fuzzy partitions. For example, the fuzzy partition with two membership functions in Fig. 1 (a) is the most interpretable among the four alternatives in Fig. 1. In this case, we can formulate the interpretability of fuzzy partitions by the number of membership functions. That is, the interpretability maximization is realized by minimizing the number of membership functions in fuzzy partitions.

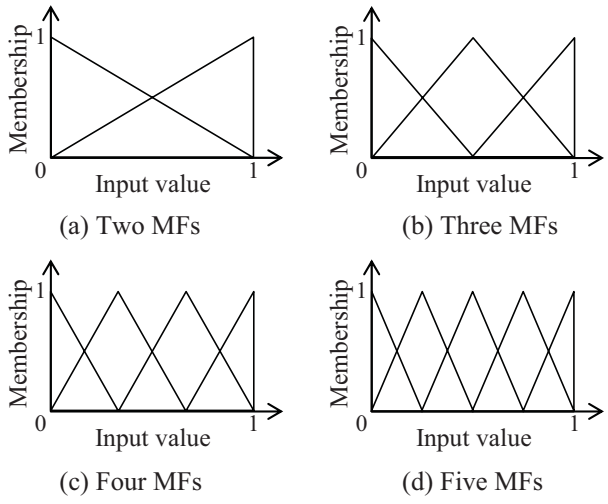


Figure 1: Fuzzy partitions with different granularities.

We can also use the number of membership functions as an interpretability measure (i.e., as a complexity measure to be minimized) when the same fuzzy partition is used for all input variables. For example, we can say that the 3x3 fuzzy grid in Fig. 2 (a) is more interpretable than the 4x4 fuzzy grid in Fig. 2 (b). The comparison, however, becomes difficult when we use different fuzzy partitions for each input variable as shown in the following two examples.

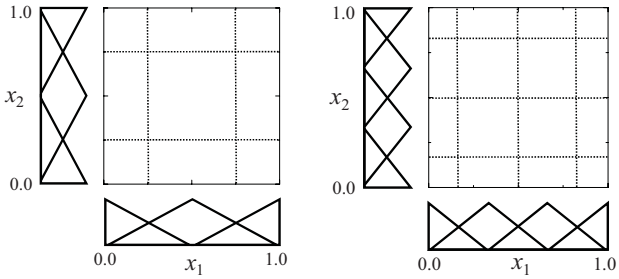


Figure 2: Comparison between the 3x3 and 4x4 fuzzy grids.

Example 1: Let us consider the 4x4 and 3x5 fuzzy grids in Fig. 3. Both fuzzy grids have eight membership functions in total (i.e., 4+4 = 3+5). Thus they are evaluated as having the same interpretability if they are compared using the number of membership functions. The 3x5 fuzzy grid is, however, viewed as being more interpretable than the 4x4 fuzzy grid if we use the number of fuzzy subspaces as an interpretability measure (i.e., 15 < 16). Since a single fuzzy rule is usually generated for each fuzzy subspace, the 3x5 fuzzy grid in Fig. 3 (b) can be viewed as being more interpretable than the 4x4 fuzzy grid in Fig. 3 (a) if we evaluate the interpretability using the number of fuzzy rules. Some human users, however, may intuitively feel that the 4x4 fuzzy grid with the same fuzzy partition for the two input variables is more interpretable than the 3x5 fuzzy grid with the different fuzzy partitions.

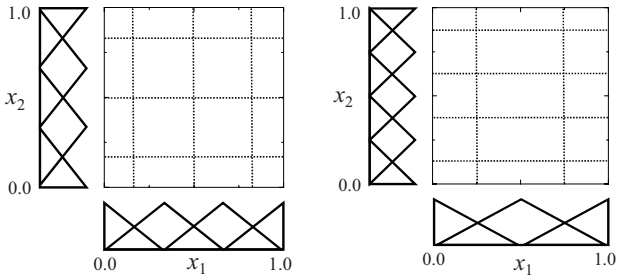
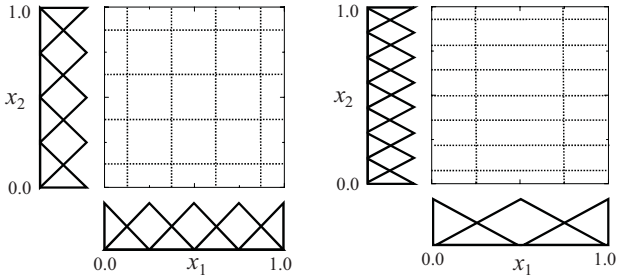


Figure 3: Example 1 with the 4x4 and 3x5 fuzzy grids.

Example 2: Let us consider the 5x5 and 3x8 fuzzy grids in Fig. 4. The 5x5 fuzzy grid in Fig. 4 (a) has less membership functions than the 3x8 fuzzy grid in Fig. 4 (b): 5+5 < 3+8. Thus the 5x5 fuzzy grid is evaluated as more interpretable than the 3x8 fuzzy grid in Fig. 4 (b) if we use the number of membership functions as an interpretability measure. The 5x5 fuzzy grid, however, has more fuzzy subspaces than the 3x8 fuzzy grid: 25 > 24. Thus the 3x8 fuzzy grid is viewed as being more interpretable than the 5x5 fuzzy grid if we use the number of fuzzy subspaces (i.e., the number of fuzzy rules) as an interpretability measure.

ISBN: 978-989-95079-6-8



(a) 5x5 fuzzy grid. (b) 3x8 fuzzy grid.

Figure 4: Example 2 with the 5x5 and 3x8 fuzzy grids.

As we have already explained in this section using the two examples, the comparison of different fuzzy grids with respect to their interpretability is not so easy. This means that the choice of an interpretability measure is difficult. If we use the number of fuzzy subspaces in Fig. 4, the 3x8 fuzzy grid is viewed as being more interpretable than the 5x5 fuzzy grid. When we use not only the number of fuzzy subspaces but also the number of membership functions, these two fuzzy grids are viewed as being non-dominated with each other with respect to the interpretability. In this case, the multi-objective formulation in (7) is handled as the three-objective problem:

$$\begin{aligned} &\text{Maximize } \mathit{Accuracy}(S), \text{ and} \\ &\quad \text{minimize } \{\mathit{Complexity}_1(S), \mathit{Complexity}_2(S)\}, \end{aligned} \quad (8)$$

where $\mathit{Complexity}_1(S)$ and $\mathit{Complexity}_2(S)$ are different complexity measures to be minimized (e.g., the number of fuzzy subspaces and the number of membership functions).

Even if we use these two measures, the 3x5 fuzzy grid is viewed as more interpretable than the 4x4 fuzzy grid. Thus we need another measure if we want to include some bias toward fuzzy grids with the same fuzzy partition for all input variables. For example, the 4x4 fuzzy grid is evaluated as more interpretable than the 3x5 fuzzy grid if the maximum number of membership functions for each input variable is used as an interpretability measure ($\max\{4, 4\} < \max\{3, 5\}$).

3 Interpretability of Fuzzy Rule-Based Systems

In genetic fuzzy systems [22]-[24], almost all aspects of fuzzy rule-based systems can be optimized since genetic algorithms perform continuous, discrete and combinatorial optimization. For example, genetic fuzzy systems can be used for choosing an appropriate type of fuzzy rules (e.g., Takagi-Sugeno, simplified Takagi-Sugeno and Mamdani). In this section, we discuss the interpretability of fuzzy rule-based systems with different fuzzy partitions and different types of fuzzy rules.

Let us consider a simple function approximation problem of a single-input and single-output system $y = f(x)$ in Fig. 5. Our task is to design an accurate and interpretable fuzzy rule-based system from the given input-output data in Fig. 5. For this task, Takagi-Sugeno fuzzy rules are written as

$$\text{Rule } R_i: \text{If } x \text{ is } A_i \text{ then } y = a_i + b_i x, \ i = 1, 2, ..., N, \quad (9)$$

where i is a rule index, A_i is an antecedent fuzzy set, a_i and b_i are real number coefficients of a consequent linear function of each fuzzy rule, and N is the total number of fuzzy rules.

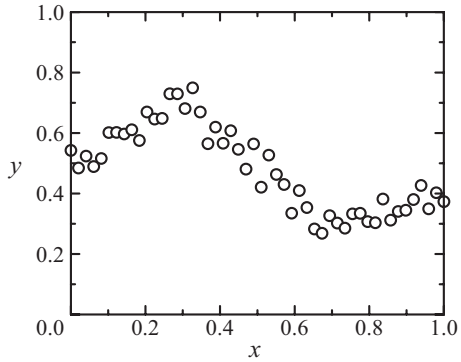


Figure 5: Input-output data in Examples 3 and 4.

When an input value x is presented to the fuzzy rule-based system with the N fuzzy rules in (9), the output value is estimated as follows:

$$y(x) = \frac{\sum_{i=1}^{N} (a_i + b_i x) \cdot \mu_{A_i}(x)}{\sum_{i=1}^{N} \mu_{A_i}(x)}, \quad (10)$$

where $y(x)$ is the estimated output value for the input value x, and $\mu_{A_i}(x)$ is the membership value of the antecedent fuzzy set A_i for the input value x.

We can use the same fuzzy reasoning mechanism for the simplified version of Takagi-Sugeno fuzzy rules:

$$\text{Rule } R_i: \text{If } x \text{ is } A_i \text{ then } y \text{ is } h_i, \ i = 1, 2, ..., N, \quad (11)$$

where h_i is a consequent real number.

Example 3: From the input-output data in Fig. 5, one may think that they can be approximated by a fuzzy rule-based system with three Takagi-Sugeno fuzzy rules. An example of such a fuzzy rule-based system is shown in Fig. 6 where each of the three lines (1), (2) and (3) is the consequent linear function of each of the three fuzzy rules with the trapezoidal antecedent fuzzy sets A_1, A_2 and A_3. The same input-output data can be also approximated by a fuzzy rule-based system with four simplified Takagi-Sugeno fuzzy rules as shown in Fig. 7. Each fuzzy rule in Fig. 7 has a triangular membership function A_i and a consequent real number h_i. The question is which is more interpretable between the two fuzzy rule-based systems in Fig. 6 and Fig. 7.

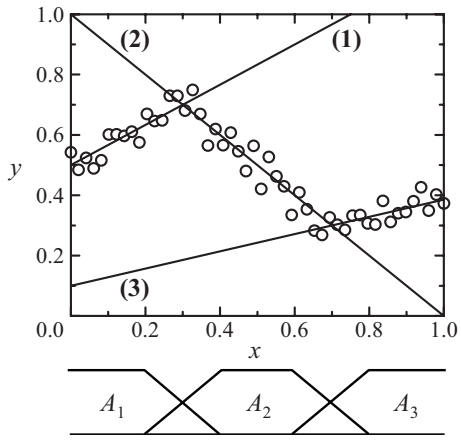


Figure 6: Three Takagi-Sugeno fuzzy rules.

ISBN: 978-989-95079-6-8

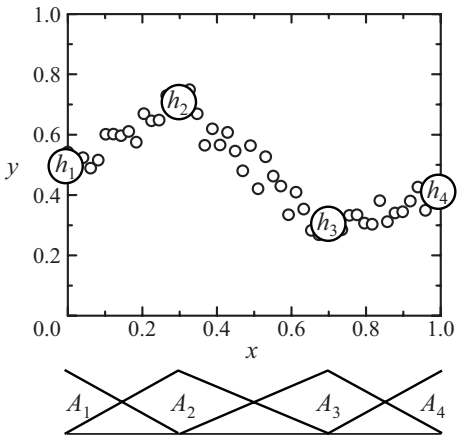


Figure 7: Four simplified Takagi-Sugeno fuzzy rules.

In Fig. 8 and Fig. 9, we show the fuzzy reasoning results by these two fuzzy rule-based systems in Fig. 6 and Fig. 7, respectively. We can see that similar results were obtained from the two fuzzy rule-based systems. Since there is no large difference in the approximation accuracy between Fig. 8 and Fig. 9, the interpretability will play an important role in the selection between the two fuzzy models in Fig. 6 and Fig. 7. If we use the number of fuzzy rules as an interpretability measure, the Takagi-Sugeno model in Fig. 6 is viewed as being more interpretable than the simplified Takagi-Sugeno model in Fig. 7. On the other hand, if we use the total number of parameters (i.e., a_i, b_i and h_i) in the consequent part of the fuzzy rules as an interpretability measure, the simplified Takagi-Sugeno model is evaluated as more interpretable than the Takagi-Sugeno model (i.e., 4 < 6).

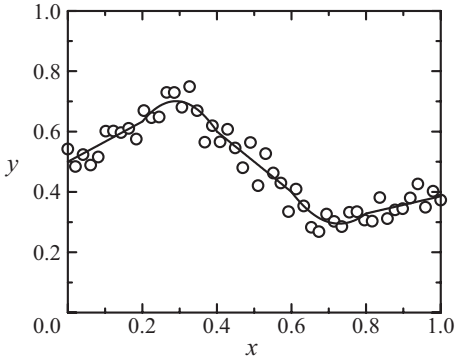


Figure 8: Results by the three Takagi-Sugeno fuzzy rules.

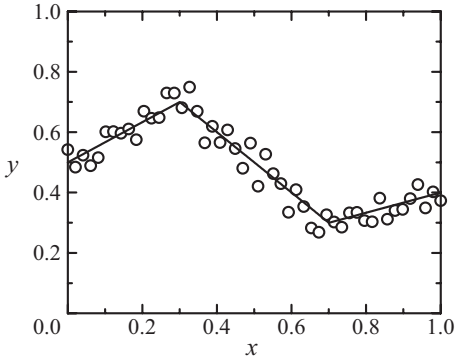


Figure 9: Results by the four simplified fuzzy rules.

Example 4: The given input-output data in Fig. 5 can be also approximated by the two Takagi-Sugeno fuzzy rules in Fig. 10. Whereas data points around $x = 0.5$ are far from the two consequent linear functions in Fig. 10, they can be approximated through the interpolation mechanism of the fuzzy reasoning in Eq. (10). Fig. 11 is the fuzzy reasoning result by the two fuzzy rules in Fig. 10. We can see from Fig. 11 that good approximation was realized by the two fuzzy rules in Fig. 10. Actually, the fuzzy reasoning result in Fig. 11 by the two Takagi-Sugeno fuzzy rules is similar to Fig. 8 and Fig. 9. The question is which is more interpretable between Fig. 6 with the three rules and Fig. 10 with the two rules.

It is clear that Fig. 10 is simpler than Fig. 6 with respect to various aspects of fuzzy rule-based systems (e.g., the number of fuzzy rules, the number of membership functions, and the number of parameters). However, one may think that Fig. 6 is more intuitive than Fig. 10. If we use the local accuracy of each liner function [5] as an interpretability measure, the three Takagi-Sugeno fuzzy rules in Fig. 6 are evaluated as being more interpretable than the two rules in Fig. 10.

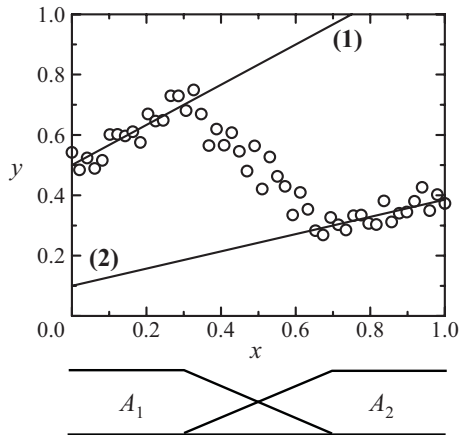


Figure 10: Two Takagi-Sugeno fuzzy rules.

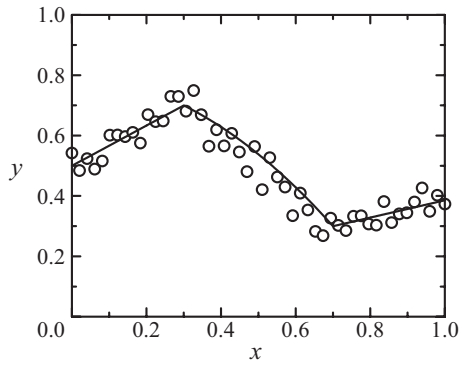


Figure 11: Results by the two Takagi-Sugeno fuzzy rules.

Example 5: We can use different types of fuzzy rules in a single fuzzy rule-based system. We show an example of such a fuzzy rule-based system in Fig. 12 where the second fuzzy rule with the antecedent fuzzy set A_2 has a consequent linear function (line (2)). Each of the other two fuzzy rules with the antecedent fuzzy sets A_1 and A_3 has a consequent real number. As we can expect, good approximation was realized by these three fuzzy rules (due to the page limitation, we can not show

ISBN: 978-989-95079-6-8

the fuzzy reasoning result). The given input-output data in Fig. 12 can be also approximated with a similar accuracy by the four simplified Takagi-Sugeno fuzzy rules in Fig. 13. Since the first two fuzzy rules with the antecedent fuzzy sets A_1 and A_2 in Fig. 13 have the same consequent real number, they can be merged into a single rule. The last two fuzzy rules with A_3 and A_4 in Fig. 13 can be also merged into a single rule. As a result, we have a fuzzy rule-based system with the two simplified Takagi-Sugeno fuzzy rules in Fig. 14.

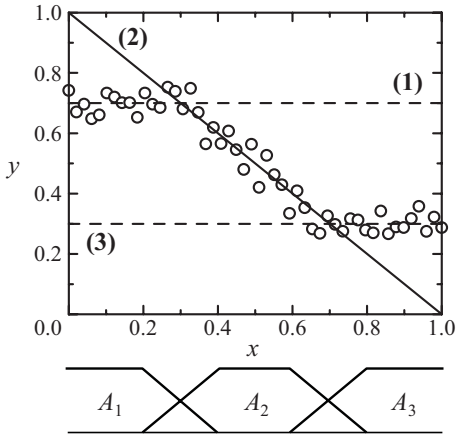


Figure 12: A Takagi-Sugeno rule and two simplified rules.

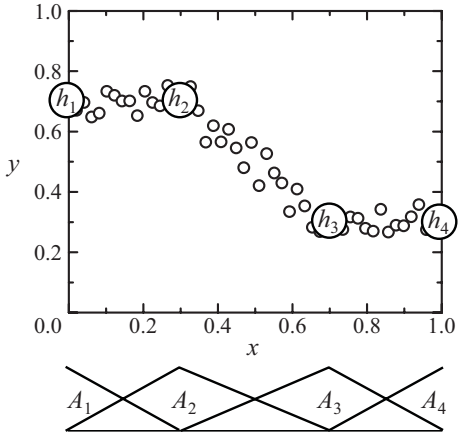


Figure 13: Four simplified Takagi-Sugeno rules.

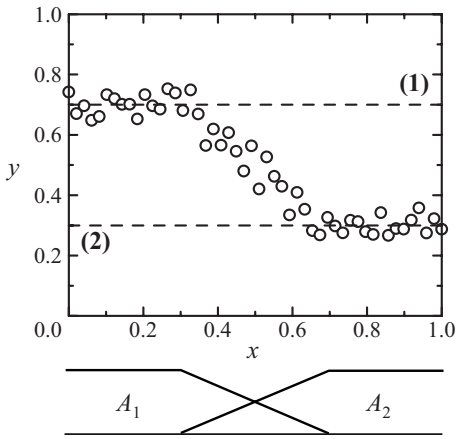


Figure 14: Two simplified Takagi-Sugeno rules.

It is clear that the fuzzy rule-based system with only the two fuzzy rules in Fig. 14 is the simplest one in Figs. 12-14. However, we have no definite answer to the question: Which is the most interpretable in the three models in Figs. 12-14?

4 Conclusions

In this paper, we demonstrated the difficulty in evaluating the interpretability of fuzzy rule-based systems. As shown in this paper, the evaluation of the interpretability is difficult even in very simple situations. Different fuzzy rule-based systems are viewed as being more interpretable according to different interpretability measures. This means that the choice of an appropriate interpretability measure is important in the design of fuzzy rule-based systems. At the same time, such a choice is difficult as shown in this paper. We hope that this paper will activate discussions on the interpretability and help to develop new approaches to fuzzy modelling based on the accuracy-interpretability tradeoff analysis.

References

[1] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka, Construction of fuzzy classification systems with rectangular fuzzy rules using genetic algorithms, *Fuzzy Sets and Systems* 65 (2/3) 237-253, 1994.

[2] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka, Selecting fuzzy if-then rules for classification problems using genetic algorithms, *IEEE Trans. on Fuzzy Systems* 3 (3) 260-270, 1995.

[3] H. Ishibuchi, T. Murata, and I. B. Turksen, Selecting linguistic classification rules by two-objective genetic algorithms, *Proc. of 1995 IEEE International Conference on Systems, Man and Cybernetics*, 1410-1415, 1995.

[4] M. Setnes, R. Babuska, and B. Verbruggen, Rule-based modeling: Precision and transparency, *IEEE Trans. on Systems, Man, and Cybernetics - Part C* 28 (1) 165-169, 1998.

[5] J. Yen, L. Wang, and G. W. Gillespie, Improving the interpretability of TSK fuzzy models by combining global learning and local learning, *IEEE Trans. on Fuzzy Systems* 6 (4) 530-537, 1998.

[6] Y. Jin, W. von Seelen, and B. Sendhoff, On generating FC^3 fuzzy rule systems from data using evolution strategies, *IEEE Trans. on Systems, Man, and Cybernetics - Part B* 29 (6) 829-845, 1999.

[7] D. Nauck and R. Kruse, Obtaining interpretable fuzzy classification rules from medical data, *Artificial Intelligence in Medicine* 16 (2) 149-169, 1999.

[8] Y. Jin, Fuzzy modeling of high-dimensional systems: Complexity reduction and interpretability improvement, *IEEE Trans. on Fuzzy Systems* 8 (2) 212-221, 2000.

[9] M. Setnes and H. Roubos, GA-based modeling and classification: Complexity and performance, *IEEE Trans. on Fuzzy Systems* 8 (5) 509-522, 2000.

[10] L. Castillo, A. Gonzalez, and R. Perez, Including a simplicity criterion in the selection of the best rule in a genetic fuzzy learning algorithm, *Fuzzy Sets and Systems* 120 (2) 309-321, 2001.

[11] H. Roubos and M. Setnes, Compact and transparent fuzzy models and classifiers through iterative complexity reduction, *IEEE Trans. on Fuzzy Systems* 9 (4) 516-524, 2001.

[12] J. Abonyi, J. A. Roubos, and F. Szeifert, Data-driven generation of compact, accurate, and linguistically sound fuzzy classifiers based on a decision-tree initialization, *International Journal of Approximate Reasoning* 32 (1) 1-21, 2003.

[13] R. Alcala, J. R. Cano, O. Cordon, F. Herrera, P. Villar, and I. Zwir, Linguistic modeling with hierarchical systems of weighted linguistic rules, *International Journal of Approximate Reasoning* 32 (2-3) 187-215, 2003.

ISBN: 978-989-95079-6-8

[14] Y. C. Hu, R. S. Chen, and G. H. Tzeng, Finding fuzzy classification rules using data mining techniques, *Pattern Recognition Letters* 24 (1-3) 509-519, 2003.

[15] J. Casillas, O. Cordon, 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. on Fuzzy Systems* 13 (1) 13-29, 2005.

[16] R. Alcala, J. Alcala-Fdez, J. Casillas, O. Cordon, and F. Herrera, Hybrid learning models to get the interpretability-accuracy trade-off in fuzzy modeling, *Soft Computing* 10 (9) 717-734, 2006.

[17] R. Alcala, J. Alcala-Fdez, F. Herrera, and J. Otero, Genetic learning of accurate and compact fuzzy rule based systems based on the 2-tuples linguistic representation, *International Journal of Approximate Reasoning* 44 (1) 45-64, 2007.

[18] E. van Broekhoven, V. Adriaenssens, and B. De Baets, Interpretability-preserving genetic optimization of linguistic terms in fuzzy models for fuzzy ordered classification: An ecological case study, *International Journal of Approximate Reasoning* 44 (1) 65-90, 2007.

[19] P. Pulkkinen and H. Koivisto, Identification of interpretable and accurate fuzzy classifiers and function estimators with hybrid methods, *Applied Soft Computing* 7 (2) 520-533, 2007.

[20] S. Destercke, S. Guillaume, and B. Charnomordic, Building an interpretable fuzzy rule base from data using orthogonal least squares - Application to a depollution problem, *Fuzzy Sets and Systems* 158 (18) 2078-2094, 2007.

[21] J. M. Alonso, L. Magdalena, and S. 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.

[22] O. Cordon, F. Herrera, F. Hoffmann, and L. Magdalena, *Genetic Fuzzy Systems*, World Scientific, 2001.

[23] O. Cordon, F. Gomide, F. Herrera, 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.

[24] F. Herrera, Genetic fuzzy systems: Status, critical considerations and future directions, *International Journal of Computational Intelligence Research* 1 (1) 59-67, 2005.

[25] H. Ishibuchi, T. Murata, and I. B. Turksen, Single-objective and two-objective genetic algorithms for selecting linguistic rules for pattern classification problems, *Fuzzy Sets and Systems* 89 (2) 135-150, 1997.

[26] H. Ishibuchi, T. Nakashima, and T. Murata, Three-objective genetics-based machine learning for linguistic rule extraction, *Information Sciences* 136 (1-4) 109-133, 2001.

[27] H. Ishibuchi and T. Yamamoto, Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining, *Fuzzy Sets and Systems* 141 (1) 59-88, 2004.

[28] H. Wang, S. Kwong, Y. Jin, W. Wei, and K. F. Man, Multi-objective hierarchical genetic algorithm for interpretable fuzzy rule-based knowledge extraction, *Fuzzy Sets and Systems* 149 (1) 149-186, 2005.

[29] 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.

[30] M. Cococcioni, P. Ducange, B. Lazzerini, and F. Marcelloni, A Pareto-based multi-objective evolutionary approach to the identification of Mamdani fuzzy systems, *Soft Computing* 11 (11) 1013-1031, 2007.

[31] A. F. Gomez-Skarmeta, F. Jimenez, and G. Sanchez, Improving interpretability in approximative fuzzy models via multiobjective evolutionary algorithms, *International Journal of Intelligent Systems* 22 (9) 943-969, 2007.

[32] R. Alcala, M. J. Gacto, F. Herrera, and J. Alcala-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

[33] J. Gonzalez, I. Rojas, H. Pomares, L. J. Herrera, A. Guillen, J. M. Palomares, and 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* 44 (1) 32-44, 2007

[34] M. J. Gacto, R. Alcala, 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.

[35] 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, 2009.

[36] L. Sanchez, J. Otero, and I. Couso, Obtaining linguistic fuzzy rule-based regression models from imprecise data with multiobjective genetic algorithms, *Soft Computing* 13 (5) 467-479, 2009.

[37] H. Ishibuchi, Multiobjective genetic fuzzy systems: Review and future research directions, *Proc. of 2007 IEEE International Conference on Fuzzy Systems*, 913-918, 2007.

[38] Y. Jin (ed.), *Multi-Objective Machine Learning*, Springer, Berlin, 2006.

[39] A. Ghosh, S. Dehuri, and S. Ghosh (Eds.), *Multi-Objective Evolutionary Algorithms for Knowledge Discovery from Data Bases*, Springer, Berlin, 2008.

[40] K. Miettinen, *Nonlinear Multiobjective Optimization*, Kluwer, Boston, 1998.

[41] A. P. Wierzbicki, M. Makowski, and J. Wessels (eds.), *Model-Based Decision Support Methodology with Environmental Applications*, Kluwer, Dordrecht, 2000.

[42] K. Miettinen and M. Makela, On scalarizing functions in multiobjective optimization, *OR Spectrum* 24 (2) 193-213, 2002.

[43] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Trans. on Evolutionary Computation* 6 (2) 182-197, 2002.

[44] E. Zitzler and L. Thiele, Multiobjective evolutionary algorithms: A comparative case study and the strength Pareto approach, *IEEE Trans. on Evolutionary Computation* 3 (4) 257-271, 1999.

[45] E. Zitzler, M. Laumanns, and L. Thiele, SPEA2: Improving the strength Pareto evolutionary algorithms, *TIK-Report 103*, Computer Engineering and Networks Laboratory (TIK), Swiss Federal Institute of Technology (ETH) Zurich, 2001.

[46] S. Guillaume, Designing fuzzy inference systems from data: An interpretability-oriented review, *IEEE Trans. on Fuzzy Systems* 9 (3) 426-443, 2001.

[47] J. Casillas, O. Cordon, F. Herrera, and L. Magdalena (eds.), *Interpretability Issues in Fuzzy Modeling*, Springer, 2003.

[48] J. Casillas, O. Cordon, F. Herrera, and L. Magdalena (eds.), *Accuracy Improvements in Linguistic Fuzzy Modeling*, Springer, 2003.

[49] R. Mikut, J. Jakel, and L. Groll, Interpretability issues in data-based learning of fuzzy systems, *Fuzzy Sets and Systems* 150 (2) 179-197, 2005.

[50] C. Mencar, G. Castellano, and A. M. Fanelli, On the role of interpretability in fuzzy data mining, *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems* 15 (5) 521-537, 2007.

[51] C. Mencar and A. M. Fanelli, Interpretability constraints for fuzzy information granulation, *Information Sciences* 178 (24) 4585-4618, 2008.

[52] S. M. Zhou and J. Q. Gan, Low-level interpretability and high-level interpretability: a unified view of data-driven interpretable fuzzy system modelling, *Fuzzy Sets and Systems* 159 (23) 3091-3131, 2008.

ISBN: 978-989-95079-6-8

Stability Analysis of Continuous PI-like Fuzzy Control Systems based on Vector Norms Approach

Anis Sakly[1] Basma Zahra[1] Mohamed Benrejeb[1]

[1]Unité de recherche L.A.R.A. Automatique[1]
Ecole Nationale d'Ingénieurs de Tunis[1]
BP 37, Belvédère 1002 Tunis, Tunisie
sakly_anis@yahoo.fr, zahra_basma@yahoo.fr, mohamed.benrejeb@enit.rnu.tn

Abstract—*This paper proposes a practical stability analysis of a particular class of continuous PI-like fuzzy control systems based on the convergence of regular vector norms. This approach is based on the comparison, the overvaluing principle and the application of Borne and Gentina criterion. The stability conditions issued from vector norms correspond to a vector Lyapunov function. A comparison system relative to a regular vector norm will be used in order to get the simple arrow form of the state matrix that yields to a suitable use of Borne and Gentina criterion for establishing sufficient conditions for global asymptotic stability.*

Keywords—Continuous PI-like fuzzy control systems, global asymptotic stability, vector norms, Borne and Gentina criterion, Arrow form state matrix.

1 Introduction

Over the last few decades, fuzzy control has received great attention from engineering and science community and has marked rapid growth. However, there isn't a general method for analysis or synthesis of such control strategy. In particular, stability analysis of control systems is an essential step before synthesizing and elaborating the control law.

In this way, many researches have been done on the stability study of fuzzy control systems since their appearance in the middle of the 70's [17].

Indeed, the majority of these studies concern TSK fuzzy systems since consequent fuzzy variables of this type (known as type III) are explicit and many important results have been obtained [13, 18, 23] based on Lyapunov stability theory. Nevertheless, there have been a limited number of stability studies of Mamdani type fuzzy systems (known as type I) due to their complexity, and for this type, fuzzy variables are linguistically understandable in both the premises and consequents. Most of these papers provide a stability analysis of a linear plant controlled by a fuzzy controller and they regard the latter as a nonlinear controller corresponding to a Lur'e system. So, the stability problem of fuzzy control systems comes down to conventional nonlinear stability theory.

Among the methods that have been used, Popov's theorem for time-invariant nonlinearity [16], circle criterion for time-variant nonlinearity [19, 20] and its extended version, the conicity criterion [1].

Other approaches have been used: ones can cite: the hyperstability approach which is equivalent to passivity [9, 10, 24], the input-output stability based on the use of small gain theorem [11, 25] and the Lyapunov function approach [8, 12].

We do remark also that the majority of these papers simplify the study by considering the consequents of the fuzzy system as singletons [22]. So, this type (known as type II) is a special case of both TSK fuzzy systems and those of Mamdani.

In this paper, the stability study of particular class of fuzzy PI controllers of type Mamdani is presented. This study is based on the application of the Borne and Gentina criterion [6] which uses Kotelyanski conditions. In [3], it has been shown that when system state matrix is in arrow form, then Borne and Gentina criterion becomes very simple to apply.

This approach has been used in many previous works [2, 4, 5, 21]. In this way, results proposed in [21] will be used and generalized for stability analysis of continuous Mamdani fuzzy systems in the case of nonlinear processes to be controlled.

The paper is organised as follows. The next section is devoted to the description of the particular class of PI-like fuzzy control system. In section 3, stability conditions of the fuzzy system are established by using Borne and Gentina criterion and vector norms approach. The stability conditions proposed are illustrated with an example presented in section 4. Finally, concluding remarks are drown in section 5.

2 Particular class of PI-like fuzzy controllers

The fuzzy PI control system considered in this study has two inputs the error e and its derivative de and one output the control derivative as shown in Fig. 1, where k_e, k_{de} et k_{du} are scale factors.

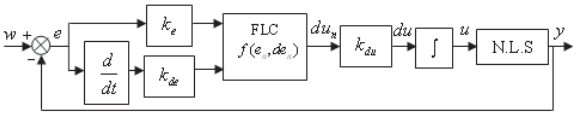


Figure 1: Fuzzy control system.

A particular class of fuzzy PI controllers of type Mamdani is obtained by considering a strong triangular partition of the normalized variables e_n, de_n and du_n presented in Fig. 2.

The rule base considered is an r×r traditional rule table that is of antidiagonal type such that the Mac Vicar-Whelan one (Table 1).

ISBN: 978-989-95079-6-8

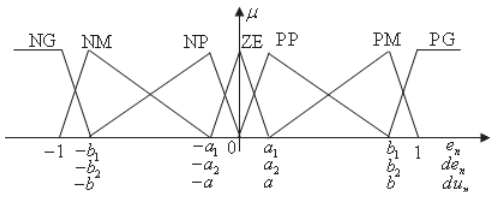


Figure 2: Fuzzy subsets partition.

Table 1 : The Mac Vicar-Whelan rule base.

e_n \ de_n	NG	NM	NP	ZE	PP	PM	PG
NG	NG	NG	NG	NG	NM	NP	ZE
NM	NG	NG	NM	NM	NP	ZE	PP
NP	NG	NM	NP	NP	ZE	PP	PM
ZE	NG	NM	NP	ZE	PP	PM	PG
PP	NM	NP	ZE	PP	PP	PM	PG
PM	NP	ZE	PP	PM	PM	PG	PG
PG	ZE	PP	PM	PG	PG	PG	PG

Let $\sigma(e_n, de_n)$ the surface in the space (e_n, de_n, du_n), verifying the two properties [21]:

$$\begin{cases} \text{i) If } \sigma = 0 \text{ then the input-output characteristic surface } du_n(e_n, de_n) = 0 \\ \text{ii) It exists } k>0 \text{ such as } du_n(k\sigma - du_n) \geq 0 \text{ for all } e_n \text{ and } de_n. \end{cases} \quad (1)$$

The first property means that the intersection of the overvaluing surface σ with the plan (e_n, de_n) is a part of the intersection of the characteristic surface $du_n(e_n, de_n)$ with the same plan. The curve $\sigma = 0$ is a straight line when the fuzzy input partition is identical for the inputs and it represents the second bissectrix as shown in Fig. 3.

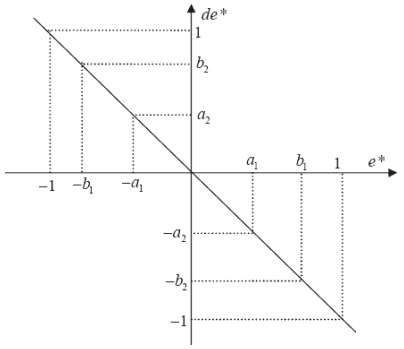


Figure 3: Curve $\sigma = 0$.

According to [21], in the case of an identical partition for the two inputs e_n and de_n (i.e. $a_1 = a_2$ and $b_1 = b_2$) and when using an antidiagonal rule table like the Mac-Vicar Whelan one. This approach is based on the overvaluing and the undervaluing of the fuzzy controller characteristic surface $du_n = (e_n, de_n)$ by two plans crossing the plan (e_n, de_n) in the second bissectrix which is the curve $\sigma = 0$, in this case given by the following equation:

$$e_n + de_n = 0 \quad (3)$$

The slopes of these plans are respectively $k_{\max}$ and $k_{\min}$, those values depend on the geometric parameters of the input variables e_n and de_n (a_i and $b_i, i = 1,2$) and the output variable du_n (a and b), on the other hand they depend on the used inference method.

In this way for each point (e_n, de_n) of $[-1,1]\times[-1,1]$ the controller output du_n is such that:

$$k_{\min}(e_n + de_n) \leq du_n \leq k_{\max}(e_n + de_n).$$

Otherwise $du_n = f(.)(e_n + de_n)$, in such way we can write:

$$k_{\min} \leq f(.) \leq k_{\max} \quad (4)$$

where $f(.)$ is a nonlinear gain.

By programming, we can obtain values of $k_{\max}$ and $k_{\min}$ that allow respectively overvaluing and undervaluing the characteristic surface of the fuzzy controller by PI plans.

In particular, for equidistant fuzzy partition of variables e_n, de_n and du_n ($a_1 = a_2 = a = 0.33$ and $b_1 = b_2 = b = 0.67$), max-min inference method and centroid defuzzification method, we obtain the action surface for the fuzzy controller given by Fig. 4.

This surface is obtained by distorting the discourse universes of e_n and de_n to 100 points, we obtain:

$$k_{\max} = 1,86 \text{ and } k_{\min} = 0,47 \quad (5)$$

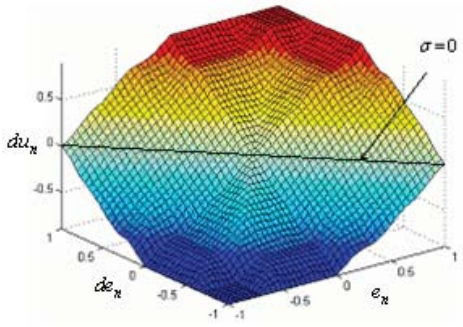


Figure 4: Action surface of the particular Mamdani fuzzy controller.

So the PI fuzzy control system of the Fig. 1 can be set in the following form (Fig. 5):

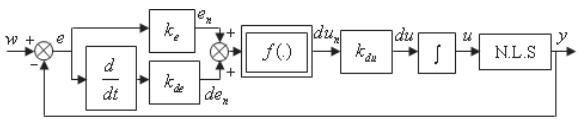


Figure 5: Equivalent fuzzy control system.

3 Proposed stability conditions

3.1 Problem formulatiom

Supposing that the system to be controlled which is nonlinear is represented by the following state matrix given in the Frobenius form such that:

$$\begin{cases} \dot{x} = A(.)x + B(.)u \\ y = C(.)x \end{cases} \quad x \in \mathbb{R}^n \quad (6.a)$$

where:

$$A(.) = \begin{bmatrix} 0 & \dots & 0 & -a_n(.) \\ 1 & \ddots & \vdots & \vdots \\ \vdots & \ddots & 0 & \vdots \\ 0 & \dots & 1 & -a_1(.) \end{bmatrix}, B(.) = \begin{bmatrix} b_n(.) \\ \vdots \\ \vdots \\ b_1(.) \end{bmatrix}, C^T(.) = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \quad (6.b)$$

From the diagram given in Fig. 5, we have $du = \dot{u}$, supposing that $v = du = \dot{u}$ and $\xi = u$, which leads to : $\dot{\xi} = v$, the nonlinear system equipped with the integration can be represented by the following state matrix:

ISBN: 978-989-95079-6-8

$$\dot{z} = \begin{bmatrix} A(.) & B(.) \\ 0 & 0 \end{bmatrix} z + \begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix} v \quad z \in \mathbb{R}^{n+1} \tag{7}$$

where: $z = \begin{bmatrix} x \\ \xi \end{bmatrix}$ and $\dot{z} = \begin{bmatrix} \dot{x} \\ \dot{\xi} \end{bmatrix}$

supposing that: $A'(.) = \begin{bmatrix} A(.) & B(.) \\ 0 & 0 \end{bmatrix}$ and $B'(.) = B' = \begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix}$.

So we can write $y = x_n = z_n$ where x_n (respectively z_n) is the n^{th} state variable of state vector x (respectively z), thus: $y = C'(.)z$ where $C'(.) = C' = \begin{bmatrix} 0 & \dots & 1 & 0 \end{bmatrix}$

On the other hand, and by referring to the diagram in Fig. 5 in the autonomous regime, we can write:

$$\begin{aligned} v &= k_{du} f(.)\left[e_n + de_n\right] = -k_{du} f(.)\left[k_e e + k_{de}\dot{e}\right] \\ &= -k_{du} f(.)\left[k_e y + k_{de}\dot{y}\right] \end{aligned} \tag{8}$$

finally: $v = -k_{du} f(.)\left[k_e C'(.)z + k_{de} C'(.)\dot{z}\right]$

then:

$$\begin{aligned} \dot{z} &= A'(.)z + B'(.)v = A'(.)z - B'(.)k_{du} f(.)\left[k_e C'z + k_{de} C'\dot{z}\right] \\ &= A'(.)z - k_{du} f(.)\left[k_e B'C'z + k_{de} B'C'\dot{z}\right] \end{aligned}$$

that leads to:

$$\left[I_{n+1} + k_{du}k_{de} f(.)B'C'\right]\dot{z} = \left[A'(.) - k_{du}k_e f(.)B'C'\right]z \tag{9}$$

supposing now: $\begin{cases} N = I_{n+1} + k_{du}k_{de} f(.)B'C' \\ M = A'(.) - k_{du}k_e f(.)B'C' \end{cases}$

let:

$$N = \begin{bmatrix} 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & 0 & \ddots & 0 \\ 0 & \dots & 0 & k_{du}k_{de} f(.) & 1 \end{bmatrix}, \; M = \begin{bmatrix} 0 & \dots & 0 & -a_n(.) & b_n(.) \\ 1 & \ddots & & \vdots & \vdots \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ \vdots & & 1 & -a_1(.) & b_1(.) \\ 0 & \dots & 0 & -k_{du}k_e f(.) & 0 \end{bmatrix}$$

where: $\det(N) = 1$ and $N^{-1} = \begin{bmatrix} 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & 0 & \ddots & 0 \\ 0 & \dots & 0 & -k_{du}k_{de} f(.) & 1 \end{bmatrix}$

Finally we obtain the following description of the closed loop system:

$$\dot{z} = A_C(.)z \tag{10}$$

where:

$$A_C(.) = N^{-1}M = \begin{bmatrix} 0 & \dots & \dots & -a_n(.) & b_n(.) \\ 1 & & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ \vdots & & 1 & -a_1(.) & b_1(.) \\ 0 & \dots & -k_{du}k_{de} f(.) & k_{du} f(.)(k_{de}a_1(.) - k_e) & -k_{du}k_{de} f(.)b_1(.) \end{bmatrix} \tag{11}$$

For establishing stability conditions for the studied system, we have to make a basic change leading to new representation of the system to get best exploitation.

3.2 *New state representation of the system*

We consider the following passage matrix P allowing passing from the matrix $A_C(.)$ to a matrix $A_C^{'}(.)$, such that:

$$P = \begin{bmatrix} 1 & \alpha_1 & \dots & (\alpha_1)^{n-1} & 0 \\ \vdots & \dots & & \vdots & \vdots \\ 1 & \alpha_{n-1} & \dots & (\alpha_{n-1})^{n-1} & \vdots \\ 0 & \dots & & 1 & 0 \\ 0 & \dots & \dots & 0 & 1 \end{bmatrix} \tag{12}$$

We note: $z' = Pz$ and let $z = P^{-1}z'$.

In this way:

$$\dot{z}' = PA_C(.)P^{-1}z' = A_C^{'}(.)z' \tag{13.a}$$

where: $A_C^{'}(.) = PA_C(.)P^{-1}$

the matrix $A_C^{'}(.)$ is given by:

$$A_C^{'}(.) = \begin{bmatrix} \alpha_1 & 0 & \dots & \delta_1(.) & \upsilon_1(.) \\ \vdots & \ddots & & \vdots & \vdots \\ 0 & \dots & \alpha_{n-1} & \delta_{n-1}(.) & \upsilon_{n-1}(.) \\ \beta_1 & \dots & \beta_{n-1} & \sigma(.) & b_1(.) \\ \lambda_1(.) & \dots & \lambda_{n-1}(.) & \varphi(.) & \psi(.) \end{bmatrix} \tag{13.b}$$

where:

$$\beta_i = \frac{1}{\prod_{\substack{j=1 \\ j \neq i}}^{n} (\alpha_i - \alpha_j)} \quad \forall i = 1,...,n-1 \tag{13.c}$$

$$\begin{cases} \delta_i(.) = -D(.,\alpha_i) & \forall i = 1,...,n-1 \\ D(.,\alpha) = \alpha^n + \sum_{j=1}^{n} a_j(.)\alpha^{n-j} \end{cases} \tag{13.d}$$

$$\begin{cases} \upsilon_i(.) = N(.,\alpha_i) & \forall i = 1,...,n-1 \\ N(.,\alpha) = \sum_{j=1}^{n} b_j(.)\alpha^{n-j} \end{cases} \tag{13.e}$$

$$\sigma(.) = -a_1(.) - \sum_{j=1}^{n-1} \alpha_j \tag{13.f}$$

$$\begin{cases} \lambda_i(.) = -k_{du}k_{de} f(.)\beta_i & \forall i = 1,...,n-1 \\ \varphi(.) = k_{du}k_{de} f(.)\left[a_1(.) + \sum_{j=1}^{n-1} \alpha_j\right] - k_{du}k_e f(.) \\ \psi(.) = -k_{du}k_{de} f(.)b_1(.) \end{cases} \tag{13.g}$$

By observing the matrix $A_C^{'}(.)$, we remark that the nonnull elements are located in the diagonal and in both two last rows and two last columns, this matrix is called in the double arrow form. The nonlinear elements are situated in the last row and the two last columns.

We can also deduce that the Borne and Gentina criterion can not be applied in this case, only the usual stability criteria can be used like Holder norms such that max norm, sum norm and module norm.

In order to use suitably the Borne and Gentina criterion, we can think of the isolation of the nonlinear elements in only one row and one column. In this way, we consider a comparison system relative to the following n regular vector norm:

ISBN: 978-989-95079-6-8

$$p(z') = \left[|z'_1|, \ldots, |z'_{n-1}|, \max\left\{ |z'_n|, |z'_{n+1}| \right\} \right]^T \quad (14)$$

The vector z' is of order $(n+1)$ and the vector $p(z')$ is of order n.

Let $Z = p(z')$, so we define the overvaluing system relative to p such that:

$$\dot{Z} = M_C(.)Z \quad (15.a)$$

The matrix $M_C(.)$ is given by:

$$M_C(.) = \begin{bmatrix} \alpha_1 & 0 & \ldots & \gamma_1(.) \\ 0 & \ddots & & \vdots \\ \vdots & \ldots & \alpha_{n-1} & \gamma_{n-1}(.) \\ \mu_1(.) & & \mu_{n-1}(.) & \mu(.) \end{bmatrix} \quad (15.b)$$

where:

$$\gamma_i(.) = |\delta_i(.)| + |\upsilon_i(.)| \qquad \forall i = 1,...,n-1$$
$$= \left| \alpha_i^n + \sum_{j=1}^{n} a_j(.)\alpha_i^{n-j} \right| + \left| \sum_{j=1}^{n} b_j(.)\alpha_i^{n-j} \right| \quad (15.c)$$

$$\mu_i(.) = \max\left\{ |\beta_i|, |\lambda_i(.)| \right\} = |\beta_i| \max\left\{ 1, k_{du}k_{de}f(.) \right\} \quad (15.d)$$

$$\mu(.) = \max\left\{ \sigma(.) + |b_1(.)|, |\varphi(.)| + \psi(.) \right\}$$
$$= \max \left\{ \begin{array}{l} (-a_1(.) - \sum_{j=1}^{n-1} \alpha_j) + |b_1(.)|, \\ \left| k_{du}k_{de}f(.) \left[a_1(.) + \sum_{j=1}^{n-1} \alpha_j \right] - k_{du}k_e f(.) \right| - k_{du}k_{de}f(.)b_1(.) \end{array} \right\} \quad (15.e)$$

The matrix $M_C(.)$ is in the arrow form. The vector norm p allows passing from the matrix in the double arrow form to a matrix in the simple arrow form by decreasing the order of the system from $(n+1)$ to n order.

To apply the Borne and Gentina criterion to the overvaluing matrices, we substitute the nonlinear elements in the last row $(\mu_i(.))$ by constant elements with the following hypothesis:

$$\mu_i(.) = |\beta_i| \text{ for } k_{du}k_{de}f(.) < 1 \quad (16)$$

In these conditions the matrix $M_C(.)$ becomes:

$$M'_C(.) = \begin{bmatrix} \alpha_1 & 0 & \ldots & \gamma_1(.) \\ 0 & \ddots & & \vdots \\ & & \alpha_{n-1} & \gamma_{n-1}(.) \\ |\beta_1| & \ldots & |\beta_{n-1}| & \mu(.) \end{bmatrix} \quad (17)$$

So the nonlinear elements of the matrix $M'_C(.)$ are isolated in the last column.

3.3 Stability conditions

Theorem 1:

If there exist $\alpha_i < 0$ for $i = 1,...,n-1$, $\alpha_i \neq \alpha_j \ \forall i \neq j$ such that $\forall Z \in S$ where S is a neighbourhood domain:

$$i)\ k_{du}k_{de}f(.) < 1$$
$$ii) - \mu(.) + \sum_{i=1}^{n-1} \gamma_i(.)\alpha_i^{-1}|\beta_i| > 0 \quad (18)$$

then the equilibrium point $Z=0$ for the continuous system is asymptotically stable.

If $S = \mathbb{R}^n$, the stability is global. ■

Proof:

The matrix $M'_C(.)$ has its off diagonal elements positive and the ones non constant are isolated in the last line.

Let the following comparison system: $\dot{Z} = M'_C(.)Z$

Thus, by referring to results obtained in [3], the conditions of the previous theorem can be deduced from the Kotelyanski conditions [6]. These conditions require having the principal minors with alternating signs (see Appendix), the (α_i) are chosen all negative.

$\alpha_1 < 0$

$\alpha_1\alpha_2 > 0$

$\vdots$

$(-1)^{n-1} \prod_{i=1}^{n-1} \alpha_i > 0$

and $(-1)^n \det(M_C) > 0$

The $(n-1)$ first conditions are checked because the α_i are negative, however the last condition yields to:

$$(-1)^n \det\left(M'_C(.)\right) = (-1)^n \begin{vmatrix} \alpha_1 & 0 & \ldots & 0 & \gamma_1(.) \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \ldots & 0 & \alpha_{n-1} & \gamma_{n-1}(.) \\ |\beta_1| & \ldots & \ldots & |\beta_{n-1}| & \mu(.) \end{vmatrix}$$
$$= (-1)^n \left[\mu(.) \prod_{i=1}^{n-1} \alpha_i - \sum_{i=1}^{n-1} \left(|\gamma_i(.)\beta_i| \prod_{\substack{j=1 \\ j \neq i}}^{n-1} \alpha_j \right) \right] > 0$$

then the theorem is obtained by dividing this condition by $\left((-1)^{n-1} \prod_{i=1}^{n-1} \alpha_i \right)$ such that: $-\mu(.) + \sum_{i=1}^{n-1} \gamma_i(.)\alpha_i^{-1}|\beta_i| > 0$. ■

In order to simplify the application of the theorem, we have to exploit the expression of $\mu(.)$. We suppose that:

$$\mu(.) = \left| k_{du}k_{de}f(.) \left[a_1(.) + \sum_{j=1}^{n-1} \alpha_j \right] - k_{du}k_e f(.) \right| - k_{du}k_{de}f(.)b_1(.).$$

Thus, we obtain the following corollary.

Corollary 1:

If there exist $\alpha_i < 0$ for $i = 1,...,n$, $\alpha_i \neq \alpha_j \ \forall i \neq j$ such that $\forall Z \in S$ the three following conditions are checked:

$$i)\ k_{du}k_{de}f(.) < 1$$
$$ii)\left(-a_1(.) - \sum_{j=1}^{n-1} \alpha_j \right) + |b_1(.)| < \mu(.) \quad (19)$$
$$iii) - \mu(.) + \sum_{i=1}^{n-1} \gamma_i(.)\alpha_i^{-1}|\beta_i| > 0$$

then the equilibrium point $Z=0$ for the system is asymptotically stable.

If $S = \mathbb{R}^n$, the stability is global. ■

Remark 1 :

For $n = 1$, the stability condition issued from Theorem 1 is such that: $\mu(.) < 0$. ■

ISBN: 978-989-95079-6-8

4 Example

For the validation of the results obtained we consider the stabilization of a fuzzy control system, where the controller is of type PI fuzzy controller and the system to be controlled is a speed of DC motor. This motor is supposed with a shunt excitation represented by a transfer function of a first order system which is preceded by a nonlinear element, corresponding to nonlinearity characteristic of the magnetic flux for example (Fig. 6).

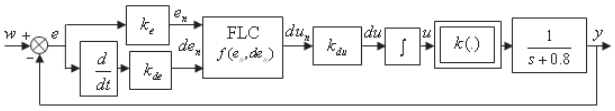


Figure 6: Fuzzy control system for the speed of DC motor with shunt excitation.

The nonlinear gain $k(.)$ is represented by the following allure (Fig. 7):

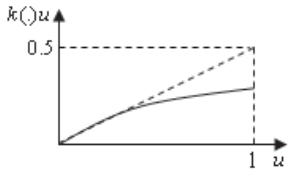


Figure 7: Characteristic of the nonlinear gain.

The system is described by the following state equation:

$$\begin{cases} \dot{x} = A(.)x + B(.)u \\ y = C(.)x \end{cases} \quad (20.a)$$

where:

$$\begin{cases} A(.) = A = -0.8 \\ B(.) = k(.) \\ C(.) = C = 1 \end{cases} \quad (20.b)$$

the matrix $A_C(.)$ is given by :

$$A_C(.) = \begin{bmatrix} -0.8 & k(.) \\ 0.8k_{du}k_{de}f(.) - k_{du}k_{de}f(.) & -k_{du}k_{de}f(.)k(.) \end{bmatrix} \quad (21)$$

By making a basic change to the previous system we obtain:

$$\dot{z}' = A'_C(.)z' \quad (22.a)$$

Such that:

$$\begin{cases} A'_C(.) = PA_C(.)P^{-1} \\ P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2 \end{cases} \quad (22.b)$$

so we can write: $A'_C(.) = A_C(.)$ and:

$$M_C(.) = \max\left\{\left|k_{du}k_{de}f(.)0.8 - k_{du}k_{de}f(.)\right| - k_{du}k_{de}f(.)k(.), -0.8 + k(.)\right\}$$

the factor scales k_e and k_{de} are chosen such that:

$k_e = k_{de} = 1$.

According to Remark 1, by applying Theorem 1 we get the following condition: $\mu(.) < 0$

which leads to:

$$\begin{cases} -0.8 + k(.) < 0 \\ k_{du}f(.)\left[0.2 - k(.)\right] < 0 \end{cases} \text{ and so: } \begin{cases} 0.2 < k(.) < 0.8 \\ k_{du}f(.) > 0 \end{cases}$$

The area representing the stability domain of the fuzzy system is given by Fig. 8:

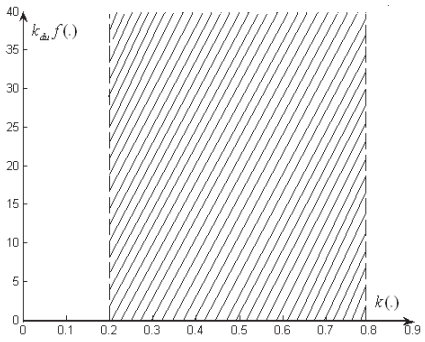


Figure 8: Stability domain of the fuzzy system obtained from Theorem 1.

The application of Corollary 1 allows to deduce the following condition: $-0.8 + k(.) < k_{du}f(.)\left[0.2 - k(.)\right] < 0$

then: $k_{du}f(.) < \dfrac{0.8 - k(.)}{k(.) - 0.2}$ and $0.2 < k(.) < 0.8$.

The area representing the stability domain of the fuzzy system is given by Fig. 9:

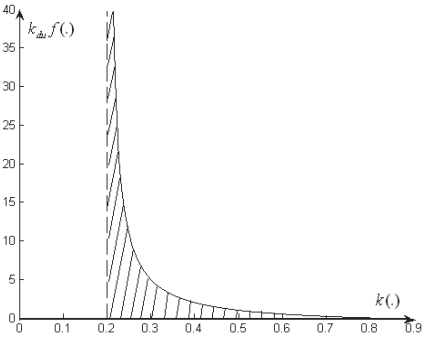


Figure 9: Stability domain of the fuzzy system obtained from Corollary 1.

5 Conclusion

The stability conditions of particular class of PI-like fuzzy control systems were presented in this paper. These conditions were deduced from stability study of overvaluing systems based on vector norms and the application of the Borne and Gentina criterion. After making a basic change on the system, we obtain a state matrix in a double arrow form, to return to the usual arrow form of the matrix and to get matrix with nonlinear elements isolated in only one row or column, we had considered a comparison system relative to a regular vector norm. In this way the Borne and Gentina criterion was used to get sufficient stability conditions. These conditions were applied to nonlinear system given by the control of the speed of DC motor with shunt excitation.

Appendix

Borne-Gentina practical stability criterion [15]

Let consider the nonlinear continuous process described in state space by: $\dot{x} = A(.)x$; $A(.)$ is an $n \times n$ matrix, $A(.) = \{a_{i,j}\}$. If the overvaluing matrix $M(A(.))$ has its non constant elements isolated in only one row, the verification of the Kotelyanski condition enables to conclude to the stability of the initial system.

As an example, if the non constant elements are isolated in only one row of $A(.)$, Kotelyanski lemma applied to the

ISBN: 978-989-95079-6-8

overvaluing matrix obtained by the use of the n regular vector norm $p(x)$ with $x=[x_1,x_2,...,x_n]^T$, such as: $p(x)=[|x_1|,|x_2|,...,|x_n|]^T$, leads to the following stability conditions of initial system:

$$a_{1,1}<0,\begin{vmatrix} a_{1,1} & |a_{1,2}| \\ |a_{2,1}| & a_{2,2} \end{vmatrix}>0,...,(-1)^n\begin{vmatrix} a_{1,1} & |a_{1,2}| & \dots & |a_{1,n}| \\ |a_{2,1}| & a_{2,2} & \dots & |a_{2,n}| \\ \vdots & \vdots & & \vdots \\ |a_{n,1}(.)| & |a_{n,2}(.)| & \dots & a_{n,n}(.) \end{vmatrix}>0$$

The Borne-Gentina practical criterion applied to continuous systems generalizes the Kotelyanski lemma for nonlinear systems and defines large classes of systems for which the linear Aizerman conjecture can be applied, either for the initial system or for its comparison system.

References

[1] J. Aracil, A. Garcia-Cerezo, A. Barreiro and A. Ollero , *Fuzzy Control of dynamical systems and stability analysis based on the conicity criterion*, IFSA World Congress, Brussels, Belgium, 1991.

[2] M. Benrejeb, M. Gasmi, *Stability and Stabilization Conditions of TSK Models using Vector Norms*, International Conference on Communication, Computer and Power (ICCCP'01), Muscat, Février 2001.

[3] M. Benrejeb, P. Borne, F. Laurent, *Sur une application de la représentation en flèche à l'analyse des processus*, RAIRO Automatique, 16 (2), pp 133-146, 1982.

[4] M. Benrejeb, A. Sakly, K. Ben Othman and P. Borne, *Choice of conjunctive operator of TSK fuzzy systems and stability domain study*, Math. And Comp. in Simul., Vol 76, pp 410-421, 2008.

[5] M. Benrejeb, D. Soudani, A. Sakly and P. Borne, *New discrete TSK fuzzy systems characterization and stability domain*, Int. Jour. Of Comp. Comm. and Cont., Vol 1, N 4, pp 9-19, 2006.

[6] P. Borne, *Non linear sytem stability*, vector norm approach, System and control Enceclopedia, t.5, 3402-3406, Perganon Press, 1987.

[7] P. Borne, M. Benrejeb, *On the Representation and the Stability Study of Large Scale Systems*, IJCCC, International Journal of Computers, Communications & Control, Vol 3, Supp. Issue: Proceedings of ICCC, pp 55-66, 2008.

[8] F. Bouslama and A. Ishikawa, *Application of limit fuzzy controllers to stability analysis*, Fuzzy Sets and Systems, Vol 49, pp 103-120, 1992.

[9] G. Calcev, *Some remarks on the stability of Mamdani fuzzy control systems*, IEEE Trans. on Fuzzy Systems, Vol 6, Issue 3, pp 436-442, 1998.

[10] G. Calcev, R. Gorez and M.D. Neyer, *Passivity approach of fuzzy control systems*, Automatica, Vol 34, N 3, pp 339-344, 1998.

[11] G. Chen and H. Ying, *Stability analysis of nonlinear fuzzy PI control systems*, Int. Conf. Fuzzy Logic Applict., Houston, TX, 1993.

[12] Y.Y. Chen, *Stability analysis of fuzzy control-A Lyapunov approach*, IEEE Conf. Syst. Man Cybern., Las Vegas, NV, 1987.

[13] H. Feng-Hsiag, H. Jing-Dong, *Stability analysis of fuzzy large-scale systems*, IEEE Trans. on Systems, Man, and Cybernetics, Part B, Vol 32, Issue 1, pp 122-126, 2002.

[14] C.-C. Fuh, P.-C. Tung, *Robust Stability analysis of fuzzy control System*, Fuzzy Sets and Systems, pp 289-298, 1997.

[15] J.C. Gentina, P. Borne, F. Laurent, "Stabilité des systèmes continus non linéaires de grande dimension, " *RAIRO*, Août, J-3, pp 69-77, 1972.

[16] R. Katoh, T. Yamashita and S. Singh, *Stability Analysis of control system having PD type of fuzzy controller*, Fuzzy Sets and Systems, Vol 74, Issue 2, pp 321-334, 1995.

[17] W.M.J. Kickert and E.H. Mamdani, *Analyses of Fuzzy Logic Controllers*, Fuzzy Sets and Systems, Vol.1, pp 29-44, 1978.

[18] J. Ch. Lo, Y. M. Chen, *Stability issues on Takagi-Sugeno fuzzy model-parametric approach*, IEEE Trans. on Fuzzy Systems, Vol 7 Issue 5, pp 597-607, 1999.

[19] K.S. Ray and D.D. Majumder, *Application of Cercle Criteria for Stability Analysis of SISO and MIMO Systems associated with Fuzzy Logic Controllers*, IEEE Trans. Syst. Man & Cyber., Vol 14, Issue 2, pp 345-349, 1984.

[20] A. Sakly, K. Ben Othman and M. Benrejeb, *Mamdani Type Fuzzy Controllers for the Stabilisation of Linear Systems*, WSEAS Transactions on Systems, Issue 4, Volume 2, pp.820-826, 2003.

[21] A. Sakly, B. Zahra and M. Benrejeb, *Stability Study of Mamdani's Fuzzy Controllers applied to Linear Plants*, Studies in Informatics and Control, SIC Journal , vol 17, Issue 4, December 2008.

[22] M. Sugeno , *On stability of fuzzy systems expressed by fuzzy rules with singleton consequences*, IEEE Trans. on Fuzzy Systems, Vol 7, Issue 2, pp 201-223, 1999.

[23] K. Tanaka, M. Sano, *A Robust Stabilization Problem of Fuzzy Control Systems and its Application to Backing up Control of a Track-Trailer*, IEEE Trans. on Fuzzy Systems, Vol 2, Issue 2, pp 119-134, 1994.

[24] L. Wang and R. Langari, *Fuzzy controller design via hyperstability approach*, IEEE Int. Conf. Fuzzy Syst., Orlando, FL, 1994.

[25] H. Ying, *Conditions on general Mamdani fuzzy controllers as nonlinear, variable gain state feedback controllers with stability analysis*, IFSA World Congress and 20th NAFIPS International Conference, 2001.

ISBN: 978-989-95079-6-8

The Fuzzy Induced Generalized OWA Operator and its Application in Business Decision Making

José M. Merigó, Anna M. Gil-Lafuente

Department of Business Administration, University of Barcelona
Av. Diagonal 690, 08034 Barcelona, Spain
Email: jmerigo@ub.edu, amgil@ub.edu

Abstract—*We present the fuzzy induced generalized OWA (FIGOWA) operator. It is an aggregation operator that uses the main characteristics of the fuzzy OWA (FOWA) operator, the induced OWA (IOWA) operator and the generalized OWA (GOWA) operator. Therefore, it uses uncertain information represented in the form of fuzzy numbers, generalized means and order inducing variables. The main advantage of this operator is that it includes a wide range of mean operators in the same formulation such as the FOWA, the IOWA, the GOWA, the induced GOWA, the fuzzy IOWA, the fuzzy generalized mean, etc. We study some of its main properties. A further generalization by using quasi-arithmetic means is also presented. This operator is called Quasi-FIOWA operator. We also develop an application of the new approach in a strategic decision making problem.*

Keywords— Decision making; OWA operator; Aggregation operators; Fuzzy numbers.

1 Introduction

Different types of aggregation operators are found in the literature for aggregating the information. A very common aggregation method is the ordered weighted averaging (OWA) operator [17]. Since its appearance, the OWA operator has been studied in a wide range of applications [1-3,5,7-8,10-13,15-22]. In [21], Yager and Filev introduced the IOWA operator. It is a generalization of the OWA operator that uses order inducing variables in the reordering of the arguments. In the last years, the IOWA operator has been studied by different authors [5,11-13,16,19,21].

When using the IOWA operator, it is assumed that the available information is exact numbers or crisp values. However, this may not be the real situation found in the decision making problem. Sometimes, the available information is vague or imprecise and it is not possible to analyze it with exact numbers. Then, it is necessary to use another approach to deal with this information such as fuzzy numbers (FN). For these situations, the IOWA is known as fuzzy number induced OWA (FN-IOWA) operator [5].

Recently, [13] have suggested a generalization of the IOWA operator by using generalized means. With this generalization, known as the induced generalized OWA (IGOWA) operator, we are able to include in the same formulation different types of induced aggregation operators such as the IOWA operator, the induced ordered weighted geometric (IOWG) operator and the induced ordered weighted quadratic averaging (IOWQA) operator, among others. Moreover, they also suggested the Quasi-IOWA operator which is a further generalization of the IGOWA operator by using quasi-arithmetic means.

Going a step further, in this paper we present the fuzzy induced generalized OWA operator which generalizes the FN-IOWA by using generalized means. We will call it the fuzzy induced generalized OWA (FIGOWA) operator. Then, we are able to obtain a wide range of fuzzy induced aggregation operators such as the FN-IOWA, the FN-IOWG operator and the FN-IOWQA operator, among others. We study some of the main properties of this generalization and we extend it to a more general formulation by using quasi-arithmetic means. The result is the Quasi-FIOWA operator. We also develop an application of the new approach in a decision making problem about selection of strategies.

This paper is organized as follows. In Section 2, we briefly review some basic concepts such as FN, the FN-IOWA and the IGOWA operator. Section 3 presents the FIGOWA operator and Section 4 studies some of its families. In Section 5 we briefly present the Quasi-FIOWA operator and in Section 6, we develop an application of the new approach in a strategic decision making problem.

2 Preliminaries

2.1 Fuzzy numbers

The FN was first introduced by [4,24]. Since then, it has been studied and applied by a lot of authors such as [6,9-11].

A FN is a fuzzy subset [23] of a universe of discourse that is both convex and normal [9]. Note that the FN may be considered as a generalization of the interval number [14] although it is not strictly the same because the interval numbers may have different meanings.

In the literature, we find a wide range of FNs [6,9]. For example, a trapezoidal FN (TpFN) A of a universe of discourse R can be characterized by a trapezoidal membership function $A = (\underline{a}, \overline{a})$ such that

$$\begin{aligned} \underline{a}(\alpha) &= a_1 + \alpha(a_2 - a_1), \\ \overline{a}(\alpha) &= a_4 - \alpha(a_4 - a_3). \end{aligned} \tag{1}$$

where $\alpha \in [0, 1]$ and parameterized by (a_1, a_2, a_3, a_4) where $a_1 \le a_2 \le a_3 \le a_4$, are real values. Note that if $a_1 = a_2 = a_3 = a_4$, then, the FN is a crisp value and if $a_2 = a_3$, the FN is

ISBN: 978-989-95079-6-8

represented by a triangular FN (TFN). Note that the TFN can be parameterized by (a_1, a_2, a_4).
In the following, we are going to review the FN arithmetic operations as follows. Let A and B be two TFN, where $A = (a_1, a_2, a_3)$ and $B = (b_1, b_2, b_3)$. Then:

1) $A + B = (a_1 + b_1, a_2 + b_2, a_3 + b_3)$
2) $A - B = (a_1 - b_3, a_2 - b_2, a_3 - b_1)$
3) $A \times k = (k \times a_1, k \times a_2, k \times a_3)$; for $k > 0$.

Note that other operations could be studied [6,9] but in this paper we will focus on these ones.

2.2 Fuzzy induced OWA operator

The FIOWA (or FN-IOWA) operator was introduced by [5]. It is an aggregation operator that uses uncertain information represented by FNs. It also uses a reordering process different from the values of the arguments. In this case, the reordering step is based on order inducing variables. It can be defined as follows.

Definition 1. *Let Ψ be the set of FN. A FIOWA operator of dimension n is a mapping FIOWA: $\Psi^n \rightarrow \Psi$ 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*:

$$FIOWA(\langle u_1,\tilde{a}_1\rangle, \ldots, \langle u_n,\tilde{a}_n\rangle) = \sum_{j=1}^{n} w_j b_j \quad (2)$$

where b_j is the $\tilde{a}_i$ value of the FIOWA pair $\langle u_i, \tilde{a}_i\rangle$ having the jth largest u_i, u_i is the order inducing variable and $\tilde{a}_i$ is the argument variable represented in the form FN.
Note that from a generalized perspective of the reordering step it is possible to distinguish between descending (DFIOWA) and ascending (AFIOWA) orders. Note also that this operator provides a parameterized family of aggregation operators that includes the fuzzy maximum, the fuzzy minimum and the fuzzy average (FA), among others.

2.3 Induced generalized OWA operator

The IGOWA operator was introduced in [13] and it represents a generalization of the IOWA operator by using generalized means. Then, it is possible to include in the same formulation, different types of induced operators such as the IOWA operator or the induced OWG (IOWG) operator. It can be defined as follows.

Definition 2. *An IGOWA operator of dimension n is a mapping IGOWA: $R^n \rightarrow R$ 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*:

$$IGOWA(\langle u_1,a_1\rangle,\ldots,\langle u_n,a_n\rangle) = \left(\sum_{j=1}^{n} w_j b_j^{\lambda}\right)^{1/\lambda} \quad (3)$$

where b_j is the a_i value of the IGOWA pair $\langle u_i,a_i\rangle$ having the jth largest u_i, u_i is the order inducing variable, a_i is the argument variable and λ is a parameter such that $\lambda \in (-\infty, \infty)$.
As we can see, if $\lambda = 1$, we get the IOWA operator. If $\lambda = 0$, the IOWG operator and if $\lambda = 2$, the IOWQA operator. Note that it is possible to further generalize the IGOWA operator by using quasi-arithmetic means. The result is the Quasi-IOWA operator.

3 Fuzzy induced generalized OWA operator

The fuzzy induced generalized OWA (FIGOWA) operator is an extension of the GOWA operator that uses uncertain information in the aggregation represented in the form of FNs. The reason for using this operator is that sometimes, the uncertain factors that affect our decisions are not clearly known and in order to assess the problem we need to use FNs. The FN is a very useful technique in decision making because it considers the different uncertain results that could happen in the future. This operator also uses a reordering process based on order inducing variables. It can be defined as follows.

Definition 3. *Let Ψ be the set of FNs. A FIGOWA operator of dimension n is a mapping FIGOWA: $\Psi^n \rightarrow \Psi$ 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*:

$$FIGOWA(\langle u_1,\tilde{a}_1\rangle,\ldots,\langle u_n,\tilde{a}_n\rangle) = \left(\sum_{j=1}^{n} w_j b_j^{\lambda}\right)^{1/\lambda} \quad (4)$$

where b_j is the $\tilde{a}_i$ value of the FIGOWA pair $\langle u_i,\tilde{a}_i\rangle$ having the jth largest u_i, u_i is the order inducing variable, $\tilde{a}_i$ is the argument variable represented in the form of FN and λ is a parameter such that $\lambda \in (-\infty, \infty)$.
Note that different types of FNs could be used in the aggregation such as TFNs, TpFNs, L-R FNs, interval-valued FNs, intuitionistic FNs, etc.
As it was explained in [13], when using FN in the OWA operator, we have the additional problem of how to reorder the arguments. In the FIGOWA operator, this is not a problem because the reordering process is developed with order inducing variables and it is independent of the values of the arguments.
From a generalized perspective of the reordering step, it is possible to distinguish between the descending FIGOWA (DFIGOWA) and the ascending FIGOWA (AFIGOWA) operator. The weights of these operators are related by $w_j = w^*_{n-j+1}$, where w_j is the jth weight of the DFIGOWA and w^*_{n-j+1} the jth weight of the AFIGOWA 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 FIGOWA operator can be expressed as:

ISBN: 978-989-95079-6-8

$$FIGOWA(\langle u_1,\tilde{a}_1\rangle,\ldots,\langle u_n,\tilde{a}_n\rangle) = W^T B \quad (5)$$

Note that if the weighting vector is not normalized, i.e., $W = \sum_{j=1}^{n} w_j \neq 1$, then, the FIGOWA operator can be expressed as:

$$FIGOWA(\langle u_1,\tilde{a}_1\rangle,\ldots,\langle u_n,\tilde{a}_n\rangle) = \frac{1}{W}\sum_{j=1}^{n} w_j b_j \quad (6)$$

The FIGOWA operator is monotonic, commutative, bounded and idempotent.

Theorem 1 (Monotonicity). *Assume f is the FIGOWA operator, if $\tilde{a}_i \geq \tilde{e}_i$, for all $\tilde{a}_i$, then*

$$f(\langle u_1,\tilde{a}_1\rangle,\ldots,\langle u_n,\tilde{a}_n\rangle) \geq f(\langle u_1,\tilde{e}_1\rangle,\ldots,\langle u_n,\tilde{e}_n\rangle) \quad (7)$$

Theorem 2 (Commutativity). *Assume f is the FIGOWA operator, then*

$$f(\langle u_1,\tilde{a}_1\rangle,\ldots,\langle u_n,\tilde{a}_n\rangle) = f(\langle u_1,\tilde{e}_1\rangle,\ldots,\langle u_n,\tilde{e}_n\rangle) \quad (8)$$

where $(\langle u_1,\tilde{e}_1\rangle,\ldots,\langle u_n,\tilde{e}_n\rangle)$ is any permutation of the arguments $(\langle u_1,\tilde{a}_1\rangle,\ldots,\langle u_n,\tilde{a}_n\rangle)$.

Theorem 3 (Boundedness). *Assume f is the FIGOWA operator, then*

$$\min\{\tilde{a}_i\} \leq f(\langle u_1,\tilde{a}_1\rangle,\ldots,\langle u_n,\tilde{a}_n\rangle) \leq \max\{\tilde{a}_i\} \quad (9)$$

Theorem 4 (Idempotency). *Assume f is the FIGOWA operator, if $\tilde{a}_i = \tilde{a}$, for all $\tilde{a}_i$, then*

$$f(\langle u_1,\tilde{a}_1\rangle,\ldots,\langle u_n,\tilde{a}_n\rangle) = \tilde{a} \quad (10)$$

Note that the proofs of Theorems 1 - 4 are omitted because they are trivial.

Another interesting issue when analysing the FIGOWA operator is the problem of ties in the order inducing variables. In order to solve this problem, we recommend to follow the policy explained in [21]. Basically, the idea is to replace each argument of the tied inducing variables by its fuzzy generalized mean. Then, different types of means may be used to replace the arguments depending on the parameter λ.

As it is explained in [21], different kinds of attributes may be used for the order inducing variables of the FIGOWA operator with the only requirement of having a linear ordering.

4 Families of FIGOWA operators

Basically, we can distinguish between two main groups of FIGOWA operators. The first family represents all the families that may be found in the weighting vector W, while the second family represents all the particular cases coming from the parameter λ.

4.1 Analysing the parameter λ

If we analyze different values of the parameter λ, we obtain another group of particular cases such as the FIOWA operator, the fuzzy IOWG (FIOWG), the fuzzy IOWQA (FIOWQA) and the fuzzy induced ordered weighted harmonic averaging (FIOWHA) operator.

When $\lambda = 1$, the FIGOWA operator becomes the FIOWA operator.

$$FIOWA(\langle u_1,\tilde{a}_1\rangle, \ldots, \langle u_n,\tilde{a}_n\rangle) = \sum_{j=1}^{n} w_j b_j \quad (11)$$

Note that it is possible to study a wide range of families of FIOWA operators by using different weighting vectors in a similar way as it has been explained in Section 4.1. For example, if $w_j = 1/n$, for all $\tilde{a}_i$, we get the FA and if the ordered position of b_j is the same than the position of the values u_i, we get the FOWA operator.

When $\lambda = 0$, we get the FIOWG operator.

$$FIOWG(\langle u_1,\tilde{a}_1\rangle, \ldots, \langle u_n,\tilde{a}_n\rangle) = \prod_{j=1}^{n} b_j^{w_j} \quad (12)$$

Note that in this case we can also study different families of FIOWG operators such as the fuzzy geometric mean or the fuzzy OWG (FOWG) operator. Note also that it is possible to distinguish between descending (DFIOWG) and ascending (AFIOWG) orders.

When $\lambda = -1$, we get the FIOWHA operator.

$$FIOWHA(\langle u_1,\tilde{a}_1\rangle, \ldots, \langle u_n,\tilde{a}_n\rangle) = \frac{1}{\sum_{j=1}^{n} \frac{w_j}{b_j}} \quad (13)$$

From a generalized perspective of the reordering step we find the descending FIOWHA (DFIOWHA) and the ascending FIOWHA (AFIOWHA) operator. Different families of FIOWHA operators are found by using different weighting vectors such as the fuzzy harmonic mean and the fuzzy ordered weighted harmonic averaging (FOWHA) operator.

When $\lambda = 2$, we get the FIOWQA operator.

$$FIOWQA(\langle u_1,\tilde{a}_1\rangle,\ldots,\langle u_n,\tilde{a}_n\rangle) = \left(\sum_{j=1}^{n} w_j b_j^2\right)^{1/2} \quad (14)$$

In this case, we can also study a wide range of families of FIOWQA operators such as the fuzzy quadratic mean and the fuzzy OWQA operator, and distinguish between the DFIOWQA and the AFIOWQA operator.

ISBN: 978-989-95079-6-8

4.2 Analysing the weighting vector W

By using a different weighting vector in the FIGOWA operator, we are able to obtain a wide range of aggregation operators. For example, we can obtain the fuzzy maximum, the fuzzy minimum, the FGM, the fuzzy weighted generalized mean (FWGM) and the FGOWA operator.

Remark 1. The fuzzy maximum is obtained if $w_p = 1$ and $w_j = 0$, for all $j \neq p$, and $u_p = \text{Max}\{\tilde{a}_i\}$. The fuzzy minimum is obtained if $w_p = 1$ and $w_j = 0$, for all $j \neq p$, and $u_p = \text{Min}\{\tilde{a}_i\}$. The FGM is found when $w_j = 1/n$, for all $\tilde{a}_i$. The fuzzy weighted generalized mean (FWGM) is obtained if $u_i > u_{i+1}$, for all i, and the FGOWA operator is obtained if the ordered position of u_i is the same than the ordered position of b_j such that b_j is the jth largest of $\tilde{a}_i$.

Remark 2. Other families of FIGOWA operators could be used in the aggregation by using a different manifestation of the weighting vector. For example, we could analyze the step-FIGOWA, the window-FIGOWA, the median-FIGOWA, the olympic-FIGOWA, the centered-FIGOWA, the S-FIGOWA, etc. For more information, see [10-13,17-22].

Remark 3. The step-FIGOWA operator is found when $w_k = 1$ and $w_j = 0$, for all $j \neq k$ and the window-FIGOWA when $w_j = 1/m$ for $k \leq j \leq k + m - 1$ and $w_j = 0$ for $j > k + m$ and $j < k$. Note that k and m must be positive integers such that $k + m - 1 \leq n$.

Remark 4. For the median-FIGOWA, we distinguish between two cases. If n is odd we assign $w_{(n+1)/2} = 1$ and $w_j = 0$ for all others, and this affects the argument $\tilde{a}_i$ with the [$(n + 1)/2$]th largest u_i. If n is even we assign, for example, $w_{n/2} = w_{(n/2)+1} = 0.5$, and this affects the arguments with the $(n/2)$th and [$(n/2) + 1$]th largest u_i.

Remark 5. The olympic-FIGOWA operator is found if $w_1 = w_n = 0$, and for all others $w_j = 1/(n - 2)$. Note that it is possible 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.

Remark 6. A further family is the centered-FIGOWA operator. This type of aggregation operator is symmetric, strongly decaying and inclusive. 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 $w_i < w_j$ which is known as softly decaying centered-FIGOWA operator. Note also the possibility of removing the third condition. Then, we shall refer to this type of aggregation as non-inclusive centered-FIGOWA operator.

Remark 7. A further interesting family is the S-FIGOWA operator. In this case, we can distinguish between three types: the "orlike", the "andlike", and the "generalized" S-FIGOWA operator. The generalized S-FIGOWA operator is obtained when $w_p = (1/n)(1 - (\alpha + \beta)) + \alpha$, with $u_p = \text{Max}\{\tilde{a}_i\}$; $w_q = (1/n)(1 - (\alpha + \beta)) + \beta$, with $u_q = \text{Min}\{\tilde{a}_i\}$; and $w_j = (1/n)(1 - (\alpha + \beta))$ for all $j \neq p,q$ where $\alpha, \beta \in [0, 1]$ and $\alpha + \beta \leq 1$. Note that if $\alpha = 0$, we get the andlike S-FIGOWA and if $\beta = 0$, the orlike S-FIGOWA.

Remark 8. Another type is the non-monotonic-FIGOWA 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.

5 The Quasi-FIOWA operator

The FIGOWA operator may be further generalized by using quasi-arithmetic means. Then, the result is the fuzzy induced ordered weighted quasi-arithmetic averaging operator or Quasi-FIOWA, for short. Note that the Quasi-FIOWA operator is an extension of the Quasi-OWA [1,3,7,11] by using order inducing variables and uncertain information represented with FNs.

Definition 4. *Let Ψ be the set of FNs. A Quasi-FIOWA operator of dimension n is a mapping $f: \Psi^n \rightarrow \Psi$ 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*:

$$f(\langle u_1,\tilde{a}_1\rangle,\ldots,\langle u_n,\tilde{a}_n\rangle) = g^{-1}\left(\sum_{j=1}^{n} w_j g(b_j)\right) \quad (15)$$

where b_j is the $\tilde{a}_i$ value of the Quasi-FIOWA pair $\langle u_i,\tilde{a}_i\rangle$ having the jth largest u_i, u_i is the order inducing variable, $\tilde{a}_i$ is the argument variable represented in the form of FN and $g(b)$ is a strictly continuous monotone function.

As we can see, when $g(b) = b^{\lambda}$, we get the FIGOWA operator. Note that it is also possible to distinguish between descending (Quasi-DFIOWA) and ascending (Quasi-AFIOWA) orders. Note also that all the properties and particular cases commented in the FIGOWA operator, are also applicable in this case.

6 Application in strategic decision making

In the following, we are going to develop a brief example where we will see the applicability of the new approach. We will focus in a decision making problem about selection of strategies. Note that other business decision making applications could be developed such as financial decision making, human resource selection, etc. Note that the FIGOWA operator may be applied in similar problems than the IOWA and the IGOWA operator.

ISBN: 978-989-95079-6-8

Assume a company that operates in North America and Europe is analyzing the general policy for the next year and they consider 5 possible strategies to follow.

1) A_1 = Expand to the Asian market.
2) A_2 = Expand to the South American market.
3) A_3 = Expand to the African market.
4) A_4 = Expand to the 3 continents.
5) A_5 = Do not develop any expansion.

In order to evaluate these strategies, the group of experts considers that the key factor is the economic situation of the next year. Thus, depending on the situation, the expected benefits will be different. The experts have considered 5 possible situations for the next year: S_1 = Very bad, S_2 = Bad, S_3 = Regular, S_4 = Good, S_5 = Very good. The expected results depending on the situation S_i and the alternative A_i are shown in Table 1. Note that the results are TFN.

Table 1: Available strategies

	S_1	S_2	S_3	S_4	S_5
A_1	(20,30,40)	(60,70,80)	(40,50,60)	(50,60,70)	(50,60,70)
A_2	(30,40,50)	(70,80,90)	(30,40,50)	(30,40,50)	(50,60,70)
A_3	(60,70,80)	(50,60,70)	(40,50,60)	(20,30,40)	(40,50,60)
A_4	(50,60,70)	(30,40,50)	(60,70,80)	(70,80,90)	(10,20,30)
A_5	(40,50,60)	(30,40,50)	(50,60,70)	(60,70,80)	(40,50,60)

In this problem, the experts consider the weighting vector W = (0.1, 0.2, 0.2, 0.2, 0.3). Due to the fact that the attitudinal character is very complex because it involves the opinion of different members of the board of directors, the experts use order inducing variables to express it.

Table 2: Order inducing variables

	S_1	S_2	S_3	S_4	S_5
A_1	7	9	6	5	8
A_2	4	3	6	8	7
A_3	2	8	4	3	6
A_4	5	6	9	2	7
A_5	8	4	3	6	5

With this information, we can aggregate it in order to take a decision. In Table 3 and 4, we show the different results obtained by using different types of FIGOWA operators in the decision process.

Table 3: First aggregation process

	FA	FWA	FOWA
A_1	(44,54,64)	(47,57,67)	(40,50,60)
A_2	(42,52,62)	(44,54,64)	(38,48,58)
A_3	(42,52,62)	(40,50,60)	(38,48,58)
A_4	(44,54,64)	(40,50,60)	(38,48,58)
A_5	(44,54,64)	(44,54,64)	(41,51,61)

Table 4: First aggregation process

	FIOWA	FIOWG	FIOWQA
A_1	(43,53,63)	(40.5,51.1,61.5)	(44.8,54.4,64.2)
A_2	(46,56,66)	(42.8,53.4,63.7)	(49.1,58.6,68.2)
A_3	(43,53,63)	(40.2,50.8,61.2)	(45.2,54.8,64.5)
A_4	(45,55,65)	(36.8,49.1,60.3)	(50.2,59.4,68.7)
A_5	(45,55,65)	(43.7,54.0,64.1)	(46.1,55.9,65.8)

If we establish an ordering of the alternatives, we get the following results shown in Table 5. Note that in this example it is not necessary to establish a criterion for ranking FNs because it is clear which alternative goes first, second and so on, in the ordering process. Note also that "$\succ$" means "preferred to" and "=" means "equal to".

Table 5: Ordering of the investments

	Ordering
FA	$A_1=A_4=A_5 \succ A_2=A_3$
FWA	$A_1 \succ A_2=A_5 \succ A_3=A_4$
FOWA	$A_5 \succ A_1 \succ A_2=A_3=A_4$
FIOWA	$A_2 \succ A_4=A_5 \succ A_1=A_3$
FIOWG	$A_5 \succ A_2 \succ A_1 \succ A_3 \succ A_4$
FIOWQA	$A_4 \succ A_2 \succ A_5 \succ A_3 \succ A_1$

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. For example, the FA gives very similar results between alternatives while the FIOWG and the FIOWQA give more differences between the alternatives.

7 Conclusions

We have presented the FIGOWA operator. It is a generalization of the OWA operator that uses the main characteristics of three well known aggregation operators: the GOWA, the IOWA and the FOWA operator. That is to say, it uses generalized means, order inducing variables and FNs in the aggregation. We have studied some of the main properties of this new aggregation operator. We have further generalized it by using quasi-arithmetic means. Then, we have obtained the Quasi-FIOWA operator. By using this approach, we get a more complete representation of the information because we are using FNs that consider the best and worst result and the possibility that the internal results will occur.

We have also developed an application of the new approach. We have focused in a strategic decision making problem and we have seen that depending on the particular FIGOWA operator used, the results and the decisions may be different.

In future research, we expect to develop further extensions by adding new characteristics [11] in the problem and applying it to other business problems such as financial decision making and human resource 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] S.S.L. Chang and L.A. Zadeh, On fuzzy mapping and control. *IEEE Transactions on Systems, Man and Cybernetics*, 2:30-34, 1972.

[5] S.J. Chen and S.M. Chen, A new method for handling multi-criteria fuzzy decision making problems using FN-IOWA operators. *Cybernetics and Systems*, 34:109-137, 2003.

[6] D. Dubois and H. Prade, *Fuzzy Sets and Systems: Theory and Applications*. Academic Press, New York, 1980.

[7] J. Fodor, J.L. Marichal and M. Roubens, Characterization of the ordered weighted averaging operators. *IEEE Transactions on Fuzzy Systems*, 3:236-240, 1995.

[8] N. Karayiannis, Soft Learning Vector Quantization and Clustering Algorithms Based on Ordered Weighted Aggregation Operators. *IEEE Transactions on Neural Networks*, 11:1093-1105, 2000.

[9] A. Kaufmann, M.M. Gupta, *Introduction to fuzzy arithmetic*. Publications Van Nostrand, Rheinhold, 1985.

[10] J.M. Merigó, Using immediate probabilities in fuzzy decision making. In: *Proceedings of the 22nd ASEPELT Conference*, Barcelona, pp. 1650-1664, 2008.

[11] 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.

[12] J.M. Merigó and M. Casanovas, Decision making with distance measures and induced aggregation operators. In: *Proceedings of the 8th FLINS International Conference*, Madrid, pp. 483-488, 2008.

[13] J.M. Merigó and A.M. Gil-Lafuente, The induced generalized OWA operator. *Information Sciences*, 179:729-741, 2009.

[14] R.E. Moore, *Interval Analysis*. Prentice Hall, Englewood Cliffs, NJ, 1966.

[15] V. Torra and Y. Narukawa, *Modeling Decisions: Information Fusion and Aggregation Operators*. Berlin: Springer-Verlag, 2007.

[16] Z.S. Xu and Q.L. Da, An overview of operators for aggregating information. *International Journal of Intelligent Systems*, 18:953-968, 2003

[17] 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.

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

[19] R.R. Yager, Induced aggregation operators. *Fuzzy Sets and Systems*, 137:59-69, 2003.

[20] R.R. Yager, Generalized OWA Aggregation Operators. *Fuzzy Optimization and Decision Making*, 3:93-107, 2004.

[21] R.R. Yager and D.P. Filev, Induced ordered weighted averaging operators. *IEEE Transactions on Systems, Man and Cybernetics B*, 29:141-150, 1999.

[22] R.R. Yager and J. Kacprzyk, *The Ordered Weighted Averaging Operators: Theory and Applications*. Norwell: Kluwer Academic Publishers, 1997.

[23] L.A. Zadeh, Fuzzy sets. *Information and Control*, 8:338-353, 1965.

[24] L.A. Zadeh, The Concept of a Linguistic Variable and its application to Approximate Reasoning. Part 1. *Information Sciences*, 8:199-249, 1975; Part 2. *Information Sciences*, 8:301-357, 1975; Part 3. *Information Sciences*, 9:43-80, 1975.

ISBN: 978-989-95079-6-8

Solution for Fuzzy Road Transport Traveling Salesman Problem Using Eugenic Bacterial Memetic Algorithm

János Botzheim[1] Péter Földesi[2] László T. Kóczy[3]

1 Department of Automation, Széchenyi István University
H-9026, Győr, Egyetem tér 1. Hungary
2 Department of Logistics and Forwarding, Széchenyi István University
H-9026, Győr, Egyetem tér 1. Hungary
3 Department of Telecommunications, Széchenyi István University
H-9026, Győr, Egyetem tér 1. Hungary
Email: botzheim@sze.hu, foldesi@sze.hu, koczy@sze.hu

Abstract— *The aim of the Traveling Salesman Problem (TSP) is to find the cheapest way of visiting all elements in a given set of cities and returning to the starting point. In solutions presented in the literature costs of travel between nodes (cities) are based on Euclidean distances, the problem is symmetric and the costs are constant. In this paper a novel construction and formulation of the TSP is presented in which the requirements and features of practical application in road transportation and supply chains are taken into consideration. Computational results are presented as well.*

Keywords— Traveling Salesman Problem, time dependent fuzzy costs, eugenic bacterial memetic algorithm

1 Introduction

The aim of the Traveling Salesman Problem (TSP) is to find the cheapest way of visiting all elements in a given set of cities where the cost of travel between each pair of them is given, including the return to the starting point. The TSP is a very good representative of a larger class of problems known as combinatorial optimization problems. If an efficient algorithm (i.e., an algorithm that will guarantee to find the optimal solution in polynomial number of steps) can be found for the traveling salesman problem, then efficient algorithms could be established for all other problems in the NP (nondeterministic polynomial time) class, thus TSP belongs to the so called NP-hard complexity class. TSP is a well studied NP-hard problem [5]. An algorithm can be considered as an effective (good) one if it has a polynomial function of the problem size n, that is, for large values of n, the algorithm runs in time at most Kn^c for some constant number K and c. The question whether or not there is a good algorithm for the TSP has not been settled. For its practical importance and wide range of application in practice [5] many approaches, heuristic searches and algorithms have been suggested [8, 10, 11, 12], while different extensions and variations of the original TSP have been investigated [6, 9, 13].

Solutions presented in the literature most frequently have the following features. Costs of travel between nodes (cities) are based on Euclidean distances, the problem is symmetric, meaning that the cost from $node_i$ to $node_j$ equals to the cost from $node_j$ to $node_i$, and the costs are constant. In this paper a novel construction and formulation of the TSP is presented in which the requirements and features of practical application in road transportation and supply chains are taken into consideration. Since the original formulation of the problem states: the aim is to find the "cheapest" tour, thus the cost matrix that represents the distances between each pair must be determined by calculating the actual costs of transportation processes. The costs of transportation consist of two main elements: costs proportional to transit distances (km) and costs proportional to transit times. Obviously the physical distances can be considered as constant values in a given relation, but transit times are subject to external factors, such as weather conditions, traffic circumstances, etc., so they should be treated as a time-dependent variable. Furthermore the actual costs are rarely constant and predictable, so fuzzy cost coefficient can be applied in order to represent the uncertainty [15, 17]. On the other hand in real road networks the actual distance between two points often alter from the Euclidean distance, furthermore occasionally some extra costs (e.g., ferriage, tunnel fare) can modify the distance-related variable costs. Considering these characteristics the original TSP should be reconstructed, so that realistic solutions can be developed.

For solving the above-mentioned fuzzy road-transport TSP (FRTTSP) in this paper we suggest a eugenic bacterial memetic algorithm (EBMA) since that algorithm is suitable for global optimization of even non-linear, high-dimensional, multi-modal, and discontinuous problems. As numerical example a modified TSP (FRTTSP) instance is considered, in which the elements of cost matrix are dependent on the steps they are selected to carry on with.

2 Formulating and solutions for the classical TSP

In the case of the traveling salesman problem, the mathematical description can be a graph where each city is denoted by a point (or node) and lines are drawn connecting every two nodes (called arcs or edges). A distance (or cost) is associated with every edge. If in a graph edges are drawn connecting any two nodes, then the graph is said to be *complete*. A round-trip of the cities corresponds to a special subset of the lines when each city is visited exactly once, and it is called a tour or a Hamiltonian cycle in graph theory. The length of a tour is the sum of the lengths of the lines in the round-trip.

ISBN: 978-989-95079-6-8

Asymmetric and symmetric TSPs can be distinguished depending on if any edge of the graph is directed or not. To formulate the symmetric case with n nodes $c_{ij} = c_{ji}$, so a graph can be considered where there is only one arc (undirected) between every two nodes. Let $x_{ij} = \{0,1\}$ be the decision variable ($i= 1,2,...,n$ and $j=1,2,...,n$), and $x_{ij} = 1$, means that the arc connecting node$_i$ to node$_j$ is an element of the tour.

Let $x_{ii} = 0 \ (i= 1,2,...,n)$ (1)

meaning that no tour element is allowed from a node to itself. Furthermore

$$\sum_{i=1}^{n}\sum_{j=1}^{n} x_{ij} = n \quad (2)$$

that is the number of decision variables where $x_{ij} = 1$ is equal to n, and

$$\sum_{i=1}^{n} x_{ij} = 1 \qquad \forall j \in \{1,2,..., n\}, \quad (3)$$

$$\sum_{j=1}^{n} x_{ij} = 1 \qquad \forall i \in \{1,2,..., n\}, \quad (4)$$

meaning that each column and row of the decision matrix has a single element with a value 1 (i.e., each city is visited once). For assuring the close circuit, an additional constraint must be set. A permutation of nodes $(p_1,p_2,....p_n)$ has to be constructed so that the total cost $C(p)$ is minimal:

$$\text{minimize} \quad C(p) = \left(\sum_{i=1}^{n-1} c_{p_i, p_{i+1}} \right) + c_{p_n, p_1} . \quad (5)$$

For a symmetrical network there are $1/2\ (n-1)!$ possible tours (because the degree of freedom is $(n-1)$ and tours describing the same sequence but opposite directions are not considered as different tours) and for asymmetric networks where $c_{ij} \neq c_{ji}$ the number of possible tours is $(n-1)!$. Some reduction can be done [14], a guarantee that is proportional to $n^2 2^n$ can be given, however it is clear that an exhaustive search is not possible for large n in practice. Rather than enumerating all possibilities, successful algorithms for solving the TSP problem have been capable of eliminating most of the roundtrips without ever actually considering them. The main groups of engines are [5-12]:

- Mixed-integer programming
- Branch-and-bound method
- Heuristic searches (local search algorithms, simulated annealing, neural networks, genetic algorithms, particle swarm optimization, ant colony optimization, etc.)

In this paper a eugenic bacterial memetic algorithm is proposed as a novel heuristic search for TSP.

3 Variable costs and consequences of their nature

3.1 Costs in real road networks

Considering real road transport networks, especially in city logistics, the actual circumstances and condition of the transit process are subject to not only the topography of the given network but to timing as well. Referring to the phenomenon of cyclic peak-hours and also to the weekly (monthly, yearly) periodicity of traffic on road, it can be stated that the unit cost of traveling is also a variable, and it can be described as a time series rather than a constant value. The main reasons of that are the following:

- The actual cost of 1 km depends on the current fuel consumption, which is partly affected by the speed, but the speed is externally determined by the current traffic.
- A relatively large proportion of transportation cost is the cost of labor (i.e., wages of drivers), which is calculated on driving time. In Europe distance-based payment for commercial drivers are not allowed according to the European Agreement Concerning the Work of Crews of Vehicles Engaged in International Road Transport (AETR).
- The return on equity capital is a vital issue for haulage companies, since the transportation sector of the economy is a capital intensive one, meaning that utilization of vehicle in time is a crucial problem.

In addition, during long-distance shipments the drivers occasionally must stop for a rest (according to AETR) at a minimum 11 hour period (often overnight), and very often week-end traffic restrictions for heavy vehicles are introduced.

Considering the uncertainty of relevant data it can be stated, that one's estimated travel time by automobile between two points is a case of possibility due to the measurement imprecision and perception. It can be seen that the circumstances and conditions are significantly changing in time that is the actual value of cost matrix element c_{ij} should be subject to timing of transit between node$_i$ and node$_j$ and an appropriate representation of imprecision can be the use of fuzzy numbers. In this sense geographical optimization alone is not appropriate, and the road transport operation has to be scheduled in time as well.

3.2 The modified TSP, the Fuzzy Road Transport Traveling Salesman Problem (FRTTSP)

For fulfilling the requirements of realistic road transport processes, we propose the following modification to the classical TSP [23].

Since the overall target is to achieve the cheapest tour (in monetary terms), the constraint that each node (city) is visited exactly once is skipped. Calculating with time-dependent cost coefficients that are not necessarily proportional to distances, a longer route can be a cheaper one, and as a consequence some nodes can be visited more than once. Thus in the FRTTSP we eliminate restrictions (2), (3) and (4).

If a significant improvement in traffic conditions can be expected, that is a future value of a cost c_{ij} will be less then its present value, it is worth waiting (suspending the tour for a while) and continuing in the next step. In this case obviously the cost of staying at a point must be calculated. In this sense we eliminate restriction (1) as well. Very often in the solutions restriction (1) is fulfilled by selecting $c_{ii} = \infty$. In our case c_{ii} is the cost of staying at node$_i$ in a given step.

ISBN: 978-989-95079-6-8

The permutation of nodes (p_1,p_2,….p_n) is being modified as well. As a city may be visited several times, objective function (5) must be rewritten:

$$\text{minimize } C(p) = \left(\sum_{i=1}^{mn-1} c_{p_i, p_{i+1}} \right) + c_{p_{mn}, p_1}, \quad (6)$$

where *m=1,2,...* is the multiplier factor. Objective function (6) is a generalized form of TSP, the multiplier factor equals to *1* in the classical cases.

In order to represent the uncertainty triangular fuzzy numbers are used as cost coefficients. Triangular fuzzy numbers have a membership function consisting of two linear segments joined at a peak, so they can be constructed easily on the basis of little information: the supporting interval $C= [c_1, c_2]$ as the smallest and the largest possible values, and c_M which is the peak value where the membership function equals to *1*. In that case the triangular fuzzy number is denoted by $\boldsymbol{C} = (c_1, c_M, c_2)$.

When the distances between the cities are described by fuzzy numbers, it must be discussed how these fuzzy numbers are summed up in a tour in order to calculate the total distance. The arithmetic of fuzzy numbers is based on the extension principle [18]. We are using triangular shaped fuzzy numbers which can be characterized by three values, the boundaries of the support and the core value. When we calculate the total distance of a tour, then instead of adding fuzzy numbers by the extension principle, we can do an easier calculation based on the defuzzified values of the fuzzy numbers. According to [16], some defuzzification method has invariance properties meaning that the result is invariant under linear transformations, thus there is no need to determine the whole outcome using the extension principle but only to compute the sum of the defuzzified values of each fuzzy number. If we are using triangular shaped fuzzy numbers then the Averaging Level Cuts (ALC) type defuzzification method used in [16] gives the same result as the Center of Gravity (COG) method. So, in the first step the fuzzy numbers are defuzzified by the COG method (which is simply the arithmetic mean of the three characteristic points of the fuzzy number) and then these crisp numbers are summed up providing the total distance of the tour.

4 Eugenic bacterial memetic algorithms

Nature inspired some evolutionary optimization algorithms suitable for global optimization of even non-linear, high-dimensional, multi-modal, and discontinuous problems. The original genetic algorithm was developed by Holland [1] and was based on the process of evolution of biological organisms. It uses three operators: reproduction, crossover and mutation. Later, new kind of evolutionary based techniques were proposed, which are imitating phenomena that can be found in nature.

Bacterial Evolutionary Algorithm (BEA) [2] is one of these techniques. BEA uses two operators; the bacterial mutation and the gene transfer operation. These new operators are based on the microbial evolution phenomenon. Bacteria share chunks of their genes rather than perform a neat crossover in chromosomes. The bacterial mutation operation optimizes the chromosome of one bacterium; the gene transfer operation allows the transfer of information between the bacteria in the population. Each bacterium represents a solution for the original problem. BEA has been applied for wide range of problems, for instance optimizing the fuzzy rule bases [2, 3] or feature selection [4].

Evolutionary algorithms are global searchers, however in most of the cases they give only a quasi-optimal solution for the problem. Local search approaches can give more accurate solution, however they are searching for the solution only in a local area of the search space. Local search approaches might be useful in improving the performance of the basic evolutionary algorithm, which may find the global optimum with sufficient precision in this combined way. Combinations of evolutionary and local-search methods are usually referred to as memetic algorithms [19]. A new kind of memetic algorithm based on the bacterial approach is the bacterial memetic algorithm (BMA) [20].

The algorithm consists of four steps. First, an initial population has to be created. Then, bacterial mutation, a local search and gene transfer are applied, until a stopping criterion is fulfilled. The bacterial mutation is applied to each chromosome one by one. First, N_{clones} copies (clones) of the bacterium are generated, then a certain segment of the chromosome is randomly selected and the parameters of this selected segment are randomly changed in each clone (mutation). Next all the clones and the original bacterium are evaluated and the best individual is selected. This individual transfers the mutated segment into the other individuals. This process continues until all of the segments of the chromosome have been mutated and tested. At the end of this process the clones are eliminated. After the bacterial mutation operator a local search is applied for each individual. This method depends on the given problem. For the TSP it is detailed in the next section.

In the next step the other evolutionary operation, the gene transfer is applied, which allows the recombination of genetic information between two bacteria. First, the population must be divided into two halves. The better bacteria are called the superior half, the other bacteria are called the inferior half. One bacterium is randomly chosen from the superior half, this will be the source bacterium and another is randomly chosen from the inferior half, this will be the destination bacterium. A segment from the source bacterium is chosen randomly and this segment will overwrite a segment of the destination bacterium or it will be added to the destination bacterium. This process is repeated for N_{inf} times. The stopping condition is usually given by a predefined maximum generation number (N_{gen}). When N_{gen} is achieved then the algorithm ends otherwise it continues with the bacterial mutation step. We use eugenic elements also in the algorithm [21]. This reflects the human's decision will, and puts some deterministic element into the algorithm. Details will be given in the next section.

The basic algorithm has four parameters: the number of generations (N_{gen}), the number of bacteria in the population (N_{ind}), the number of clones in the bacterial mutation (N_{clones}), and the number of infections (N_{inf}) in the gene transfer operation.

ISBN: 978-989-95079-6-8

5 EBMA for the modified traveling salesman problem

When applying evolutionary type algorithms first of all the encoding method must be defined. The evaluation of the individuals has to be discussed, too. The operations of the algorithm have to be adapted to the given problem.

5.1 Encoding method and evaluation of the individuals

In the modified traveling salesman problem one city may be visited more than once. Because each city must be visited at least once, one solution of the problem does not need to be a permutation of the cities. The evident encoding of the problem into a bacterium is simply the enumeration of the cities in the order they should be visited. Therefore, a length of the bacterium may be greater than the number of cities (N_{cities}), but an upper bound for the bacterium length has to be defined too, we allow bacteria not longer than $m \cdot N_{cities}$, where m is the multiplier factor, which is a parameter of the algorithm (usually m=2). The initial city is not represented in the bacterium.

The length of the bacteria can be changeable. It can be changing during the evolutionary process and the individuals can have different lengths. In the initial population generation, the length of the bacteria is a random number greater than or equal to N_{cities} and less than or equal to $m \cdot N_{cities}$.

The evaluation of a bacterium is based on the time dependent distance matrix. The distance between the first element of the bacterium and the initial city is taken from the distance matrix at the zeroth time step, the distance between the second element of the bacterium and the first element of the bacterium is taken from the distance matrix at the first time step, and so on, these distances are summed up, and the total distance is obtained in this way.

5.2 Bacterial mutation

In the bacterial mutation there is an additional parameter, the length of the segment to be mutated in the clones. First, the segments of the bacterium are determined, and a random segment order is created. In the clones, the mutation of the given segment is executed. For example in Figure 1, the length of the segment is 3, and there are 4 clones. The random segment order is e.g., {3rd segment, 1st segment, 4th segment, 2nd segment}. This means that in the first sub-cycle of bacterial mutation, the 3rd segment is mutated in the clones. After the mutation of the clones, the best one is being chosen, and this clone (or the un-mutated original bacterium) transfers the mutated segment to the other individuals.

The segments of the bacterium do not need to consist of consecutive elements. The elements of the segments can come from different parts of the bacterium as it can be seen in Figure 2.

Because in the modified TSP the number of visited cities is not predefined, bacteria with different length can occur in the population. Although in the initial population generation bacteria with different length can arise we would like to allow the changes in the length within the bacterial operations, too. Therefore before a clone is mutated a random value is used for determining that after the mutation the length of the clone will increase, decrease or remain the same.

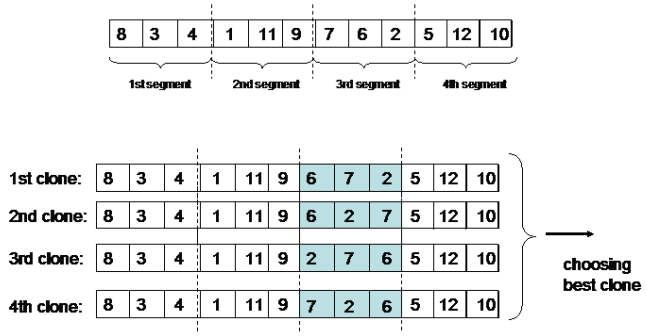


Figure 1: Bacterial mutation

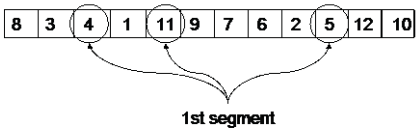


Figure 2: Segment in the bacterial mutation

Increasing is allowed only in the case, when the maximum bacterium length ($m \cdot N_{cities}$) is not exceeded, similarly, decreasing is allowed only in the case, when the minimum bacterium length (N_{cities}) is guaranteed. If the length will increase, then besides changing the positions of the cities in the selected segment of the clone, new cities are added to this clone randomly. If it will decrease, then some cities are deleted from the clone taking care that only those cities are allowed to be deleted, which have at least one other occurrence in the clone. If the length remains the same, then only the positions of the cities in the selected segment are changed.

5.3 Local search method

A tour can be improved by some local heuristics. One of the most successful methods is the Lin-Kernighan algorithm [22] which based on the k-opt algorithm. The k-opt algorithm removes k edges from the tour and reconnects the k paths optimally. We applied the 2-opt and 3-opt technique in our EBMA algorithm. For higher value of k the algorithm would take more time and would provide only small improvements on the 2-opt and 3-opt techniques.

5.4 Gene transfer

In the gene transfer operation there is also an additional parameter, the length of the segment to be transferred from the source bacterium to the destination bacterium. In contrast with the bacterial mutation, in the gene transfer, the segment can contain only consecutive elements within the bacterium. The reason for that is the segment containing consecutive elements representing sub-tours in the bacterium, and transferring good sub-tours is the main goal of the gene transfer operation.

Figure 3 shows the gene transfer in the case of time independent distance matrix. In the case of time dependent

ISBN: 978-989-95079-6-8

distance matrix, the position where the transferred segment goes to in the destination bacterium must be the same as the position of the segment in the source bacterium.

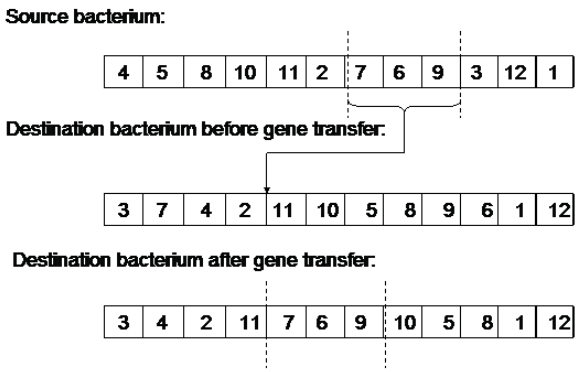


Figure 3: Gene transfer

The different individual lengths are allowed also in the gene transfer operation. After the segment was transferred to the destination bacterium, the elements that occur in the transferred segment and are already in the destination bacterium can be deleted from the destination bacterium. If the same number of elements is deleted from the destination bacterium as the length of the segment, then the length of the destination bacterium remains the same. If less elements are deleted, then its length will be increasing. If more elements are deleted (taking care that each city must have at least one occurrence in the bacterium), then the length will be decreasing. We must also take care that the length of the destination bacterium must be at least N_{cities} and at most $m \cdot N_{cities}$.

5.5 *Eugenic elements*

Eugenic is used in the initial population creation and in the bacterial mutation operator. This means that we put more determinism into the algorithm which contains normally the deterministic local search and the stochastic evolutionary operators. During the initial population creation not only random individuals are generated but also some deterministic ones according to the following rule: there is an individual, which represents the tour in which always the nearest unvisited city is visited. There can be another initial bacterium, which represents the tour in which alternating the nearest and the second nearest city is visited. There can be a third individual, where always the second nearest city is taken.

In the bacterial mutation not only randomly mutated clones are produced, but there will always be a deterministic clone, which performs a reverse ordering permutation on the selected segment. According to our experiences this can be effective in solving TSP like problems.

6 Computational results

First the efficiency of the proposed algorithm is presented. Fig. 4 and 5 compare the computational result against a "classical" reference instance TSP (www.tsp.gatech.edu, XQF131). The total length of the optimal tour is *564*, the solution of our proposed algorithm is *566*, the computational error is *0.35%*.

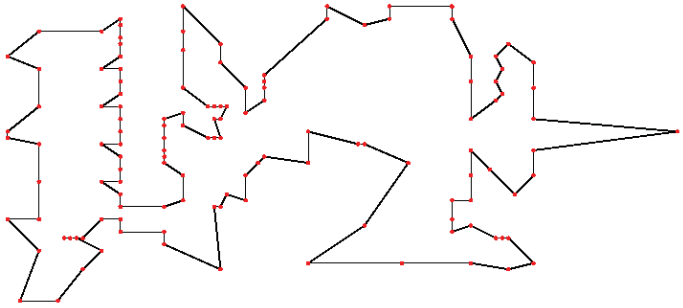


Figure 4: Reference instance (www.tsp.gatech.edu)

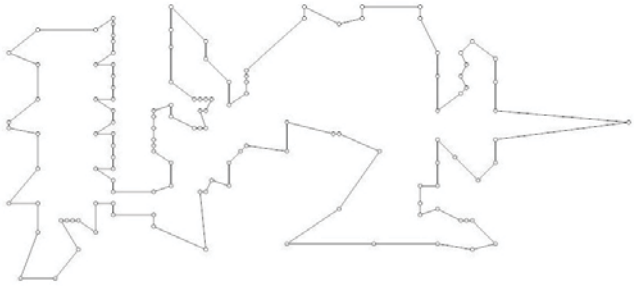


Figure 5: Solution of reference instance by the EBMA

Next in the evaluation process a classic symmetric graph is compared to an asymmetric and time dependent version of the same graph. Solution for the symmetric case is shown in Fig. 6.

The RTTSP version is modified in the following points:

$Node_{16}$ has only one connection to $node_6$.

$Node_6$ has two connections, to $node_{16}$ and to $node_9$.

$$C_{1,23}(t) = C_{1,23} + 0.05t$$
$$C_{20,0}(t) = C_{20,0} + 0.01t$$
$$C_{13,0}(t) = C_{13,0} + 0.02t$$

where t is the step in which the tour visits the node. This is a very simple representation of the time dependency of the graph. The results are shown in Fig. 7.

Since the running time of EBMA is about proportional to the number of generations it is a crucial task to select the most efficient parameters:

- number of bacteria in the population (N_{ind})
- number of clones in the bacterial mutation (N_{clones})
- number of infections (N_{inf}) in the gene transfer operation
- length of the segment to be mutated
- length of the segment to be transferred.

Finally the fuzzy cost coefficients are considered, results are shown in Fig. 8. Let the fuzzy values are:

$$\boldsymbol{C}_{2,0} = (0.015, 0.1986, 0.4)$$
$$\boldsymbol{C}_{0,2} = (0.001, 0.1986, 0.2)$$
$$\boldsymbol{C}_{23,4} = (0.1, 0.4504, 0.5)$$
$$\boldsymbol{C}_{4,23} = (0.4, 0.4504, 0.7)$$
$$\boldsymbol{C}_{26,29}(t) = (0.05, 0.1655, 0.2)\ 0.05\ t$$

ISBN: 978-989-95079-6-8

$C_{29,26}(t)$= (0.11, 0.1655, 0.3) 0.05 t
$C_{10,4}(t)$= (0.03, 0.1449, 0.16) 0.06 t
$C_{4,10}(t)$= (0.12, 0.1449, 0.3) 0.06 t
$C_{17,18}(t)$= (0.15, 0.267, 0.3)+0.05 t
$C_{18,17}(t)$= (0.15, 0.267, 0.3)-0.01 t
$C_{7,23}(t)$= (0.1, 0.18, 0.25)+0.01 t
$C_{23,7}(t)$= (0.1, 0.18, 0.25)-0.01 t
$C_{0,21}(t)$= (0.01, 0.066, 0.10)+0.02 t
$C_{21,0}(t)$= (0.01, 0.066, 0.10)-0.02 t

(Remark: peak values are kept from the crisp matrix.)

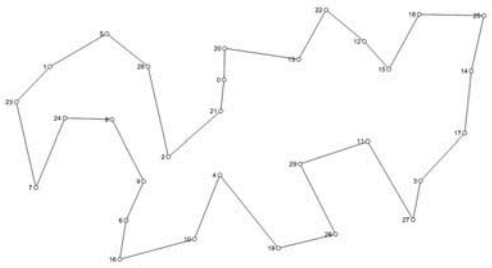


Figure 6: Graphical representation of the best tour of TSP

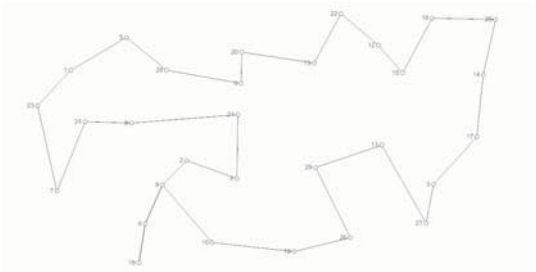


Figure 7: Graphical representation of the best tour in RTTSP



Figure 8: Graphical representation of the best tour in FRTTSP

The scope of future research activity is to set general rules that can give instructions in order to find the most efficient parameters of EBMA according to the size and other (e.g. topographical) features of a given FRTTSP, since running time is significantly affected by those parameters.

Acknowledgment
This paper was supported by the Széchenyi University Main Research Direction Grant 2009, and a National Scientific Research Fund Grant OTKA T048832 and K75711.

References

[1] Holland, J. H.: *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, The MIT Press, Cambridge (1992)

[2] Nawa, N. E., Furuhashi, T.: *Fuzzy System Parameters Discovery by Bacterial Evolutionary Algorithm*, IEEE Tr. Fuzzy Systems 7 (1999) pp. 608-616

[3] Botzheim, J., Hámori, B., Kóczy, L. T., Ruano, A. E.: *Bacterial algorithm applied for fuzzy rule extraction*, in Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems, Annecy, France (2002) pp. 1021-1026

[4] Botzheim J., Drobics M., Kóczy L. T.: *Feature selection using bacterial optimization*, in Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems, IPMU 2004, Perugia, Italy, (July 2004) pp. 797–804

[5] Applegate, D.L., Bixby, R.E., Chvátal, V., Cook, W.J.: *The Traveling Salesman Problem, A Computational Study*, Princeton University Press, Princeton and Oxford (2006) pp. 10-11

[6] Yu-Hsin Liu : *A hybrid scatter search for the probabilistic traveling salesman problem*, Computers & Operations Research 34 (2007) pp. 2949-2963

[7] Teoh, E.J.,Tan, K.C., Tang, H.J., Xiang, C., Goh, C.K.: *An asynchronous recurrent linear threshold network approach to solving the traveling salesman problem*, Neurocomputing 71 (2008) pp. 1359-1372

[8] Ding, C., Cheng, Y., He, M.: *Two-Level Genetic algorithm for Clustered Traveling Salesman Problem with Application in Large-Scale TSPs*, Tsinghua Science and Technology, Vol.12.No.4 (2007) pp. 459-465

[9] Hsiao-Fan Wang, Yu-Pin Wen: *Time-Constrained Chinese Postman Problems*, Computers and Mathematics with Applications 44 (2002) pp. 375-387

[10] Shi, X.C., Liang, Y.C., Lee, H.P., Lu, C., Wang, Q.X.: *Particle swarm optimization-based algorithms for TSP and generalized TSP*, Information Processing Letters 103 (2007) pp. 169-176

[11] Bontoux, B., Feillet, D.: *Ant colony optimization for the traveling purchaser problem,* Computers & Operations Research 35 (2008) pp. 628-637

[12] Yu-Wan Chen, Yong-Zai Lu, Penf Chen: *Optimization with extremal dynamics for the traveling salesman problem*, Physica A 385 (2007) pp. 115-123

[13] Guan-Chun Luh, Shih-Wei Lee: *A bacterial evolutionary algorithm for the job shop scheduling problem,* Journal of the Chinese Institute of Industrial Engineers, Vol. 23, No. 3, (2006) pp. 185-191

[14] Held, M.R., Karp R.M.: *A dynamic programming approaches to sequencing problems*, Journal of the Society of Industrial and Applied Mathematics 10 (1962) pp. 196-210

[15] Ammar,E.E., Youness, E.A.: *Study on multiobjective transportation problem with fuzzy numbers*, Applied Mathematics and Computation 166 (2005) pp. 241-253

[16] Oussalah, M.: *On the compability between defuzzification and fuzzy arithmetic operations*, Fuzzy Sets and Systems 128 (2002) pp. 247-260

[17] Kikuchi, S., Chakroborty, P.: *Place of possibility theory in transportation analysis*, Transportation Research Part B 40 (2006) pp 595-615

[18] Zadeh, L.A.: *The concept of linguistic variable and its application to approximate reasoning*, Inform. Sci. Part 1, 8 (1975) 199-249, Part 2, 301-357, Part 3, 9 (1975) 43-80.

[19] Moscato P. *On Evolution, Search, Optimization, Genetic Algorithms and Martial Arts: Towards Memetic Algorithms*. Caltech Concurrent Computation Program, Tech. Rep., California, 1989.

[20] Botzheim J., Cabrita C., Kóczy L. T., and Ruano A. E., "*Fuzzy rule extraction by bacterial memetic algorithm*," IFSA 2005, Beijing, China, 2005, pp.1563-1568.

[21] Ye J., Tanaka M., Tanino T.: *Eugenics-based genetic algorithm, IEICE Transactions on Information and Systems*, Vol. E79-D No. 5 pp. 600-607.

[22] Lin S., Kernighan B. W.: *An Effective Heuristic Algorithm for the Traveling-Salesman Problem*, Operations Research, 21, pp. 498-516 (1973).

[23] Földesi P., Botzheim J.: *Solution for modified traveling salesman problem with variable cost matrix using bacterial evolutionary algorithm*. Acta Technica Jaurinensis Series Logistica, 1(2):159–171, 2008.

Feature Selection Using Fuzzy Objective Functions

Susana M. Vieira[1] João M. C. Sousa[1] Uzay Kaymak[2]

1. Instituto Superior Técnico, Technical University of Lisbon
Lisbon, Portugal
2.Econometric Institute, Erasmus School of Economics, Erasmus University of Rotterdam
Rotterdam, The Netherlands
Email: susana@dem.ist.utl.pt, jmsousa@ist.utl.pt, u.kaymak@ieee.org

Abstract *— One of the most important stages in data preprocessing for data mining is feature selection. Real-world data analysis, data mining, classification and modeling problems usually involve a large number of candidate inputs or features. Less relevant or highly correlated features decrease, in general, the classification accuracy, and enlarge the complexity of the classifier. Feature selection is a multi-criteria optimization problem, with contradictory objectives, which are difficult to properly describe by conventional cost functions. The use of fuzzy decision making may improve the performance of this type of systems, since it allows an easier and transparent description of the different criteria used in the feature selection process. In previous work an ant colony optimization algorithm for feature selection was presented, which minimizes two objectives: number of features and classification error. Two pheromone matrices and two different heuristics are used for each objective. In this paper, a fuzzy objective function is proposed to cope with the difficulty of weighting the different criteria involved in the optimization algorithm.*

Keywords— Feature selection, fuzzy decision functions, ant colony optimization.

1 Introduction

Feature selection has been an active research area in data mining, pattern recognition and statistics communities. The main idea of feature selection is to choose a subset of available features, by eliminating features with little or no predictive information and also redundant features that are strongly correlated. Many practical pattern classification tasks (e.g., medical diagnosis) require learning of an appropriate classification function that assigns a given input pattern (typically represented by using a vector of feature values) to one of a set of classes. The choice of features used for classification has an impact on the accuracy of the classifier and on the time required for classification.

The challenge is selecting the minimum subset of features with little or no loss of classification accuracy. The feature subset selection problem consists of identifying and selecting a useful subset of features from a larger set of often mutually redundant, possibly irrelevant, features with different associated importance [1].

Like many design problems, feature selection problem, is characterized by multiple objectives, where a trade-off amongst various objectives must be made, leading to under or over-achievement of different objectives. Moreover, some flexibility may be present for specifying the constraints of the problem. Furthermore, some of the objectives in decision making may be known only approximately. Fuzzy set theory provides ways of representing and dealing with flexible or soft constraints, in which the flexibility in the constraints can be exploited to obtain additional trade-off between improving the objectives and satisfying the constraints.

Various fuzzy optimization methods have been proposed in the literature in order to deal with different aspects of soft constraints. In one formulation of fuzzy optimization due to Zimmermann [2], which is used in the rest of this paper, concepts from Bellman and Zadeh model of fuzzy decision making [3] are used for formulating the fuzzy optimization problem.

In this paper is used a feature selection algorithm which is based in ant colony optimization. The ant feature selection algorithm uses two cooperative ant colonies, which are used to cope with two different objectives. The two objectives we consider are minimizing the number of features and minimizing the classification error. Two pheromone matrices and two different heuristics are used for each objective. These goals are translated into fuzzy sets.

The paper is organized as follows. Section 2 presents a brief description of fuzzy optimization. A brief consideration of the fuzzy models we use is presented in Section 3. The ACO feature selection algorithm is presented in Section 4. In Section 5 the results are presented and discussed. Some conclusions are drawn in Section 6 and the possible future work is discussed.

2 Fuzzy Optimization

Fuzzy optimization is the name given to the collection of techniques that formulate optimization problems with flexible, approximate or uncertain constraints and goals by using fuzzy sets. In general, fuzzy sets are used in two different ways in fuzzy optimization.

1. To represent uncertainty in the constraints and the goals (objective functions).
2. To represent flexibility in the constraints and the goals.

In the first case, fuzzy sets represent generalized formulations of intervals that are manipulated according to rules similar to interval calculus by using the α-cuts of fuzzy sets. In the second case, fuzzy sets represent the degree of satisfaction of the constraints or of the aspiration levels of the goals, given the flexibility in the formulation. Hence, the constraints (and the goals) that are essentially crisp are assumed to have some flexibility that can be exploited for improving the optimization objective. This framework is suitable for the representation of interaction and possible trade-off amongst the constraints and the objectives of the optimization [4].

ISBN: 978-989-95079-6-8

2.1 General formulation

The general formulation for *fuzzy optimization* in the presence of flexible goals and constraints is given by

$$\begin{array}{ll} \underset{\mathbf{x}\in X}{\text{fuzzy maximize}} & [f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_n(\mathbf{x})] \\ \text{subject to} & g_i(\mathbf{x}) \tilde{\leq} 0, \quad i = 1, 2, \ldots, m. \end{array} \tag{1}$$

In (1), the tilde sign denotes a fuzzy satisfaction of the constraints. The sign $\tilde{\leq}$ thus denotes that $g_i(\mathbf{x}) \leq 0$ can be satisfied to a degree smaller than 1. The fuzzy maximization corresponds to achieving the highest possible aspiration level for the goals $f_1(\mathbf{x})$ to $f_n(\mathbf{x})$, given the fuzzy constraints to the problem. This optimization problem can be solved by using the approach of Bellman and Zadeh to fuzzy decision making [3].

Consider a decision making problem where the decision alternatives are $\mathbf{x} \in X$. A fuzzy goal F_j, $j = 1, 2, \ldots, n$ is a fuzzy subset of X. Its membership function $F_j(\mathbf{x})$, with $F_j : X \rightarrow [0, 1]$ indicates the degree of satisfaction of the decision goal by the decision alternative $\mathbf{x}$. Similarly, a number of fuzzy constraints G_i, $i = 1, 2, \ldots, m$ can be defined as fuzzy subsets of X. Their membership functions $G_i(\mathbf{x})$, denote the degree of satisfaction of the fuzzy constraint G_i by the decision alternative $\mathbf{x} \in X$. According to Bellman and Zadeh's fuzzy decision making model, the fuzzy decision D is defined as the confluence of the fuzzy goals and constraints, i.e.

$$D(\mathbf{x}) = F_1(\mathbf{x}) \circ \cdots \circ F_n(\mathbf{x}) \circ G_1(\mathbf{x}) \circ \cdots \circ G_m(\mathbf{x}), \tag{2}$$

where $\circ$ denotes an aggregation operator for fuzzy sets. Since the goals and the constraints must be satisfied simultaneously, Bellman and Zadeh proposed to use an intersection operator, i.e. a fuzzy t-norm for the aggregation. The optimal decision alternative $\mathbf{x}^*$ is then the argument that maximizes the fuzzy decision, i.e.

$$\mathbf{x}^* = \arg\max_{\mathbf{x}\in X} D(\mathbf{x}). \tag{3}$$

The optimization problem is then defined by

$$\max_{\mathbf{x}\in X} [F_1(\mathbf{x}) \wedge \cdots \wedge F_n(\mathbf{x}) \wedge G_1(\mathbf{x}) \wedge \cdots \wedge G_m(\mathbf{x})]. \tag{4}$$

Note that both the goals and the constraints are aggregated. Hence, the goals and the constraints are treated equivalently, which is why the model is said to be symmetric. The symmetric model is not always appropriate, however, since the aggregation of the goals and the constraints may have different requirements. Often, for example, some trade off amongst the goals is allowed or may even be desirable, which may be modeled by an averaging operation. The constraints, however, should not be violated, i.e. their aggregation must be conjunctive. In that case, the goals and the constraints cannot be combined uniformly by using a single aggregation operator. In the simplest case, the goals must be combined by using one operator and the constraints must be combined by using another operator. The aggregated results must then be combined at a higher level by using a third aggregation operator, which has to be conjunctive (i.e. both the aggregated goals and the aggregated constraints should be satisfied).

Clearly, the above formulation of fuzzy optimization is closely related to the penalty function methods known from classical optimization theory. The aggregated goals correspond to an overall objective function, which is maximized. The constraints are added to this objective function by using fuzzy t-norms, which is similar to the addition of a penalty function to an optimization objective function in classical optimization. After combining the objectives and the constraints, the resulting optimization is unconstrained, but possibly non-convex. Furthermore, gradient descent methods may not be suitable for the maximization due to possible and likely discontinuity in the first derivative of the final aggregated function. Derivative-free search and optimization algorithms such as simulated annealing, evolutionary algorithms or other bio-inspired algorithms, such as ant colony optimization, can be used to solve this type of optimization problems.

2.2 Weighted aggregation in fuzzy optimization

Weighted aggregation has been used quite extensively especially in fuzzy decision making, where the weights are used to represent the relative importance that the decision maker attaches to different decision criteria (goals or constraints). Almost always an averaging operator has been used for the weighted aggregation, such as the generalized means [5].

The averaging operators are suitable for modeling compensatory aggregation. They are not suitable, however, for modeling simultaneous satisfaction of aggregated criteria. Since the goal in fuzzy optimization is the simultaneous satisfaction of the optimization objectives and the constraints, t-norms must be used to model the conjunctive aggregation. In order to use the weighted aggregation in fuzzy optimization, weighted aggregation using t-norms must thus be considered.

The axiomatic definition of t-norms does not allow for weighted aggregation. In order to obtain a weighted extension of t-norms, some of the axiomatic requirements must be dropped. Especially the commutativity and the associativity properties must be dropped, since weighted operators are by definition not commutative.

Weighted t-norms. In [4], the authors used weighted counterparts of several t-norms for fuzzy optimization. The specific operators considered are the weighted extension of the product t-norm given by

$$D(\mathbf{x}, \mathbf{w}) = \prod_{i=1}^{m} [G_i(\mathbf{x})]^{w_i}, \tag{5}$$

the extension of the Yager t-norm given by

$$D(\mathbf{x}, \mathbf{w}) = \max\left(0, 1 - \sqrt[s]{\sum_{i=1}^{m} w_i (1 - G_i(\mathbf{x}))^s}\right), \quad s > 0. \tag{6}$$

The term fuzzy optimization in the remainder of this paper also refers to a formulation in terms of the flexibility of the goals.

3 Fuzzy Models for Classification

Fuzzy modeling for classification, is a technique that allows an approximation of nonlinear systems when there is none or few knowledge of the system to be modeled [6]. The fuzzy modeling approach has several advantages when compared to

ISBN: 978-989-95079-6-8

other nonlinear modeling techniques. In general, fuzzy models can provide a more transparent model and can also give a linguistic interpretation in the form of rules, which is appealing when dealing with classification systems. Fuzzy models use rules and logical connectives to establish relations between the features defined to derive the model. This paper uses Takagi-Sugeno (TS) fuzzy models [7], which consist of fuzzy rules where each rule describes a local input-output relation, typically in an affine form. The affine form of a TS model is given by:

$$R_i : \textbf{If } x_1 \text{ is } A_{i1} \textbf{ and } \ldots \textbf{ and } x_n \text{ is } A_{in} \textbf{ then } y_i = a_{i1}x_1 + \ldots + a_{in}x_n + b_i\,, \quad (7)$$

where $i = 1, \ldots, K$, K denotes the number of rules in the rule base, R_i is the i^{th} rule, $\mathbf{x} = [x_1, \ldots, x_n]^T$ is the antecedent vector, n is the number of features, $A_{i1}, \ldots, A_{in}$ are fuzzy sets defined in the antecedent space, y_i is the output variable for rule i, $\mathbf{a}_i$ is a parameter vector and b_i is a scalar offset. The consequents of the affine TS model are hyperplanes in the product space of the inputs and the output. The model output, y, can then be computed by aggregating the individual rules contribution:

$$y = \frac{\sum_{i=1}^{K} \beta_i y_i}{\sum_{i=1}^{K} \beta_i}, \quad (8)$$

where β_i is the degree of activation of the ith rule:

$$\beta_i = \prod_{j=1}^{n} \mu_{A_{ij}}(x_j), \quad (9)$$

and $\mu_{A_{ij}}(x_j) : \mathbb{R} \rightarrow [0, 1]$ is the membership function of the fuzzy set A_{ij} in the antecedent of R_i.

The performance criterion used to evaluate the fuzzy model is the classification accuracy C_a, given by the percentage of correct classifications:

$$C_a = \frac{(N_n - N_e)}{N_n} \times 100\%, \quad (10)$$

where N_n is the number of used samples and N_e is the number of classification errors in test samples (misclassifications).

4 Ant Feature Selection

Ant algorithms were first proposed by Dorigo [8] as a multi-agent approach to difficult combinatorial optimization problems, such as traveling salesman problem, quadratic assignment problem or supply chain management [9, 10]. The ACO methodology is an optimization method suited to find minimum cost paths in optimization problems described by graphs [11]. This paper presents a new implementation of ACO applied to feature selection, where the best number of features is determined automatically. In this approach, two objectives are considered: minimizing the number of features and minimizing the classification error. Two cooperative ant colonies optimize each objective. The first colony determines the number (cardinality) of features and the second selects the features based on the cardinality given by the first colony. Thus, two pheromone matrices and two different heuristics are used. A novel approach for computing a heuristic value is proposed to determine the cardinality of features. The heuristic value is computed using the Fisher discriminant criterion for feature selection [12], which ranks the features giving them a given relative importance and it is described in more detail in section 4.2.3. The best number of features is called *features cardinality* N_f. The determination of the *features cardinality* is addressed in the first colony sharing the same minimization cost function with the second colony, which in this case aggregates both the maximization of the classification accuracy and the minimization of the features cardinality. Hence, the first colony determines the size of the subsets of the ants in the second colony, and the second colony selects the features that will be part of the subsets.

The algorithm used in this paper deals with the feature selection problem as a multi–criteria problem with a single objective function. Therefore, a pheromone matrix is computed for each criterion, and different heuristics are used.

The objective function of this optimization algorithm aggregate both criteria, the minimization of the classification error rate and the minimization of the features cardinality:

$$J^k = w_1 \frac{N_e^k}{N_n} + w_2 \frac{N_f^k}{n} \quad (11)$$

where $k = 1, \ldots, g$, N_n is the number of used data samples and n is the total number of features. The weights w_1 and w_2 are selected based on experiments.

To evaluate the classification error, a fuzzy classifier is built for each solution following the procedure described in Section 3.

4.1 Fuzzy decision function

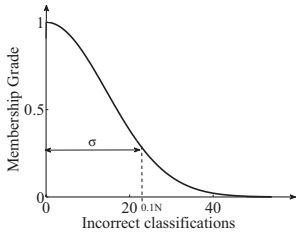
The objective function (11) can be interpreted as follows. The term containing the predicted errors indicates that these should be minimized, while the term containing the size of the features subset used indicates that the number of features should be reduced. Hence, minimizing the output errors and the size of the features subset can be regarded as forcing the model to be less complex and maintain a good performance at the same time. The parameters containing the weights, w_1 and w_2, can be changed so that the objective function is modified in order to lead to a solution where the accuracy of the model is more important or to a much simpler model where the reduction of the number of the features is imperative. This balance always depends on the application.

When fuzzy multicriteria decision making is applied to determine the objective function, additional flexibility is introduced. Each criterion ζ_i is described by a fuzzy set, where $i = 1, \ldots, T$, stands for the different criteria defined. Fuzzy criteria can be described in different ways. The most straightforward and easy way is just to adapt the criteria defined for the classical objective functions. Fig. 1 shows examples of general membership functions that can be used for the error N_e^k and for the features cardinality N_f^k, with $k = 1, \ldots, g$. In this example, the minimization of the classification error is represented by an exponential membership function, given by

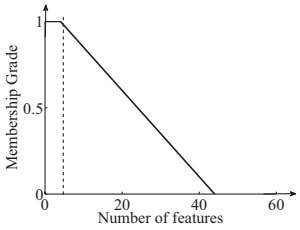
$$\mu_e^k = \exp^{-\left(\frac{N_e^k - a}{2\sigma}\right)^2} \quad (12)$$

This well-known function has the nice property of never reaching the value zero, and the membership value is still quite considerable for an error of 10%. Therefore, this criterion is

ISBN: 978-989-95079-6-8



(a) Membership function for the number of incorrect classifications.



(b) Membership function for the number of selected features

Figure 1: Membership functions for the feature selection goals.

considered to be a fuzzy goal. The features cardinality N_f can be represented, for example, by a trapezoidal membership function around zero, as shown in Fig. 1b. A reduced number of features is considered to be a desired outcome of the optimization algorithm. The membership function is defined so that for a low number of features the membership degree is one and linearly decreases to zero. The membership degree should be zero outside the maximum number of features available. The parameters defining the range of the trapezoidal membership function are application dependent. Sometimes it is convenient to make the upper limit of the membership function much lower than the maximum number of features allowed, specially if a very large number of features is being tested. In general, all the parameters of the different membership functions are application dependent. However, it is possible to derive some tuning guidelines, as will be described here. The membership functions quantify how much the system satisfies the criteria given a particular feature subset solution, bringing various quantities into a unified domain. The use of the membership functions introduces additional flexibility to the goals, and it leads to increased transparency as it becomes possible to specify explicitly what kind of solution is preferred. For instance, it becomes easier to penalize more severely a subset of features that have larger classification errors. Or if we prefer a solution with less features, a higher number of features can also be penalized.

Note that there is no need to scale the several parameters and as in (11), when fuzzy objective functions are used, because the use of membership functions introduce directly the normalization required. For this particular aspect, this feature reduces the effort on defining the objective function, when compared to classical objective functions. After the membership functions have been defined, they are combined by using a decision function, such as a parametric aggregation operator from the fuzzy sets theory (see Section 2).

Table 1: List of symbols.

Variable	Description
General	
n	Number of features
N	Number of samples
N_n	Number of samples used for validation
I	Number of iterations
K	Number of rules/clusters of the fuzzy model
N_c	Number of existing classes in database
g	Number of ants
$\mathbf{x}$	Set with all the features
$\mathbf{w}$	Subset of features selected to build classifiers
J^k	Cost of the solution for each ant k
J^q	Cost of the winner ant q
Ant colony for cardinality of features	
N_f	Features cardinality (number of selected features)
$N_f(k)$	Features cardinality of ant k
I_n	Number of iterations with same feature cardinality
α_n	Pheromone weight of features cardinality
β_n	Heuristic weight of features cardinality
τ_n	Pheromone trails for features cardinality
η_n	Heuristic of features cardinality
ρ_n	Evaporation of features cardinality
Γ_n^k	Feasible neighborhood of ant k (features cardinality availability)
$\mathcal{Q}_i$	Amount of pheromone laid in the features cardinality of the best solution
Ant colony for selecting subset of features	
$L_f^k(t)$	Feature subset for ant k at tour t
α_f	Pheromone weight of features
β_f	Heuristic weight of features
τ_f	Pheromone trails for feature selection
η_f	Heuristic of features
ρ_f	Evaporation of features
Γ_f^k	Feasible neighborhood of ant k (features availability)
$\mathcal{Q}_j$	Amount of pheromone laid in the features of the best solution

4.2 Algorithm description

Let $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T$ be the set of given n features, and $\mathbf{w} = [w_1, w_2, \ldots, w_{N_f}]^T$, be a subset of features where $(\mathbf{w} \subseteq \mathbf{x})$. It is desirable that $N_f << n$. Table 1 describes the variables used in the algorithm.

4.2.1 Probabilistic rule

Consider a problem with N_f nodes and two colonies of g ants. First, g ants of the first colony randomly select the number of nodes N_f to be used by the g ants of the second colony. The probability that an ant k chooses the features cardinality $N_f(k)$ is given by

$$p_i^k(t) = \frac{[\tau_{n_i}]^{\alpha_n} \cdot [\eta_{n_i}]^{\beta_n}}{\sum_{l \in \Gamma_n^k} [\tau_{n_l}]^{\alpha_n} \cdot [\eta_{n_l}]^{\beta_n}} \tag{13}$$

where τ_{n_i} is the pheromone concentration matrix and η_{n_i} is the heuristic function matrix, for path (i). The values of the

ISBN: 978-989-95079-6-8

pheromone matrix are limited to $[\tau_{n_{min}}, \tau_{n_{max}}]$, with $\tau_{n_{min}} = 0$ and $\tau_{n_{max}} = 1$. Γ_n^k is the feasible neighborhood of ant k (available number of features to be selected), which acts as the memory of the ants, and contains all the trails that the ants have not passed and can be chosen. The parameters α_n and β_n measure the relative importance of trail pheromone and heuristic knowledge, respectively.

After all the g ants from the first colony have chosen the features cardinality $N_f(k)$, each ant k from the second colony select $N_f(k)$ features (nodes). The probability that an ant k chooses feature j as the next feature to visit is given by

$$p_j^k(t) = \frac{[\tau_{f_j}(t)]^{\alpha_f} \cdot [\eta_{f_j}]^{\beta_f}}{\sum_{l \in \Gamma_f^k} [\tau_{f_l}(t)]^{\alpha_f} \cdot [\eta_{f_l}]^{\beta_f}} \quad (14)$$

where τ_{f_j} is the pheromone concentration matrix and η_{f_j} is the heuristic function matrix for the path (j). Again, the pheromone matrix values are limited to $[\tau_{f_{min}}, \tau_{f_{max}}]$, with $\tau_{f_{min}} = 0$ and $\tau_{f_{max}} = 1$. Γ_f is the feasible neighborhood of ant k (available features), which contains all the features that the ants have not selected and can be chosen. Again, the parameters α_f and β_f measure the relative importance of trail pheromone and heuristic knowledge, respectively.

4.2.2 Updating rule

After a complete tour, when all the g ants have visited all the $N_f(k)$ nodes, both pheromone concentration in the trails are updated by

$$\tau_{n_i}(t+1) = \tau_{n_i}(t) \times (1 - \rho_n) + \Delta\tau_{n_i}(t) \quad (15)$$

$$\tau_{f_j}(t+1) = \tau_{f_j}(t) \times (1 - \rho_f) + \Delta\tau_{f_j}(t) \quad (16)$$

where $\rho_n \in [0, 1]$ is the pheromone evaporation of the features cardinality , $\rho_f \in [0, 1]$ is the pheromone evaporation of the features and $\Delta\tau_{n_i}$ and $\Delta\tau_{f_j}$ are the pheromone deposited on the trails (i) and (j), respectively, by the ant q that found the best solution J^q for this tour:

$$\Delta\tau_{n_i}^q = \begin{cases} \mathcal{Q}_i & \text{if node } (i) \text{ is used by the ant } q \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

$$\Delta\tau_{f_j}^q = \begin{cases} \mathcal{Q}_j & \text{if node } (j) \text{ is used by the ant } q \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

The number of nodes $N_f(k)$ that each ant k has to visit on each tour t is only updated every I_n tours (iterations), in order to allow the search for the best features for each features cardinality N_f. The algorithm runs I times. Both colonies share the same cost function given in (11).

4.2.3 Heuristics

The heuristic value used for each feature (ants visibility) for the second colony, is computed as

$$\eta_{f_j} = 1/N_{e_j} \quad (19)$$

for $j = 1, \ldots, n$. For the features cardinality (first colony), the heuristic value is computed using the

Fisher discriminant criterion for feature selection [12]. Considering a classification problem with two possible classes, class 1 and class 2, the Fisher discriminant criterion is described as

$$F(i) = \frac{|\mu_1(i) - \mu_2(i)|^2}{\sigma_1^2 + \sigma_2^2} \quad (20)$$

Algorithm 1 Ant Feature Selection

```
/*Initialization*/
set the parameters ρ_f, ρ_n, α_f, α_n, β_f, β_n, I, I_n, g.
for t = 1 to I do
  for k = 1 to g do
    Choose the subset size N_f(k) of each ant k using (13)
  end for
  for l = 1 to I_n do
    for k = 1 to g do
      Build feature set L_f^k(t) by choosing N_f(k) features using (14)
      Compute the fuzzy model using the L_f^k(t) path selected by ant k
      Compute the cost function J^k(t)
      Update J^q
    end for
    Update pheromone trails τ_{n_i}(t+1) and τ_{f_j}(t+1), as defined in (15) and (16).
  end for
end for
```

where $\mu_1(i)$ and $\mu_2(i)$ are the mean values of feature i for the samples in class 1 and class 2, and σ_1^2 and σ_2^2 are the variances of feature i for the samples in class 1 and 2. The score aims to maximize the between-class difference and minimize the within-class spread. Other currently proposed rank-based criteria generally come from similar considerations and show similar performance [12]. Since our goal is to work with several classification problems, which can contain two or more possible classes, a one versus-all strategy is used to rank features. Thus, for a C-class prediction problem, a particular class is compared with the other $C-1$ classes that are considered together. The features are weighted according to the total score summed over all C comparisons:

$$\sum_{j=1}^{C} F_j(i), \quad (21)$$

where $F_j(i)$ denotes the Fisher discriminant score for the i^{th} feature at the j^{th} comparison. Algorithm 1 presents the description of the ant feature selection algorithm.

5 Experimental Results

The effectiveness of the proposed method is applied to a data set taken from the UCI repository [13].

The classification error rates of the used classifiers are obtained by performing 10 independent runs. The data set is divided in test and training sets. The experimental results are presented as the best, the worst and the mean value of correct classifications C_a out of ten independent runs. These ten runs were obtained using the same test data set.

Wine data set The wine data set is a widely used classification data available online in the repository of the University of California [13], and contains the chemical analysis of 178 wines grown in the same region in Italy, derived from three different cultivars. Thirteen continuous attributes are available for classification: alcohol, malic acid, ash, alkalinity of ash, magnesium, total phenols, flavanoids, nonflavanoids phenols, proanthocyanism, color intensity, hue, OD280/OD315 of

ISBN: 978-989-95079-6-8

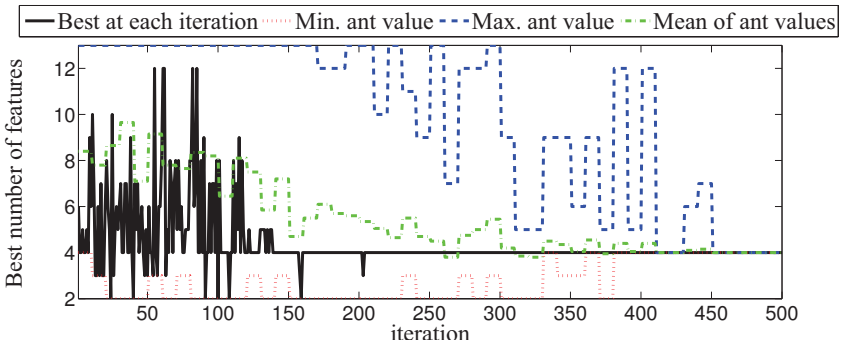


Figure 2: Convergence of the best number of feature during an algorithm run.

Table 2: Classification rates on the Wine data.

Method	**Number of features**	**Classification accuracy** (%)		
		Max.	**Mean**	**Min.**
AFS classic obj. func.	4-8	100	99.8	98.8
AFS fuzzy obj. func.	4	100	99.7	98.6

dilluted wines and proline. This data set has 13 features, three different classes and 178 samples. The AFS algorithm with fuzzy objective function is applied to select the relevant features within the wine classification data set and is compared to the same algorithm with a non fuzzy objective function.

As can be seen in Table 2, the obtained results are better than those in feature selection with classical objective function, once fewer features are selected and because the algorithm always converges to the same number of features.

An example of the process of the ant feature selection with fuzzy objectives searching for optimal solutions for wine data set is given in Fig. 2, where it is possible to observe how all the ants converge to the same solution.

6 Conclusions

A fuzzy objective function for ant feature selection is proposed in this paper. The problem is divided into two objectives: choosing the features cardinality and selecting the most relevant features. The feature selection algorithm uses fuzzy classifiers. The proposed algorithm was applied to a well known classification database that is considered a benchmark. The performance of the proposed algorithm was compared to previous works. The ant based feature selection algorithm yielded similar or better classification rates and the convergence of the solution is better than the previous approach.

In the near future, the proposed feature selection algorithm will be applied to classification problems with a larger number of features.

Acknowledgements

This work is supported by the Portuguese Government and FEDER under the programs: Programa de financiamento Plurianual das Unidades de I&D da FCT (POCTI-SFA-10-46-IDMEC) and by the FCT grant SFRH/25381/2005, Fundação para a Ciência e a Tecnologia, Ministério do Ensino Superior, da Ciência e da Tecnologia, Portugal.

References

[1] H. Motoda H. Liu. *Feature Selection for Knowledge Discovery and Data Mining*. Kluwer Academic Publishers, 1998.

[2] H. J. Zimmermann. Description and optimization of fuzzy systems. *International Journal of General Systems*, 2:209–215, 1976.

[3] R. E. Bellman and L. A. Zadeh. Decision-making in a fuzzy environment. *Management Science*, 17(4):141–164, 1970.

[4] U. Kaymak and J.M. Sousa. Weighted constraint aggregation in fuzzy optimization. *Constraints*, 8(1):61–78, January 2003.

[5] U. Kaymak and H. R. van Nauta Lemke. A sensitivity analysis approach to introducing weight factors into decision functions in fuzzy multicriteria decision making. *Fuzzy Sets and Systems*, 97(2):169–182, July 1998.

[6] L. F. Mendonça, S. M. Vieira, and J. M. C. Sousa. Decision tree search methods in fuzzy modeling and classification. *International Journal of Approximate Reasoning*, 44(2):106–123, 2007.

[7] 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, 1985.

[8] M. Dorigo. *Optimization, Learning and Natural Algorithms (in Italian)*. PhD thesis, 1992.

[9] C. A. Silva, J. M. C. Sousa, and T. A. Runkler. Rescheduling and optimization of logistic processes using GA and ACO. *Engineering Applications of Artificial Intelligence*, 21(3):343–352, 2007.

[10] C. A. Silva, J. M. C. Sousa, T. A. Runkler, and J. M. G. Sá da Costa. Distributed optimization of a logistic system and its suppliers using ant colony optimization. *International Journal of Systems Science*, 37(8):503–512, 2006.

[11] Marco Dorigo, M. Birattari, and T. Stützle. Ant colony optimization. *IEEE Computational Intelligence Magazine*, 1(4):28–39, 2006.

[12] R. O. Duda, P. E. Hart, and D. G Stork. *Pattern Classification, Second Edition*. Wiley–Interscience Publication, 2001.

[13] A. Asuncion and D.J. Newman. UCI machine learning repository, 2007.

ISBN: 978-989-95079-6-8

Clustering of Fuzzy Shapes by Integrating Procrustean Metrics and Full Mean Shape Estimation into K-Means Algorithm

Vasile Georgescu

Faculty of Economics and Business Administration, University of Craiova
Craiova, Romania
Email: vgeo@central.ucv.ro, v_geo@yahoo.com

Abstract—*In this paper we propose a generalization of K-means algorithm, which is adapted to integrate Procrustean metrics and full mean shape estimation, with the aim of clustering objects with either multiple or fuzzy contours. First we are concerned with the representation of fuzzy shapes and introduce appropriate shape metrics and descriptors. Next, we discuss Procrustean methods for aligning shapes, finding mutual dissimilarities and estimating shape class centroid. In the case of multiple-contour crisp shapes, we can benefit from the Extended Orthogonal Procrustes method to find mutual distances between shape pairs and from the Generalized Orthogonal Procrustes technique to estimate the Procrustes mean shape of a collection of shapes. On the other hand, dealing with the case of fuzzy shapes needs more advanced Procrustean techniques to consider weighted distances between points placed on* α *– level contours with different membership degrees. This leads to solve a Weighted Orthogonal Procrustes problem, which typically needs to introduce a weighting matrix of residuals (distances). As an application, we suggest using such methods to cluster ultrasound images of lymph nodes, which typically appear as double-contour shapes.*

Keywords— Clustering of fuzzy shapes, Fuzzy shape metrics and descriptors, Procrustes analysis, Mixing K-means algorithm with Procrustean metrics and mean shape estimation.

1 Shape analysis

Shapes and textures are extremely important features in human as well as machine vision and understanding systems. Shape analysis is concerned with two main classes of algorithms: boundary-based (when only the shape boundary points are used for the description) and region-based (when the whole interior of a shape is used). There are many imaging applications where image analysis can be reduced to the analysis of shapes, in contrast to texture analysis. However, many shape/edge detection techniques use texture information during the segmentation process.

There are several methods for extracting data from shapes, each with their own benefits and weaknesses. These include measurement of lengths and angles, landmark analysis and outline analysis. A landmark is a point of correspondence on each object that matches higher dimensionalities between and within populations. Landmark placement consists of locating a finite number of points on the outline.

More advanced techniques have been designed for semiautomatic and automatic feature extractions. Active contour modeling techniques are commonly used for shape analysis and detection. Some of the techniques for texture feature extraction use gray level co-occurrence matrices, fractal dimension, etc.

Morphometric analysis aims to describe the shape of an object in a way that removes extraneous information and thereby facilitates comparison between different objects. In these terms, a shape is referred to as an invariant to similarity transformations (such as scaling, rotation and translation).

The image fuzzification plays a pivotal role in all image processing systems. Several kinds of image fuzzification can be distinguished:

- histogram-based grey-level fuzzification (e.g. brightness in image enhancement);
- local fuzzification (e.g. edge detection);
- feature fuzzification (scene analysis, object recognition).

2 Representation of fuzzy shapes

2.1 Crisp shapes

Crisp shapes represent objects with crisp borders. Furthermore, if a texture is associated with the object, it has to be uniformly represented (e.g. a digitized image, where all pixels are classified as object pixels, or as background pixels).

The coordinates of selected landmarks for a crisp shape can be arranged in a $n \times p$ configuration matrix A, or equivalently on a $np \times 1$ configuration vector $a = vec(A)$.

2.2 Continuous fuzzy shapes

This paper primarily focuses on the representation of fuzzy shapes with fuzzy contour, which are commonly obtained through fuzzy segmentation techniques. In particular, we also consider the case of crisp shapes with multiple contours. In the same way as it is convenient to model binary images as crisp objects, it is possible to model grey-level images directly as fuzzy sets. If the grey-level values of an image are scaled to be between 0 and 1, the grey-level of a pixel can be seen as its membership to the set of high-valued (bright) pixels.

Fuzziness of an image representation can arise from various reasons, such as limited acquisition conditions (scanning resolution), but also as intrinsic property of the image, which may have imprecise borders. In such cases, pixels close to the border of the object have assigned to them a fuzzy membership value according to the extent of their belongingness to the object.

ISBN: 978-989-95079-6-8

Continuous fuzzy shapes can be described as fuzzy geometric objects. A continuous fuzzy geometric object A in $\Re^p$ is defined as a set of pairs $\{(x, \mu_A(x)) \mid x \in \Re^p\}$ where $\mu_A : \Re^p \to [0,1]$ is the membership function of A in $\Re^n$. It is assumed to have a bounded support.

An alternative representation of fuzzy geometric objects is given by a set of $\alpha-$cuts: $C(A) = \{A_\alpha \mid \alpha \in [0,1]\}$, where $A_\alpha = \{x \in \Re^p \mid \mu_A(x) \geq \alpha\}$ is a crisp object, whose $\alpha-$level contour is obtained for $\mu_A(x) = \alpha$. As a characteristic of fuzzy geometric objects, the membership function is non-increasing away from the interior of the object. For example, in figure 1 is shown a fuzzy disk. Its core is a crisp disk defined by $A_1 = \{x \in \Re^2 \mid (x_1^2 + x_2^2 \leq r_1)\}$ and its contour is the circle defined by $A_1^C = \{x \in \Re^2 \mid (x_1^2 + x_2^2 = r_1)\}$, where r_1 is the length of the corresponding radix. In general, for any $\alpha \in [0,1]$, the $\alpha-$cut is defined by $A_\alpha = \{x \in \Re^2 \mid (x_1^2 + x_2^2 \leq r_\alpha)\}$, and the $\alpha-$level contour is defined by $A_\alpha^C = \{x \in \Re^2 \mid (x_1^2 + x_2^2 = r_\alpha)\}$.

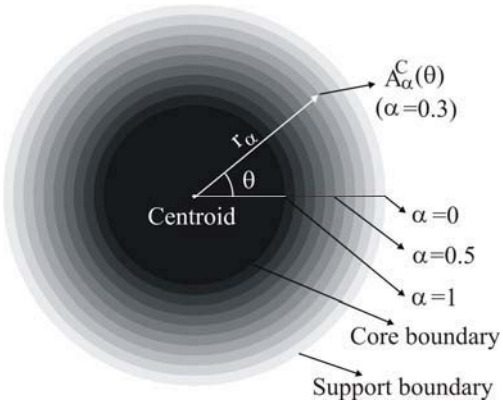


Figure 1: A fuzzy disk: centroid, core, support, α-level contours, radial distance

A shape descriptor based on a one-dimensional functional representation of the two-dimensional shape boundary is called a signature of the shape. The simplest way to generate a signature is to express the radial distance from the centroid to the boundary, as a function of the angle. This is called the centroid distance function. Thus, for crisp objects, the shape signature function corresponds to the Euclidean distance between each boundary point $A(t) = (x(t), y(t))$ and the centroid $A_c = (x_c, y_c)$ of the shape:

$$CD(t) = \sqrt{(x(t) - x_c)^2 + (y(t) - y_c)^2}$$

This shape signature function based on the centroid distance is a convenient choice in the case of star-shaped objects with respect to the centroid (i.e., for each point $y \in A$, the line segment connecting y with the centroid is contained in A).

In the case of a fuzzy object, boundary points are not strictly defined; there is a progressive transition of the membership values from the support outline to the core outline.

The shape signature can be generalized for a continuous fuzzy shape in two possible ways:

- as s radial integral of the membership function:

$$CD_{fuzzy-1}(t) = \int_{A_c}^{A(t)} \mu_A(x(\rho), y(\rho))\, d\rho$$

where $\rho = \rho(t)$ is a parameterization of the straight path between a boundary point and the centroid.

- as an average signature obtained from the $\alpha-$cuts:

$$CD_{fuzzy-2}(t) = \int_0^1 CD_\alpha(t)\, d\alpha$$

where fuzzy star-shaped objects are considered, with all the boundaries of there $\alpha-$cuts jointly indexed by the same parameter t.

A path π in $\Re^p$ from a point $x \in \Re^p$ to another point $y \in \Re^p$ is a continuous function $\pi : [0,1] \to \Re^p$, such that $\pi(0) = x$ and $\pi(1) = y$. The length of a path π in A, denoted by $\Pi_A(\pi)$, is the value of the following integration

$$\Pi_A(\pi) = \int_0^1 \mu_A(\pi(t)) \left| \frac{d\pi(t)}{dt} \right| dt$$

where $\Pi_A(\pi)$ is the integral of membership values (in A) along π.

2.3 Discrete fuzzy shapes

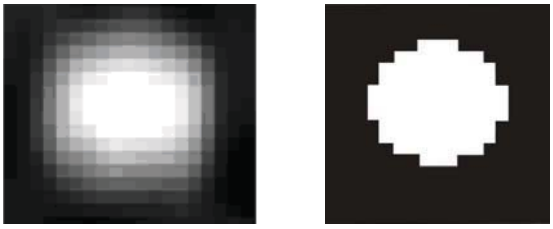
Discrete fuzzy objects can arise from the digitization of scanned images. Generally, the gray-level images will be thresholded to calculate geometrical measures. Since the images or their segments have ill-defined or non-crisp boundaries, it is sometimes appropriate to consider them as fuzzy sets. The concept of fuzzy digital geometry has been introduced by Rosenfeld and plays a key role in many image processing applications: "The standard approach to image analysis and recognition begins by segmenting the image into regions and computing various properties of and relationships among these regions. However, the regions are not always 'crisply' defined; it is sometimes more appropriate to regard them as fuzzy subsets of the image... It is not always obvious how to measure geometrical properties of fuzzy sets, but definitions have been given and basic properties established for a variety of such properties and relationships, including connectedness and surroundedness, convexity, area, perimeter and compactness, extent and diameter" ([12]).

The application areas of fuzzy geometry are image representation, enhancement and segmentation. The process of converting the input image into a fuzzy set by indicating, for each pixel, the degree of membership to the object, is referred to as "fuzzy segmentation". The most straightforward way to perform fuzzy segmentation is to scale grey-levels of an image to be between 0 and 1. Such

ISBN: 978-989-95079-6-8

grey levels reflect the area coverage of a pixel by the object, and can be naturally used as membership values. However, in most cases, more advanced segmentation methods are required, especially since it is rarely sufficient to use only the brightness of pixels to calculate fuzzy membership values. For example, fully segmented image can be generated by combining the optimum automatic thresholding procedure with edge detection to produce continuously connected object border.

The object of interest is represented as a discrete spatial fuzzy subset of a grid. It should be noted, however, that the discrete fuzzy objects obtained from the digitization of scanned images (say, using a grey-level scale) are affected by multiple distortions, due to limited representation resolution. Consequently, their properties are significantly different with respect to those of corresponding continuous fuzzy objects. Figure 2 shows a discrete fuzzy disk (a) and, for comparison, its crisp counterpart (b).



(a) discrete fuzzy disk (b) crisp (binary) disk

Figure 2: Thresholding a fuzzy (gray-level) image: pixels with a membership degree below the threshold are lost

The mapping $\mu_A : \Re^p \to [0,1]$ of a continuous fuzzy shape becomes, by discretization, $\mu_A : Z^p \to \left\{0, \frac{1}{k}, \ldots, 1\right\}$, where k is the maximal number of grey levels available (e.g., k = 255 for 8-bit pixel representation).

A configuration matrix for a discrete p-dimensional fuzzy shape A can be represented by vertical concatenation of its $\alpha-$level contours into a $nk \times p$ block matrix. Each one of the k sub-matrices defined at each level α collects $n \times p$ landmarks: $\mathrm{A}_\alpha = \left(x_1^A(\alpha) \quad x_2^A(\alpha) \quad \ldots \quad x_p^A(\alpha)\right)$, where $\alpha \in \{\alpha_i \mid 0 = \alpha_1 < \cdots < \alpha_i < \cdots < \alpha_k = 1\}$.

A similar $nk \times p$-dimensional configuration matrix can be defined for a crisp shape with multiple (say k) contours.

The shape signature can be also generalized for a discrete fuzzy shape in two possible ways:

- using the distance between the boundary points and the centroid:

$$CD_{discrete_fuzzy-1}(k) = \mu_A(x_c, y_c) + \\ + \sum_{j=1}^{N_k} \delta_k(j) \cdot \mu_A(x_k(j), y_k(j))$$

- as an average signature obtained from the $\alpha-$cuts:

$$CD_{discrete_fuzzy-2}(k) = \frac{1}{\alpha_{total}} \sum_{\alpha=1}^{\alpha_{total}} CD_{\alpha-resampled}(k)$$

where $CD_{\alpha-resampled}(k)$ is the k th sample of the resampled signature obtained for one $\alpha-$cut.

2.4 Lymph nodes as an example of crisp double-contour shapes

The ultrasound image of a lymph node (see figure 3) is a typical example of a crisp double-contour shape. It appears as an ovoid-shaped masse with an echogenic center, representing the medullary, and a peripheral, hypoechogenic cortical region, interrupted on the hyllum, which give him a reniform shape. Usually, the normal lymph nodes present a thin cortical peripheral zone, while benign inflammatory changes in lymphadenitis may enlarge the node but with preservation of the ovoidal shape and of the ratio cortical/medullary thickness less than 1.0. Malignant metastatic or infiltrated nodes are more apparent than normal ones as they become larger, rounder, and more uniformly hypoechoic by the regularly/irregularly thickening of the cortical zone with progressive restriction of the hyperechogenic medullar area. The Computed Aided Diagnosis develops ultrasound applications, especially for breast imaging, but the complete characterization must include the analysis of the satellite lymph nodes appearance. Because of the large variability of the shape and cortical-to-medullary ratio, computer vision applications are needed to make possible the automatic diagnosis, especially in breast cancer screening.

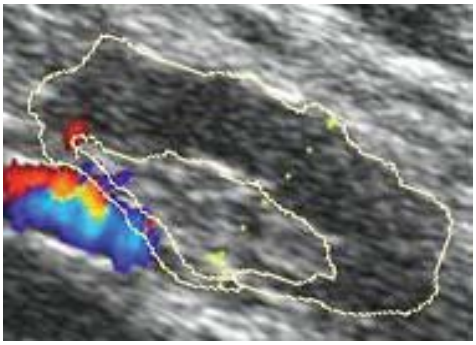


Figure 3: A crisp double-contour shape: the ultrasound image of a lymph node

3 Procrustean shape analysis

A configuration matrix A is not a proper shape descriptor, because it is not pose invariant. For any similarity transformation, i.e. $s \in \Re^+$, $R \in SO(p)$ (the special orthogonal group, i.e. R is $(p \times p)$ matrix, s.t. $R'R = I$) and $t \in \Re^p$, the configuration given by $sAR + 1_p t'$ describes the same shape as A, where 1_p is the $p \times 1$ vector $(1\ 1 \ldots 1)'$. To obtain a true shape representation, location, scale and rotational effects need to be filtered out. This is carried out by shape alignment, i.e. by establishing a *coordinate reference*, commonly known as *pose*. A very popular alignment procedure is *Procrustes shape analysis,* which provides a measure, Procrustes distance, that quantifies the dissimilarity of two configurations, and which is invariant with respect to translation, scaling, and rotation.

ISBN: 978-989-95079-6-8

Procrustes shape analysis also provides a way to define the average shape, the Procrustes mean shape, which can be viewed as a representative class template.

The *Extended Orthogonal Procrustes* (EOP) problem is a least squares method for fitting a given configuration matrix A to another given matrix B. It is based on the functional model $E = sAR + 1_p t' - B$ and consists of minimizing the Procrustes distance between A and B (i.e. $\|E\|_F^2$), under choice of unknown similarity transformation parameters R, t and s. This leads to solving the problem $\min_{R,t,s} E'E$, subject to the orthogonality restriction $R'R = I$.

Generalized Orthogonal Procrustes (GOP) *analysis* is a technique that provides least-squares correspondence of more than two model points. The solution of the problem can be thought as the search of the unknown optimal matrix W (also named *consensus matrix*), defined as follows:

$$M + E_i = \hat{A}_i = s_i A_i R_i + 1_p t_i'; \quad i = 1, \ldots, m$$
$$vec(E_i) \sim N\left(0, \Sigma = \sigma^2 \left(Q_n \otimes Q_p\right)\right)$$

where E_i is the random error matrix in normal distribution, Σ is the covariance matrix, Q_n is the cofactor matrix of the n points, Q_p is the cofactor matrix of the p coordinates of each point, $\otimes$ stands for the Kronecker product, and σ^2 is the variance factor.

Let $C = \sum_{i=1}^{m} \hat{A}_i / m$ be the *geometrical centroid* of the transformed matrices. Therefore, Generalized Orthogonal Procrustes problem can be solved minimizing

$$m \sum_{i=1}^{m} \left\| \hat{A}_i - C \right\|^2 = m \sum_{i=1}^{m} tr\left\{ \left(\hat{A}_i - C\right)' \left(\hat{A}_i - C\right) \right\}$$

Crosilla and Beinat (2002) proved that the shape mean (centroid) C corresponds to the least squares estimation $\hat{M}$ of the true value M: $C = \hat{M} = \sum_{i=1}^{m} \hat{A}_i$.

In the case of *multiple-contour crisp shapes* we can benefit from the *Extended Orthogonal Procrustes* method in order to find mutual distances between shape pairs and from the *Generalized Orthogonal Procrustes* technique in order to estimate the Procrustes mean shape of a collection of shapes. This is illustrated in figures 4 and 5.

On the other hand, dealing with the case of fuzzy shapes needs more advanced Procrustean techniques, which allow us to consider *weighted distances* between points placed on α – level contours with different membership degrees. This leads to solve a *Weighted Orthogonal Procrustes* (WOP) problem.

A weighting matrix W of the residual E (defined above) is now introduced and the minimization problem becomes:

$$\min \|W E\|_F^2$$

subject to orthogonality restriction $R'R = I$, $\det(R) = 1$.

Typically, an iterative method is needed to derive a solution to WOP.

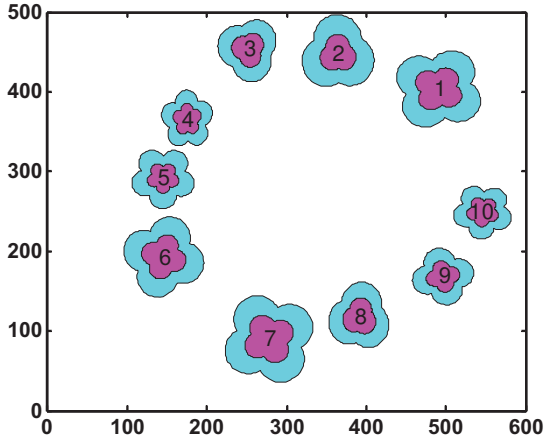


Figure 4: Ten double-contour star-shaped 2D objects with 3, 4 and 5 "lobes"

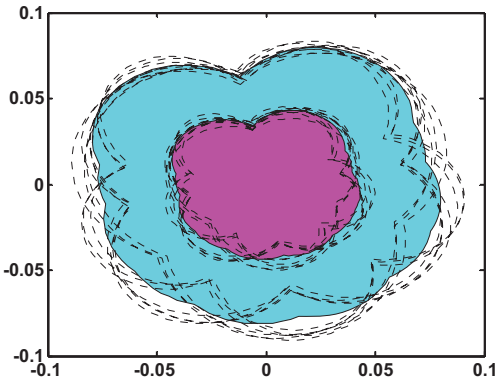


Figure 5: Procrustes mean shape (shape centroid)

4 A generalization of k-means algorithm for clustering fuzzy shapes

K-means is a commonly used data clustering for partitioning data points into disjoint groups such that data points belonging to same cluster are similar, while data points belonging to different clusters are dissimilar. The main idea is to define *k* centroids, one for each cluster, and to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, an early groupage is done. Next, we need to re-calculate *k* new centroids of the clusters resulting from the previous step. After we have these *k* new centroids, a new binding has to be done between the same data set points and the nearest new centroid. We continue this loop until no more changes are done.

Clustering of objects or images of objects, according to the shapes of their boundaries is of a key importance in computer vision and pattern recognition. This paper was intended to pay attention to this reason by proposing a generalization of K-means algorithm in order to integrate Procrustean metrics and full mean shape estimation, in a way

ISBN: 978-989-95079-6-8

making it able of clustering objects with either multiple or fuzzy contours.
We first present the algorithm in pseudo-code, as follows:

- Make initial guesses for the mean shapes v_1, v_2, …, v_k, by choosing the first k shapes from a random permutation.
- While any change still exists in any mean shape
  - Calculate all pair-wise Procrustes distances between shapes using the *Extended Orthogonal Procrustes* algorithm
  - Use the estimated mean shapes to assign the shape samples into clusters
  - For i from 1 to k
    - Replace v_i with the mean shape of all of the samples for cluster i, using the *Generalized Orthogonal Procrustes* algorithm
  - end_for
- end_while

The resulting mean shapes for each one of the 3 clusters are shown below.

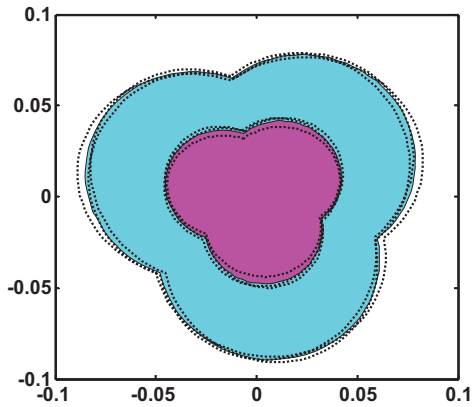


Figure 8: First cluster. Mean shape and three cluster members: {2, 3, 8}

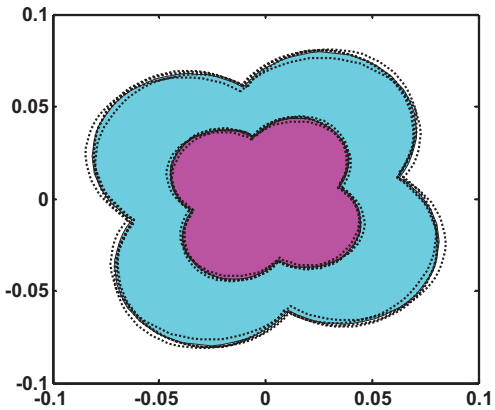


Figure 9: Second cluster. Mean shape and four cluster members: {1, 6, 7, 9}

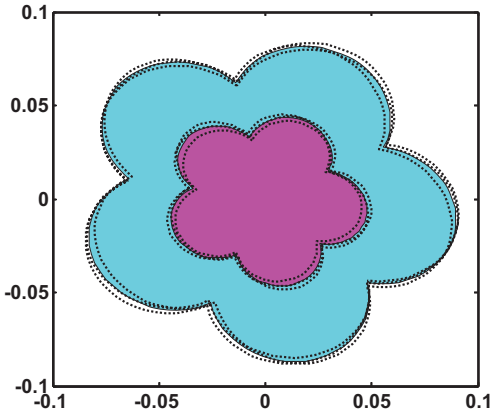


Figure 10: Third cluster. Mean shape and three cluster members: {4, 5, 10}

Our method is thus graphically validated.
As an alternative, one can use a "linkage" method to perform hierarchical shape clustering, i.e. to create a hierarchical tree of clusters starting from the symmetric matrix of Procrustean mutual distances between pairs of shapes. We obtained the same clusters as in the case of using K-means: {6, 7, 9, 1}, {2, 3, 8}, {4, 5, 10}. The dendrogram is shown below.

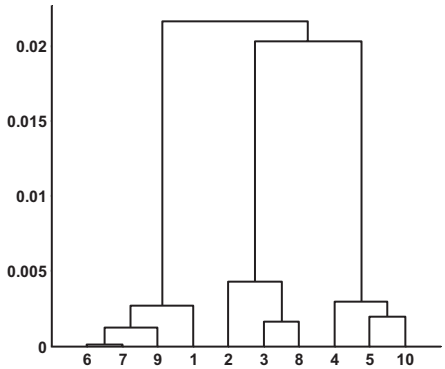


Figure 11: The dendrogram

5 Further remarks

We propose a two stage procedure for landmark placement. In the first stage, landmarks are placed arbitrarily (figure 10).

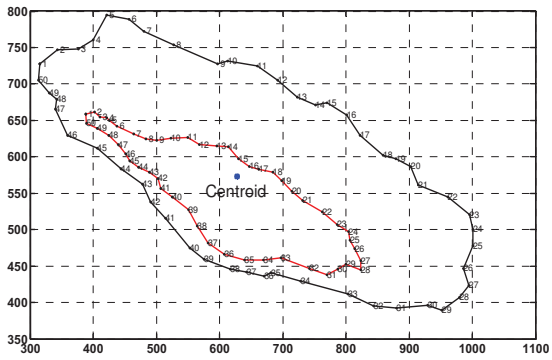


Figure 10: First stage: landmarks are placed arbitrarily

ISBN: 978-989-95079-6-8

However, we need a frame of reference to compare and display the differences in shape. Thus, we use Principal Component Analysis (PCA), which uses similarity transformations to produce a standard shape orientation, based on decomposing the overall variation of data. Each axis on a PCA representation of transformed data is an eigenvector of the covariance matrix of shape variables. In this morphospace, the first axis accounts for maximum variation in the sample, with further axes representing further decreasing variations (figure 11).

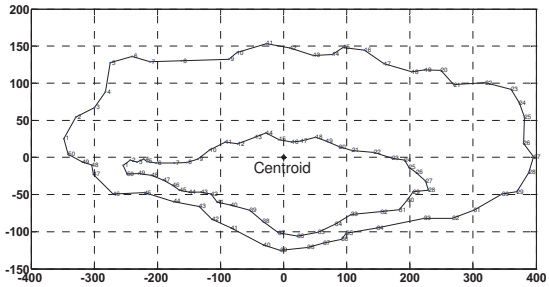


Figure 11: PCA-based alignment along the axes of maximum variation

In the second stage we replace all landmarks based on a standard procedure. The right most object point landmark is taken in the horizontal direction from the centroid. For each pair of α – level contours and for each pair of landmarks on these contours, distances are computed starting from the corresponding points, where corresponding points are those located in the same direction from the centroid. It is essential to use an angular landmark placement procedure on the boundaries, to provide appropriate correspondence between the points on the (parts) of the boundaries having different lengths, when the corresponding boundary subparts of different α – level contours are matched (figure 12).

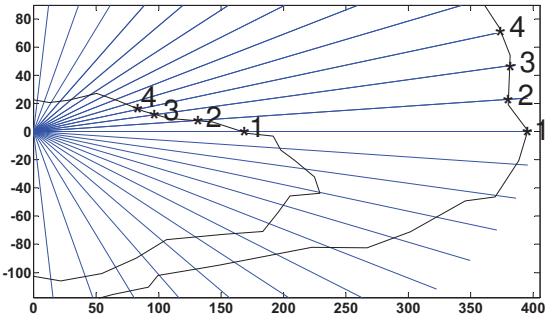


Figure 12: Corresponding points placed on radial directions from the centroid, starting from the right most object point

References

[1] I. Bloch and H. Maître. Fuzzy mathematical morphologies: A comparative study. Pattern Recognition, 28(9):1341–1387, 1995.

[2] J. Buckley and E. Eslami. Fuzzy plane geometry I: Points and lines. Fuzzy Sets and Systems, 86:179–187, 1997.

[3] P. Diamond. A note on fuzzy starshaped fuzzy sets. Fuzzy Sets and Systems, 37:193–199, 1990.

[4] J. Chanussot, I. Nyström, and N. Sladoje. Shape signatures of fuzzy star-shaped set based on distance from the centroid, *Pattern Recognition Letters*, 26(2005) 735-746

[5] B. Chaudhuri. Some shape definitions in fuzzy geometry of space. Pattern Recognition Letters, 12:531–535, 1991.

[6] F. Crosilla, A. Beinat. Use of generalized Procrustes analysis for the photogrammetric block adjustment by independent models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 56(3), 2002, pp. 195-209.

[7] I.L. Dryden, K.V. Mardia. *Statistical shape analysis*. John Wiley and Sons, Chichester, England, 1998, 83-107.

[8] J. Hartigan, M. Wang. A K-means clustering algorithm. *Applied Statistics*, 28, 1979, 100–108.

[9] M. Machtegael and E. Kerre. Connections between binary, grey-scale and fuzzy mathematical morphologies. Fuzzy Sets and Systems, 124:73–85, 2001.

[10] A. Rosenfeld. Fuzzy digital topology. Information and Control, 40:76–87, 1979.

[11] A. Rosenfeld. The diameter of a fuzzy set. Fuzzy Sets and Systems, 13:241–246, 1984.

[12] A. Rosenfeld. The fuzzy geometry of image subsets. Pattern Recognition Letters, 2:311–317, 1984.

[13] A. Rosenfeld and S. Haber. The perimeter of a fuzzy subset. Pattern Recognition, 18:125–130, 1985.

[14] A. Rosenfeld. Fuzzy geometry: An updated overview. *Information Sciences,* vol. 110, 1998, 127–133.

[15] P. K. Saha, F. W. Wehrli, and B. R. Gomberg. Fuzzy distance transform: Theory, algorithms, and applications. *Computer Vision and Image Understanding*, 86:171–190, 2002.

[16] P.H. Schoenemann. A generalized solution of the orthogonal Procrustes problem. *Psychometrika*, 31(1) , 1966, pp. 1-10.

[17] P.H. Schoenemann, R. Carroll. Fitting one matrix to another under choice of a central dilation and a rigid motion. *Psychometrika*, 35(2), 1970, pp. 245-255.

ISBN: 978-989-95079-6-8

Development of a Fuzzy Expert System for a Nutritional Guidance Application

Petri Heinonen, Marjo Mannelin, Hannu Iskala[1] Aki Sorsa, Esko Juuso [2]

1 Flow-Team Oy
Oulu, Finland
2 Control Engineering Laboratory, University of Oulu
Oulu, Finland
Email: firstname.lastname@flow-team.fi, firstname.lastname@oulu.fi

Abstract— *The importance of nutritional guidance grows as nutritional problems, such as obesity and type-2 diabetes, are becoming more common. Nutritional guidance is carried out by mapping the nutritional state of an individual using a food diary and by comparing the nutrition intake levels to the recommended reference values. Typically, the expert knowledge of a nutritionist is required to balance the diet.*

This paper presents a fuzzy expert system for a nutritional guidance application. Expert knowledge acquisition is carried out using variable tabulation creating the basis for the rulebase of the fuzzy system. The recommended nutrition intake values are used to generate membership functions for fuzzification. The development of a Mamdani-type fuzzy system is illustrated using a hierarchical structure. The first level system refers to groups of similar foods and the second level model defines the added and reduced foodstuffs.

The validation of the fuzzy model was carried out in three phases: first two types of sensitivity analysis were performed, and then the output was analysed with expert knowledge.. The results from the validation schemes are promising.

Keywords— expert system, fuzzy logic, nutritional guidance

1 Introduction

Unhealthy diets and physical inactivity are the main causes of non-communicable diseases such as Cardiovascular disease, type 2 diabetes, and osteoporosis. Obesity is a large and growing problem in developed countries. [1] There have been regular surveys of Finnish health behaviour since 1978. According to the latest survey, over 50 percent of Finnish men and women are overweight [2]. The best way to avoid nutritional problems is to have a well-balanced diet and to ensure the sufficient intake of essential nutrients [3].

Individual nutrition guidance is becoming more and more important as problems with obesity and type-2 diabetes are increasing rapidly. The nutrition guidance is based on information from food diaries kept by the users where they record all the foods they consume. This information is converted into nutrient intake levels with the help of a food composition database. The intake levels are compared to the recommended intake levels and the expert knowledge of a nutritionist is used to balance the diet to meet the recommendations. There are restrictions, for example allergies, diseases, and special diets which must also be taken in to account when balancing diets.

Some software has been developed for the purpose of nutritional guidance. For example, a French research group developed a nutrition software application for hospital use [4]. They use fuzzy arithmetic to handle imprecise and uncertain information in the computations. Fuzzy logic is first used by the application to determine how well the current meal is balanced, after which a heuristic search is performed to provide nutritional guidance. Fuzzy logic is also used in a nutritional guidance application in [5]. They define the optimality of a nutrient intake level as a fuzzy set (Fig. 1). The values of the points from "a" to "e" in Fig. 1 are based on the recommended intake levels. In [5], the Prerow value (PV) is used to measure how well the diet is balanced. In the calculation of PV, the nutrient with the lowest value has the strongest influence on the result. The gradient optimization method is used to find the shortest route to maximize PV in the discrete multidimensional search space.

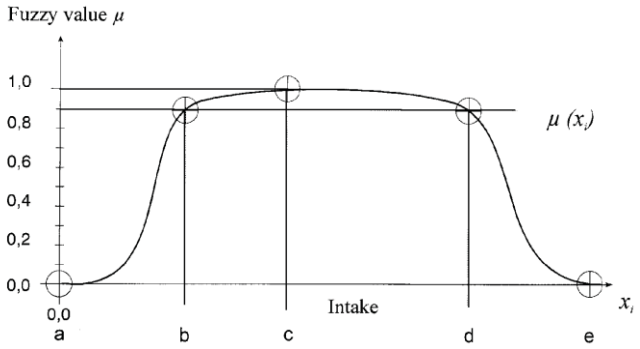


Figure 1: Membership function of optimal nutrient intake level [5].

In [6], the optimization results of two approaches to the nutrition calculation are compared. The first approach considers the use of non-fuzzy numbers and the second approach the use of fuzzy numbers in calculations and optimization. They concluded that optimization with plain numbers always yields results which meet the selected recommendations. The downside of this method is that it recommends drastic changes to eating habits. According to [6], the optimization with fuzzy numbers is not Pareto optimal. The results from fuzzy optimization are sometimes unbalanced. While the most of the nutrient intake levels are on an optimal level, some intake levels are either insufficient or excessive, and some are even less in balance

ISBN: 978-989-95079-6-8

than initially. However, the instructions given are closer to the original eating habits and are therefore easier to follow.

Dietary guidance applications usually present their results in the form of large tables where the suitable amounts of each nutrient are listed. However, this information is not very informative if you want to know what foods you should cut back on and what should you eat more. Therefore, there is a need for software which interprets the nutrient level results into the foodstuff level guidance.

The aim of this study is to develop a fuzzy model for the Nutri-Flow dietary analysis application. The Nutri-Flow application provides nutrition guidance in a user-friendly way: the balanced diet does not alter the original diet too much and the recommended adjustments to the diet are given on a foodstuff level in text format. This paper presents the fuzzy model developed for the Nutri-Flow application.

2 Nutrition recommendations and guidance

The prerequisite for good health is a well balanced diet. In a well balanced diet energy intake is in proportion to energy consumption and the nutrition intake is balanced. This often implies increasing the amount of dietary fibre in the carbohydrate intake and reducing the intake of purified sugars. Dietary guidance is based on the calculation of the user's nutrient intake levels and the nutrition recommendations.

2.1 Nutrition recommendations

The first nutrition recommendations were given to prevent deficiency disorders. The basis of traditional nutrition recommendations is physiological data on nutrient intake requirements. The Finnish Nutrition Recommendations are based on the Nordic recommendations which in turn are based on scientific data. The main objective of the national recommendations is to get Finnish people to balance their diets and to improve their health. The recommendations are defined for healthy and moderately physically active people. Therefore, they should only be used as guidelines for individuals and with great precaution. Nutrition recommendations represent the average daily values of long term intake. [3] Public dietary guidelines are criticized in [7]. They conclude that an adequate scientific background must first be established before public dietary guidelines are declared.

The nutrition recommendations are given as reference values which are defined separately for each nutrient. If there is evidence that the nutrient has an effect on health, a reference value is defined for it. The reference values are determined using estimates and evaluations based on current knowledge of nutritional needs and health. Thus, there will always be a degree of uncertainty and imprecision in the values. The reference values should not be considered as fixed points. Reference values distinguish between the lower intake level (LI), the average requirement (AR), the recommended intake (RI), and the upper intake level (UL). [3] The definitions of the reference values are:

- LI – the minimum requirement of a nutrient without the deficiency symptoms,
- AR – the nutrient intake level that is adequate for maintaining good health for healthy individuals,
- RI – the nutrient intake level that is adequate for maintaining the good nutritional status among 98 % of healthy individuals. RI is calculated based on AR and
- UL – an estimate of the highest level of intake that carries no appreciable risk of adverse health effects [8].

Depending on the nutrient, some reference values may be missing. For example, only RI is defined for magnesium, because the effects of magnesium are not that well known. Whereas, all the reference values have been set for vitamin B_6, because there is reliable scientific data available on how vitamin B_6 affects health with different intake levels. The reference values for vitamin B_6 are presented in Table 1. [3]

Table 1: Reference values for vitamin B_6 [3].

Vitamin B_6 level	Women [mg/d]	Men [mg/d]
LI	0.8	1
AR	1.0	1.3
RI	1.2	1.6
UL	25	25

2.2 Dietary guidance

Nutrient intake is usually assessed using a food composition databases. Food composition information is needed when calculating the composition of menus and recipes. Primary sources of food composition data are government databases, databases provided by academic and other institutions, the food industry, and scientific literature [9]. Finnish food composition information is collected into the Fineli® Nutrition Database which is provided by the National Public Health Institute. The commercial version of the database provides information on 2000 foods and 68 nutrients. The composition data consists of averaged information on Finnish food, and thus the composition of individual foods can vary depending on the given data.

Food composition data and the expert knowledge of a nutritionist are used to balance the user's diet on the basis of a food diary and personal nutrition intake recommendations (reference values). The national recommendations are adjusted individually according to the user's health data to form personal recommendations for evaluation. Typically, a nutritionist is needed to compare the data and to balance a suitable diet. The task is very complex because the information on a nutrient level is not useful in its self and must be converted onto a foodstuff level. In other words, people should know what foods they should eat more and what less. A balanced diet should also be reasonable in comparison to the original diet. Furthermore, some restrictions, for example allergies, diseases, and special diets, must be taken in to account when balancing diets.

3 Fuzzy logic

A rule based fuzzy logic system (FLS) consists of three interconnected main blocks for mapping crisp inputs to crisp outputs. The main blocks are the fuzzifier, the inference machine with the rulebase, and the defuzzifier [10]. The

ISBN: 978-989-95079-6-8

design parameters for a FLS are scaling factors, fuzzification and defuzzification methods, rulebase, and membership function construction, and representation.

Fuzzification is needed to convert the crisp input data into fuzzy sets for the inference machine. The fuzzification is carried out by using membership functions with defined shapes and parameters. In some cases, the input value needs to be normalised before fuzzification.

The mapping from inputs to outputs is done in the fuzzy inference module. The module contains all the necessary information for forming the output of the system. The heart of the inference module is the rulebase. The definition data used to generate and collect the output of the system is stored in the inference machine. The selected implication method and the aggregation method are a part of the definition data. The rulebase is a collection of IF-THEN rules which represent the expert information used to define the behaviour of the system. The antecedent of the rule defines the state of the system and also determines if the rule is triggered. The consequent of the rule is the output of each rule guiding the system output towards the solution. The membership grade of the rule antecedent is utilized through implication to produce the output of each rule. The minimum and the product are typically used as implication methods in Mamdani-type fuzzy systems. In Mamdani-type fuzzy systems, the output of each rule is fuzzy and aggregation is required to obtain the fuzzy output of the system. The fuzzy output needs to be defuzzified to obtain the crisp output of the system. In Takagi-Sugeno and Singleton -type fuzzy systems, the rule outputs are linear functions or constants, respectively. The crisp output of the system in these cases is typically obtained by taking the weighted average of the rule outputs. [11]

The defuzzifier module is needed if the output of the inference machine is fuzzy. There are several defuzzification methods. The most commonly used method is the centre of gravity method (COG). Other typical methods are the first (FOM), the last (LOM), and the middle (MOM) of maxima, and the first (FOS), the last (FOL), and the middle (FOM) of support. The defuzzification method should be chosen carefully because it affects the crisp result considerably. [12]

4 Data acquisition

The Nutri-Flow application uses a commercial version of the Fineli® Nutrition Database. The database is used to convert the information in the food diaries to averaged daily nutrient intake levels. The Finnish recommended nutrient levels (the reference values) are used to define the membership functions for the system inputs. The reference values are updated by nutritionist, and the personal recommendations are revised according to the user's health data.

Expert data acquisition was one of the most challenging parts of this study. The first task was to find an efficient way to convert expert knowledge into the fuzzy rule IF-THEN language. Variable tabulation proved to be a very good tool for this. With tabulation, all the input set conditions with appropriate consequents can be examined separately. Variable tabulation is possible because the number of variables in most of the rules is two or three. If the number of variables connected to a single rule is greater than three, tabulation becomes difficult. Table 2 shows an example of tabulation for carbohydrate and vitamin-C. The "fruits and berries" group is the consequent variable. The notations -, 0, and + represent the fuzzy values "too little", "ideal", and "too much", respectively. In this case, the consequents are limited to "add" and "no action". Naturally, the action "reduce" is also used, however, it is not feasible for the output variable "fruits and berries" group.

Table 2: Variable tabulation for "fruits and berries" group.

	Carbohydrate		
Vitamin-C	-	0	+
-	add	no action	no action
0	no action	no action	no action
+	no action	no action	no action

5 Developed model

The developed fuzzy model is used in the Nutri-Flow application. The schematic diagram of the overall system is provided in Fig. 2. The calculations start by analysing the user's nutrient intake levels based on the food diary and the Fineli® database. The intake levels are then fed into the developed fuzzy system. The fuzzy system is divided into two hierarchical levels. The first level model uses the nutrient intake levels as inputs and produces output on the level of main food groups such as "fruits and berries", "vegetables" and "dairy products". These outputs are then fed into the 2nd level model together with the nutrient intake levels. The 2nd level model produces outputs on a food subgroup level. Fig. 3 illustrates the division of foods into main groups and subgroups.

The fuzzy system requires user information (personal recommendations) to define the membership functions for the input variables. The output variables of the complete fuzzy system are continuous. The behaviour of the fuzzy system is not regulated by user specific information, such as the original diet, allergies, and diseases. The user related specific information is used in the optimization process. As depicted in Fig. 2, the focus of the future development of the Nutri-Flow application is to find an optimization algorithm which takes the outputs of the complete fuzzy system and the user specific information into account. Also, the nutrient density of the recommended foods should be considered when evaluating the final results of the system. Furthermore, the optimization algorithm would be used to search for solutions within a discrete space. The Nutri-Flow application uses the jFuzzyLogic java package for fuzzy inference. This Chapter presents the properties of the fuzzy system in more detail.

5.1 Membership functions

The values of the system inputs are the average daily intake levels of nutrients. The values are crisp numbers and they need to be fuzzified for the fuzzy inference machine. To map the crisp input values into the fuzzy domain, three linguistic variables are used: "too little", "ideal", and "too much" (Fig. 4). The fuzzy output variables are "reduce", "no action", and "add". The selected forms of the MFs are trapezoidal or triangular. The points "A", "B", and "C"

ISBN: 978-989-95079-6-8

correspond to LI, RI, and UL, respectively. If the points "A", "B", and "C" depicted in Fig. 4 are defined, three MFs are used to fuzzify the selected input variables. For example, in case of alcohol, there is no LI value and point “A” is not defined, thus the "too little" MF is not used. It should be noticed that the values "A", "B", and "C" are influenced by the user’s gender, age, mass, height, body mass index, and the level of physical activity. Some other factors, such as diseases and pregnancy, also affect the reference values. The output membership functions are produced similarly but the points from "A" to "C" are kept fixed.

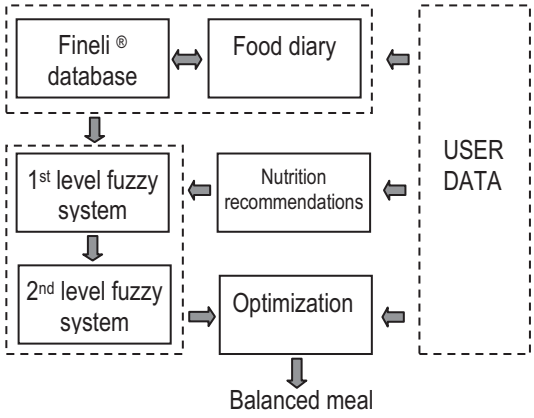


Figure 2: Schematic diagram of overall system.

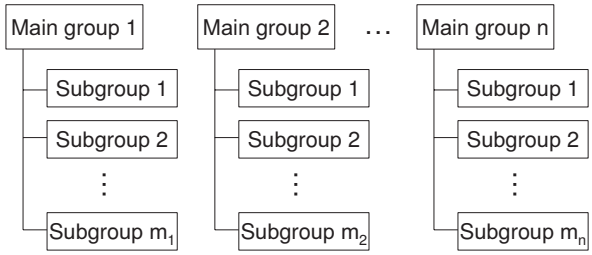


Figure 3: Division of foods to main groups and subgroups.

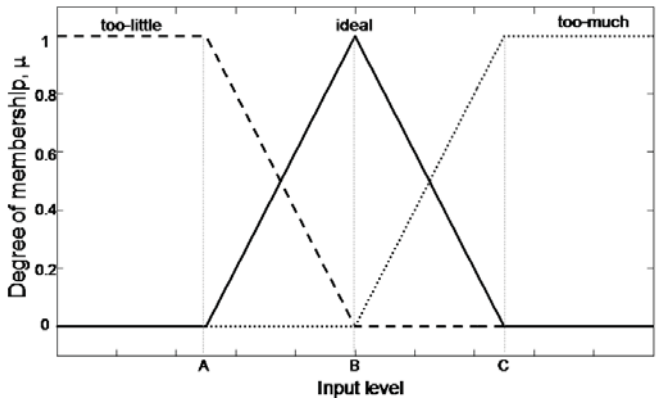


Figure 4: Input membership functions.

5.2 Rulebase and fuzzy inference

The most important and most time consuming phase of this study was building the rulebase. It is the heart of the fuzzy system, and therefore it must work correctly. Every input set condition must trigger at least one rule for each of the output variables. In this study, the rulebase was built on two levels. The first level of the fuzzy system operates on the main food group level and the second one on the subgroup level. The hierarchical structure is used because of the complexity of the system. The purpose of the first level is to roughly evaluate the nutritional status of the user and to produce additional, informative variables for the second level model, which in turn contains more detailed information but uses the additional variables to handle the complexity of the overall system.

As mentioned earlier, expert knowledge is obtained through variable tabulation. As an example, the tabulation of carbohydrate and vitamin-C was presented in Table 2. The consequent variable in this example was the "fruits and berries" -group. The first level rules corresponding to Table 2 are

Rule 1: IF Vitamin-C is "too little" AND Carbohydrate is "too little" THEN fruits and berries group is "add"
Rule 2: IF Vitamin-C is NOT "too little" OR Carbohydrate is NOT "too little" THEN fruits and berries group is "no action"

The output of the first level model directly influences the structure of the second level rules. The second level rules corresponding to the "fruits and berries" group have a general structure which can be written as

IF fruits and berries group is "Add" AND additional propositions THEN ...

If the consequent of the first level model is "no action", then the consequent of the corresponding food subgroups determined on the second level is also "no action".

5.3 Design parameters

In this study, the variables are not scaled and thus no scaling factors are needed. The fuzzy inference uses the minimum as an implication method and the maximum as an aggregation method. The used T-norm is a minimum. The consequents cannot be put in the form of an equation or defined by a single number. Therefore, a Mamdani-type inference is used with COG as the defuzzification method.

6 Results and discussion

All the calculations in this study were carried out in Matlab®. The data for membership function generation was extracted from the Nutri-Flow application. The data contains personal nutrient intake recommendations and the nutrient intake levels of the selected test cases. Also, a set of test cases with varying dietary habits were selected from the software for testing purposes.

6.1 Model validation with expert knowledge

The validation of the developed model was challenging because no "right answers" exist, and thus no meaningful error describing values (such as the sum of squared error) could be calculated. Therefore, only qualitative validation was carried out.

The model outputs are calculated using the developed model with the Nutri-Flow application. The output of the application is in text form and contains information on the

ISBN: 978-989-95079-6-8

diet and instructions on how the diet should be improved. The output is obtained in text form even though the optimization algorithm mentioned earlier has not yet been implemented. The validity of the output of each test case was evaluated by a nutritionist with promising results.

6.2 Uniformly distributed sensitivity test

The input-output behaviour of the model was tested with an uniformly distributed sensitivity test. The system sensitivity of each input variable was examined as follows: the supplies of all nutrients except the one under examination are fixed on the ideal intake level and the test variable gets random values in the range of [0, UL]. The number of test rounds conducted on one variable was 500. The test range is set so that the variable gets values from all the fuzzy sets. The quadratic error of the output is recorded as a sum of the output variables. The test results show how large the independent impact of an input variable is on the output of the whole system.

The value resulting from the test is the sum of the quadratic error which is the Euclidean distance from the system's ideal state. The test result value is zero for a variable with no individual impact on the output variable. The larger the value of a variable the greater its individual impact on the output variable.

A variable affects the output if it is the only variable in the antecedent part or the OR-connective is used. Also, if the tested variable is connected with an AND-connective to other variables and the other variables have a NOT-term in the proposition, the tested variable affects the output.

After the test, the results were given to a nutritionist for analysis. The analysis showed that the input variables only affected the desired output variables with the desired magnitude. An example of the results from the sensitivity analysis is shown in Table 4.

Table 4: Example results from sensitivity analysis.

	output 1	output 2	output 3	output 4
hard fat	0	0	1	0
soft fat	0	0	1	1
fibre	0.51	0	0.51	0

6.3 Normally distributed sensitivity test

Another type of sensitivity analysis was carried out by changing a group of input variables randomly at the same time. In this approach, the values of the input variables vary around the ideal value and they are taken from a normal distribution. The distributions of the output variables are obtained as a result and they are visually inspected. Fig. 5 illustrates a random test set of a variable. The mean and the standard deviation of the random changes are defined individually for all the input variables. The outputs of the random tests are analyzed by visually inspecting the distributions of the output variables. An example of the distribution of an output variable is given in Fig. 6. From the figure, it can be seen that the histogram does not quite follow a normal distribution. That is anticipated because no adding rules are associated with alcoholic beverages. Thus the histogram of the output variable is slightly skewed, and only values equal to or smaller than zero are obtained. Negative values of the output variables indicate that the consumption of a corresponding foodstuff should be reduced.

Some statistical values can also be calculated to evaluate the shape of the histograms. Mean, standard deviation, skewness, and kurtosis are used in this study. The mean and standard deviation values give an overview of the output data behaviour. The skewness is a measure of the asymmetry of the data around the mean value. If the data is spread out more to the left, the skewness value is negative and if the data is spread out more to the right, the skewness value is positive. For normal distribution the skewness value is zero. The distribution can also be evaluated with the kurtosis value which is a measure of how outlier-prone the distribution is. Table 5 presents the statistical values for some output variables. The evaluation of the histograms, both visually and on the basis of statistical values, shows that the developed model behaves correctly.

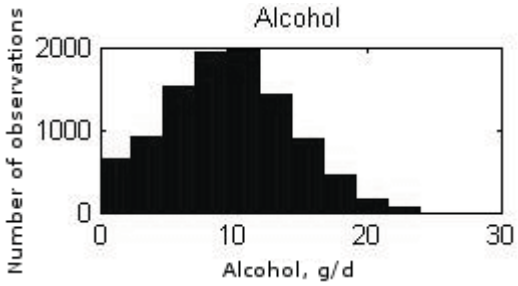


Figure 5: Test set for intake levels of alcohol.

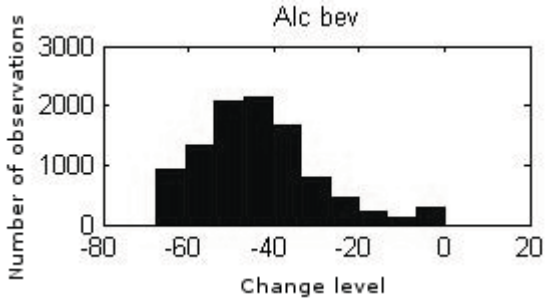


Figure 6: Histogram of output variable.

Table 5: Example statistical values of output variables.

	Mean	Std	Skewness	Kurtosis
Fruits and berries	11.9	19.1	1.1	2.5
Vegetables	-5.0	26.0	-0.2	2.3
Cereals and bakeries	-39.2	17.5	1.4	3.4
Alcoholic beverages	-43.5	14.1	0.9	4.1

6.4 Evaluation of the results

The results obtained from the performed validation are promising but more validation with more efficient methods is required. All of the validation procedures showed that the model performs in correctly, but no solid proof for this was obtained. The addition of the optimization algorithm will

ISBN: 978-989-95079-6-8

make validation easier because the numerical values (distance from the original and the recommended diet) can be calculated.

7 Conclusions

The importance of nutritional guidance is growing as nutritional problems, such as obesity and type-2 diabetes, are becoming more common. Nutrition guidance is carried out by mapping the nutrition intake levels from the user's food diary and comparing them to the recommended levels. Typically, balancing the diet requires expert knowledge on nutrition.

This paper presented the development of a fuzzy expert system for a nutrition guidance application. The expert knowledge acquisition was carried out using variable tabulation, thus creating a basis for the development of the rulebase for the fuzzy system. The recommended nutrition intake values were used to generate the membership functions for fuzzification. A Mamdani-type fuzzy system was developed with a hierarchical structure. The first level system only referred to the groups of similar foods while the second level model referred to food subgroups.

The validation of the fuzzy model was carried out in three phases: first two types of sensitivity analysis were performed, and then the output was analysed with expert knowledge. The results from all the validation schemes are promising.

References

[1] A. Waxman, WHO's global strategy on diet, physical activity and health: response to a worldwide epidemic of non-communicable diseases. Scandinavian Journal of Nutrition, 48(2):58, 2004.

[2] Peltonen et al. Kansallinen FINRISKI 2007-terveystutkimus, Tutkimuksen toteutus ja tulokset. Kansanterveyslaitoksen julkaisuja B34/2008, Helsinki, National Public Health Institute, 2008. (In Finnish)

[3] Nordic Council of Ministers, Nordic Nutrition Recommendations. Nordic Council of Ministers, Copenhagen, 4th, 435, 2004.

[4] J-C. Buisson and A. Garel, Balancing Meals Using Fuzzy Arithmetic and Heuristic Search Algorithms. IEEE Transactions on Fuzzy Systems, 11(1), Feb 2003.

[5] Wirsam et al. Fuzzy sets and fuzzy decision making in nutrition. European Journal of Clinical nutrition, 51(5): 286-296, 1997.

[6] Gedrich et al. How optimal are computer-calculated optimal diets? European Journal of Clinical Nutrition, 53: 309-318, 1999.

[7] Marantz et al. A Call For Higher Standards of Evidence for Dietary Guidelines. American Journal of Preventive Medicine, 34(3): 234-240, 2008.

[8] Institute of Medicine, Dietary Reference Intakes: applications in dietary planning. Washington D.C.: National Academies Press, 2003.

[9] Pennington et al. Food Composition Data: the foundation of dietic practice and research. Journal of the American Dietic Association, 107(12): 2105-2113, 2007

[10] J.M. Mendel, Fuzzy logic systems for engineering: a tutorial. Proceedings of the IEEE, 83(3): 345-377, 1995.

[11] Driankov et al. An Introduction to Fuzzy Control. Springer, Berlin, 2.rev, 316, 1996.

[12] X. Liu, Parameterized defuzzification with maximum entropy weighting function- Another view of the weighting function expectation method. Mathematical and Computer Modelling, 45(1-2): 177-188, 2007.

ISBN: 978-989-95079-6-8

Possible and Necessary h-indices

Marek Gągolewski[1,2] Przemysław Grzegorzewski[1,2]

1. Systems Research Institute, Polish Academy of Sciences
ul. Newelska 6, 01-447 Warsaw, Poland
2. Faculty of Mathematics and Information Science, Warsaw University of Technology
pl. Politechniki 1, 00-661 Warsaw, Poland
Email: {gagolews,pgrzeg}@ibspan.waw.pl

Abstract— *The problem of measuring scientific impact is considered. A class of so-called p-sphere indices, which generalize the well known Hirsch index, is used to construct a possibility measure of scientific impact. This measure might be treated as a starting point for prediction of future index values or for dealing with right-censored bibliometric data.*

Keywords— Hirsch's h-index, p-sphere indices, scientific impact, possibility theory, scientometrics.

1 Introduction

Fair and objective assessment methods of individual scientists had become the focus of scientometricians' attention since the very beginning of their discipline. A quantitative expression, i.e. measurement, of some publication-citation process characteristics is assumed to be a predictor of broadly conceived scientific competence.

Among the most popular scientific impact indicators is the h-index, proposed by J. Hirsch in 2005 [1]. It has been defined as follows. An author who had published n papers has the Hirsch index equal to H, if each of his H publications were cited at least H times, and each of the other $n - H$ items were cited no more than H times. This simple indicator quickly received much attention in the academic community [2, 3] and started to be a subject of intensive research. It was noted (see for example [4, 5]) that contrary to earlier approaches, i.e. publication count, citation count etc., the measure both concerns productivity and impact of an individual.

Many modifications of the h-index were later proposed, e.g. Egghe's g-index [6], Kosmulski's $h(2)$-index [7], Jin's R-index [8] or Schreiber's h_m-index [9]. It was also a matter of more formal studies (e.g. [10, 11, 12, 13, 14, 15]). It is worth noting that the h-index can be expressed as the Sugeno integral of some function with respect to a fuzzy counting measure [16]. For more details we refer the reader to the extensive scientometric literature.

2 Difficulties related to the h-index

From now on $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ denotes the set of all natural numbers and zero, while $\mathbb{R}_0^+ = \mathbb{R}^+ \cup \{0\}$ stands for the set of all nonnegative real numbers. For any sets X, Y and Z ($Z \subset X$) and $f : X \to Y$ by $f|_Z$ we mean a mapping satisfying $f|_Z(x) = \mathbf{1}_{x \in Z} f(x)$, where $\mathbf{1}$ is an indicator function.

Let us assume that an individual has published exactly $n \in \mathbb{N}$ papers. The list of papers generates an ordered *citation sequence* $\mathbf{C} = (c_1, c_2, \ldots, c_n)$ such that $c_i \geq c_j$ for $1 \leq i < j \leq n$, where $c_i \in \mathbb{N}_0$ is the number of unique citations received by the i-th article.

The problem with the h-index is twofold. Firstly, it assumes perfect knowledge of the author's citation sequence. In practice we gather bibliometric data from large online academic services, such as Thomson *Web of Science*, Elsevier *Scopus* or Google *Scholar*. The coverage of all digital libraries is limited, so in most cases we are dealing with right-censored data.

On the other hand, citing is a dynamic process. If you were applying for academic tenure and were asked to determine your h-value, would you be sure that the index is not about to increase in a while? That is because the Hirsch coefficient totally ignores the number of citations received by publications represented by $c_1, c_2, \ldots, c_H$ (we only know their citation counts are $\geq H$) and how close to H are $c_{H+1}, c_{H+2}, \ldots, c_n$.

As an illustration of the raised issues, consider the following citation sequences: $\mathbf{C}_1 = (4, 4, 4, 4, 0, 0, 0)$ and $\mathbf{C}_2 = (10, 9, 8, 7, 4, 4, 3, 1)$. Both have the h-index of 4, but in the latter case there is not much needed for $\mathbf{C}_2$'s h to increase even to the value of 6. Such property of the citation sequence could be called *saturation* or even *instability*.

In the next section we recall necessary information on a class of so-called "geometric" scientific impact indices [17]. In Section 4 we suggest how to evaluate the h-index stability by means of possibility theory. In our model a given, but usually uncertain, citation sequence will be treated as a source of information to estimate possible and necessary index values. Section 5 illustrates the construction of the suggested possibility measure for individual's h-index.

3 p-sphere indices

Let us recall some definitions from the paper [17]. The following function, based on the citation sequence, will be useful.

Definition 1. A *citation function* based on a citation sequence $\mathbf{C}$ is a mapping $\pi_{\mathbf{C}} : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ given by

$$\pi_{\mathbf{C}}(x) = \begin{cases} c_i & \text{if} \quad x \in [i-1, i), \quad i = 1, 2, \ldots, n, \\ 0 & \text{if} \quad x \geq n. \end{cases} \tag{1}$$

An exemplary citation function for sequence $\mathbf{C} = (5, 4, 3, 3, 3, 1)$ is depicted in Fig. 1.

Definition 2. Consider a function $f : \mathbb{R}_0^+ \to \mathbb{R}_0^+$. We say that the citation function π *dominates* the function f (denoted $\pi \succeq f$) if $\pi(x) \geq f(x)$ for every $x \in \mathbb{R}_0^+$.

ISBN: 978-989-95079-6-8

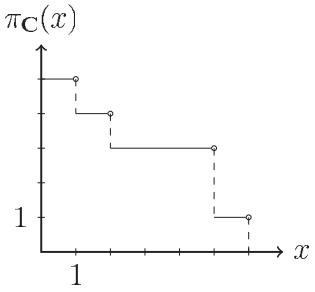


Figure 1: Citation function for $\mathbf{C} = (5, 4, 3, 3, 3, 1)$.

A set of all functions $f : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ dominated by the citation function π will be denoted by L_π.

Among the two new classes of scientific impact indices, the authors defined the p-sphere index.

Definition 3. Given an arbitrary real number $1 \leq p < \infty$ and any real number $r \geq 0$ let $s_{p,r} : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ denote a function

$$s_{p,r}(x) = \begin{cases} (r^p - x^p)^{\frac{1}{p}} & \text{for} \quad x \in [0, r)\,, \\ 0 & \text{for} \quad x \geq r. \end{cases} \tag{2}$$

Moreover, for $p = \infty$ we have

$$s_{\infty,r}(x) = \begin{cases} r & \text{for} \quad x \in [0, r)\,, \\ 0 & \text{for} \quad x \geq r. \end{cases} \tag{3}$$

Intuitively, for $x \in [0, r)$, the graph of $s_{p,r}(x)$ determines a part of an L^p-sphere (i.e. the boundary of an L^p-ball on a plane) of radius r (see Fig. 2) and centered at $(0, 0)$. Therefore $s_{p,r}$ is further on called the *p-sphere function* of *radius* r.

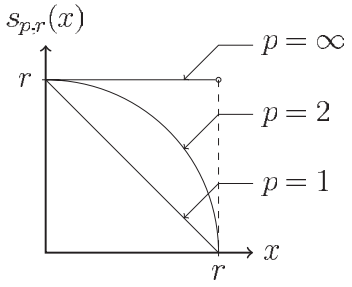


Figure 2: Exemplary p-spheres, for $p = 1, 2, \infty$.

Definition 4. The *p-sphere index* (or, originally, the *maximal p-radius*) of a citation function π is the greatest number $r \geq 0$, for which π still dominates $s_{p,r}$, i.e.

$$r_\pi(p) := \max\{r : \ \pi \succeq s_{p,r}\}\,. \tag{4}$$

The p-sphere index was primarily denoted $r_p(\pi)$. In this article we mainly consider the maximal radius to be a function of p, hence the change. For abbreviation, the p-sphere function of maximal p-radius $s_{p,r_\pi(p)}$ will be denoted by $S_{p,\pi}$ and called the *maximal p-sphere function* for π.

Here are some properties of the p-sphere index.

Lemma 5. *For any given $p \geq 1$ and a citation sequence $\mathbf{C} = (c_1, c_2, \ldots, c_n)$, let $\pi = \pi_{\mathbf{C}}$. Then the following properties hold:*

(i) $r_\pi(p) \leq n$.

(ii) $r_\pi(p) \leq c_1$.

(iii) If $p = \infty$, then $r_\pi(\infty) = H$, where H is the h-index of an individual (see [1]).

(iv) If $p = 1$, then $r_\pi(1) = W$, where W is the individual's w-index, as defined by Woeginger in [13].

(v) For any $q > p$, $r_\pi(q) \leq r_\pi(p) \leq 2r_\pi(q)$.

(ii) $r_\pi(p)$ is nonincreasing with respect to p.

Their proofs were given in [17].

4 Possible and necessary h-index

The theory of fuzzy measures and evidence has been well established. Therefore only definitions and properties that are necessary are recalled (for more details and further references the reader is referred, e.g., to [18]).

Further on we consider measures with respect to $\mathbb{R}_0^+$ and a family of all its subsets $\mathcal{P}(\mathbb{R}_0^+)$.

Definition 6. A function $\mu : \mathcal{P}(\mathbb{R}_0^+) \to [0, 1]$ is a *fuzzy measure* if it satisfies the following requirements:

M1. $\mu(\emptyset) = 0$ and $\mu(\mathbb{R}_0^+) = 1$ (boundary conditions),

M2. for all $A, B \in \mathcal{P}(\mathbb{R}_0^+)$, if $A \subseteq B$, then $\mu(A) \leq \mu(B)$ (monotonicity).

Definition 7. Let Pos denote a fuzzy measure. Then Pos is called a *possibility measure* iff for any family $\{A_k \in \mathcal{P}(\mathbb{R}_0^+) : k \in K\}$ and an arbitrary index set K,

$$\mathrm{Pos}\left(\bigcup_{k\in K} A_k\right) = \sup_{k\in K} \mathrm{Pos}(A_k). \tag{5}$$

For each possibility measure Pos we may associate another fuzzy measure Nec, called *necessity measure*, defined by

$$\mathrm{Nec}(A) := 1 - \mathrm{Pos}(\overline{A}), \tag{6}$$

where $A \in \mathcal{P}(\mathbb{R}_0^+)$.

It may be shown, that for every $A \in \mathcal{P}(\mathbb{R}_0^+)$ and for any possibility measure Pos and the associated necessity measure Nec following relations hold:

(i) $\mathrm{Nec}(A) > 0 \Rightarrow \mathrm{Pos}(A) = 1$,

(ii) $\mathrm{Pos}(A) < 1 \Rightarrow \mathrm{Nec}(A) = 0$.

Proposition 8. *Every possibility measure* Pos *may be uniquely determined by a* possibility distribution function *(abbreviated Pos.D.F.) $R : \mathbb{R}_0^+ \to [0, 1]$ by the formula*

$$\mathrm{Pos}(A) = \sup_{x\in A} R(x), \tag{7}$$

where $A \in \mathcal{P}(\mathbb{R}_0^+)$.

ISBN: 978-989-95079-6-8

Now we are ready to present a class of possibility measures for the h-index values.

Suppose we are given a citation sequence $\mathbf{C} = (c_1, c_2, \ldots, c_n)$ of some individual X. Some of the elements of $\mathbf{C}$ are either right-censored data or are potentially due to increase in a "short time". Let $H = h(\mathbf{C}) = \max\{i \in \mathbb{N}_0 : c_i \geq i\}$ be equal to X's Hirsch index. Thus we have sure evidence that the true value of X's Hirsch-index is *at least* H.

We propose some minimal requirements for a possibility distribution to describe hypothetical values of the h-index.

Definition 9. A *Pos.D.F. for the Hirsch index* is a mapping $R_h : \mathbb{R}_0^+ \to [0,1]$ which satisfies the following axioms:

H1. $R_h(x) = 0$ for $x < H$ or $x \notin \mathbb{N}_0$,

H2. $R_h(H) = 1$ (normalization),

H3. For any $x, x' \in \mathbb{N}_0$, if $H \leq x < x'$ then $R_h(x) > R_h(x')$ or $R_h(x') = 0$.

These axioms seem quite natural. Let us look at some properties of the fuzzy measures defined by such Pos.D.F. The following proposition might be easily proved.

Proposition 10. *Let R_h be an arbitrary Pos.D.F. for the Hirsch index. Furthermore, let* Pos *be the possibility measure determined by R_h and* Nec *be the associated necessity measure. Then for any $[a,b] \in \mathcal{P}(\mathbb{R}_0^+)$ the following properties are satisfied.*

- *If $H \in [a,b]$ then* $\mathrm{Pos}([a,b]) = 1$.
- *If $H \notin [a,b]$ then* $\mathrm{Nec}([a,b]) = 0$.
- *If $H \in [a,b]$ then for any $b < b'$ we get* $\mathrm{Nec}([a,b]) \leq \mathrm{Nec}([a,b'])$.

The postulated measures may give clues for questions such as: What is the (broadly conceived) possibility that the true value of X's h-index is really "equal to $H+1$" or "greater than H". Note that publishing-citing is an extremely complicated process and without drastic simplifications and idealizations it cannot be modeled using stochastic methods. Thus, in general, it is not reasonable to consider the results of individual scientometric measurements by means of the *probability* theory. Therefore, our problem is how to construct appropriate possibilistic measures. In the next section we propose a Pos.D.F. for the Hirsch index defined by means of the $r(p)$-indices.

5 Example

Lemma 11. *Assume we are given a citation function $\mathbf{C} = (c_1, c_2, \ldots, c_n)$, $\pi = \pi_{\mathbf{C}}$ and $r_\pi(\infty) = H \in \mathbb{N}$. Then the properties below are satisfied:*

(i) $c_i \geq H$ *for every* $i \leq H$,

(ii) $c_i \leq H$ *for every* $i > H$,

(iii) $r_\pi(p) \leq 2^{\frac{1}{p}} H$ *for any* $1 \leq p < \infty$.

Proof. Only (iii) is proved here. Let $1 \leq p < \infty$. By Lemma 5, $H \leq r_\pi(p) \leq 2H$. Let $a = \max\{c_1, 2H\}$. Consider a citation sequence $\mathbf{C}' = (c_1', c_2', \ldots, c_n')$, such that $c_i' = a$ for $i \leq H$, and $c_i' = H$ for $H < i \leq n$. Clearly, $r_{\pi_{\mathbf{C}'}}(\infty) = H$, and $c_i \leq c_i'$ for any i. For every $1 \leq p < \infty$, $S_{p,\pi_{\mathbf{C}'}}$ goes through (H, H), so $r_{\pi_{\mathbf{C}'}}(p)$ is the largest possible for fixed H. Solving (2) for r gives $r_{\pi_{\mathbf{C}'}}(p) = 2^{\frac{1}{p}} H$. Hence $r_{\pi_{\mathbf{C}}}(p) \leq 2^{\frac{1}{p}} H$ as stated. □

Let us define the inverse function of r_π. As r_π is not necessarily an injection, we need a special formula. In the sequel we assume that

$$r_\pi^{-1}(\varrho) := \max\{p : r_\pi(p) = \varrho\}, \tag{8}$$

for $\varrho \in [r_\pi(\infty), r_\pi(1)]$, and undefined otherwise. This definition is sensible, as it may be easily shown that for any $1 \leq p' < p''$ and $x \in \mathbb{R}_0^+$, $r_\pi(p') = r_\pi(p'')$ implies $S_{p',\pi}(x) \leq S_{p'',\pi}(x)$, hence the knowledge of maximal p satisfying $r_\pi(p) = \varrho$ is the most informative.

Now, let $R_\pi : \mathbb{R}_0^+ \to [0,1]$ be a function given by:

$$R_\pi(x) = \begin{cases} 0 & \text{for } x < r_\pi(\infty), \\ 2 - 2^{1/r_\pi^{-1}(x)} & \text{for } x \in [r_\pi(\infty), r_\pi(1)], \\ 0 & \text{for } x > r_\pi(1). \end{cases} \tag{9}$$

It is easily seen that for $x \in [r_\pi(\infty), r_\pi(1)]$, $R_\pi(x)$ is a strictly decreasing, but not necessarily continuous function.

For any $1 \leq p < \infty$, if $r_\pi(p) = 2^{\frac{1}{p}} H$, then $r_\pi^{-1}(x) = [\log_2(\frac{1}{H}x)]^{-1}$ and

$$R_\pi(x) = \begin{cases} 2 - \frac{1}{H}x & \text{for } x \in [H, 2H], \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

Lemma 12. *Let $\mathbf{C}$ and $\mathbf{C}'$ be citation sequences for which $\pi = \pi_{\mathbf{C}}$, $\pi' = \pi_{\mathbf{C}'}$, $r_\pi(\infty) = r_{\pi'}(\infty) = H$ and for any $1 \leq p < \infty$, $r_\pi(p) = 2^{\frac{1}{p}} H$.*

(i) $R_{\pi'}(x) \leq R_\pi(x)$ *for any x,*

(i) If $R_{\pi'}$ is given by (10)*, then $r_{\pi'}(p) = 2^{\frac{1}{p}} H$ for any $1 \leq p < \infty$.*

The proof is left to the reader.

It may be shown that $R_\pi|_{\mathbb{N}_0}$ is a Pos.D.F. for the Hirsch index for π. Thus the class of p-sphere indices may be used to construct a possibility measure having the properties of our interest. Obviously, one may find uncountably many transforms of r_π^{-1} to $[0,1]$, so that is just an example. Generating meaningful possibility measures by means of the p-sphere indices dependent on a type of the process affecting citation sequences is in scope of our future research.

Let us discuss the behavior of the proposed possible h-index on real-world data and see how it can be used to differentiate between individual's citation information. We consider the output of 3 Polish computer scientists, Prof. A having $n_A = 19$ publications, Prof. B with $n_B = 29$ publications and Prof. C with his $n_C = 18$ publications. Their citation sequences, according to *Scopus*[1], are given in Table 1. See Fig. 3 for the citation functions π_A, π_B, π_C and the maximal p-spheres for $p = 1, 2, \infty$. Each author's h-index equals $H = 7$, but they differ in values of other p-sphere indices, e.g. $r_{\pi_A}(1) = 12$, $r_{\pi_B}(1) = 14$ and $r_{\pi_C}(1) = 10$.

[1] The publication data were gathered on January 20, 2009 and are limited to the field "Computer Science". Authors' names have been intentionally masked.

ISBN: 978-989-95079-6-8

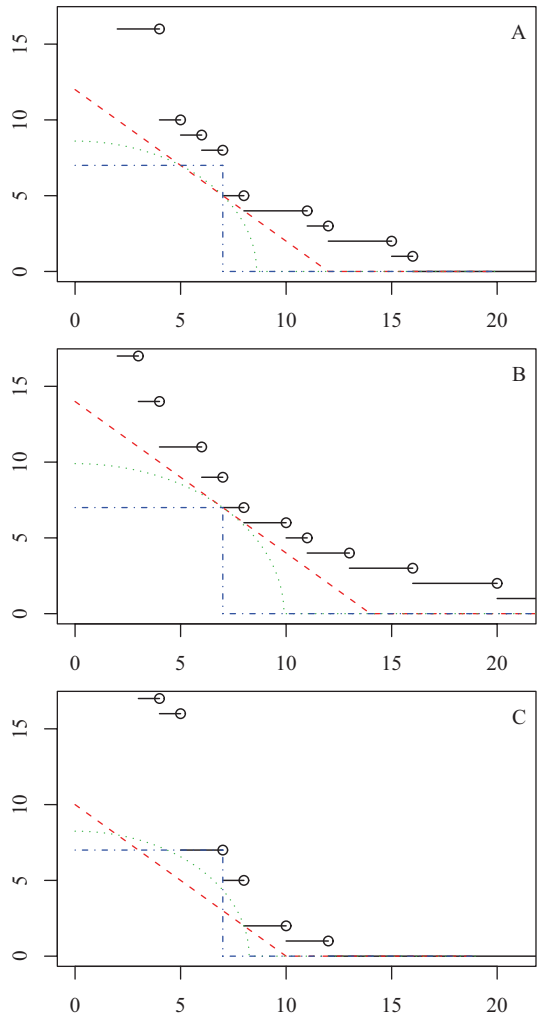


Figure 3: Citation functions of the 3 authors.

The author's R_π functions and resulting possibility distribution functions for the Hirsch-index are depicted in Fig. 4. Please note that R_{π_B} is of the form (10).

We see that Prof. B has the greatest possibility of increasing his h-index. On the contrary, some papers of Prof. C has not got a required number of citations so he can not have high expectations of a greater h-value. The proposed Poss.D.F. clearly differentiates between all the authors and can be used as a complement to the Hirsch index.

6 Conclusions

In the paper we discussed an important problem related to the Hirsch index: it assumes perfect knowledge of author's citation sequence and does not take into account a dynamic essence of the publication/citation process. Therefore we proposed a possibilistic approach to the indicator by setting several axioms for the possibility distribution function for the h-index.

In this model, given citation data are treated as an evidence for the minimal h-value and thus are just a starting point for speculation on its likely value in the case of perfect information.

We used a recently-proposed class of scientometric coefficients, the p-sphere indices, which is a generalization of Hirsch's index, to construct an exemplary possibility measure. A real-world example was presented for illustration.

As it was already mentioned, future work should definitely encompass the construction of different possibility measures, according to the type of a process affecting citation sequences.

Table 1: Citation sequences of the 3 authors.

$\mathbf{C}_A$	(103, 20, 16, 16, 10, 9, 8, 5, 4, 4, 4, 3, 2, 2, 2, 1, 0, 0, 0),
$\mathbf{C}_B$	(56, 30, 17, 14, 11, 11, 9, 7, 6, 6, 5, 4, 4, 3, 3, 3, 2, 2, 2, 2, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0),
$\mathbf{C}_C$	(39, 34, 23, 17, 16, 7, 7, 5, 2, 2, 1, 1, 0, 0, 0, 0, 0, 0).

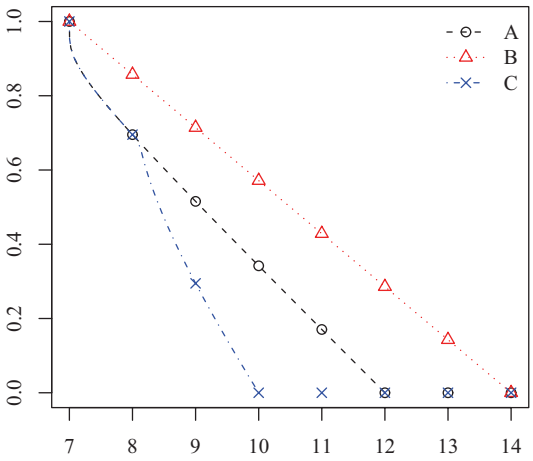


Figure 4: R_π and Pos.D.F. for the h-index of the 3 authors.

References

[1] Jorge E. Hirsch. An index to quantify individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46):16569–16572, 2005.

[2] Philip Ball. Index aims for fail ranking of scientists. *Nature*, 436:900, 2005.

[3] Lutz Bornmann and Hans-Dieter Daniel. The state of h index research. *EMBO Reports*, 10(1):2–5, 2009.

[4] Wolfgang Glänzel. On the h-index — A mathematical approach to a new measure of publication activity and citation impact. *Scientometrics*, 67(2):315–321, 2006.

[5] Lutz Bornmann and Hans-Dieter Daniel. What do we know about the h index? *Journal of the American Society for Information Science and Technology*, 58(9):1381–1385, 2007.

[6] Leo Egghe. Theory and practise of the g-index. *Scientometrics*, 69(1):131–152, 2006.

[7] Marek Kosmulski. A new Hirsch-type index saves time and works equally well as the original h-index. *ISSI Newsletter*, 2(3):4–6, 2006.

[8] BiHui Jin, LiMing Liang, Ronald Rousseau, and Leo Egghe. The R- and AR-indices: Complementing the h-index. *Chinese Science Bulletin*, 52(6):855–863, 2007.

[9] Michael Schreiber. A modification of the h-index: The h_m-index accounts for multi-authored manuscripts. *Journal of Informetrics*, 2(3):211–216, 2008.

ISBN: 978-989-95079-6-8

[10] Quentin L. Burrell. On the h-index, the size of the Hirsch core and Jin's A-index. *Journal of Informetrics*, 1:170–177, 2007.

[11] Leo Egghe. Modelling successive h-indices. *Scientometrics*, 77(3):377–387, 2008.

[12] Wolfgang Glänzel. On some new bibliometric applications of statistics related to the h-index. *Scientometrics*, 77(1):187–196, 2008.

[13] Gerhard J. Woeginger. An axiomatic characterization of the Hirsch-index. *Mathematical Social Sciences*, 56(2):224–232, 2008.

[14] Ronald Rousseau. Woeginger's axiomatisation of the h-index and its relation to the g-index, the $h(2)$-index and the r^2-index. *Journal of Informetrics*, 2(4):335–340, 2008.

[15] Antonio Quesada. Monotonicity and the Hirsch index. *Journal of Informetrics*, 3(2):158–160, 2009.

[16] Vicenç Torra and Yasuo Narukawa. The h-index and the number of citations: two fuzzy integrals. *IEEE Transactions on Fuzzy Systems*, 16(3):795–797, 2008.

[17] Marek Gągolewski and Przemysław Grzegorzewski. A geometric approach to the construction of scientific impact indices. *Scientometrics*, 2009. In press. DOI:10.1007/s11192-008-2253-y.

[18] George J. Klir and Bo Yuan. *Fuzzy sets and fuzzy logic. Theory and applications*. Prentice Hall PTR, New Jersey, 1995.

ISBN: 978-989-95079-6-8

Similarity based fuzzy interpolation applied to CAGD*

Joan Jacas Amadeu Monreal Jordi Recasens

Sec. Matemàtiques i Informàtica. E.T.S. d'Arquitectura de Barcelona
Univ. Politècnica de Catalunya. Diagonal, 649 - 08028 Barcelona, Spain
Email: joan.jacas@upc.edu, amadeo.monreal@upc.edu, j.recasens@upc.edu

Abstract— *This paper presents an approach to a system based on fuzzy logic for the* **design** *of interpolative curves and surfaces in the context of Computer Aided Geometric Design (CAGD). Some problems arising from a previous model are studied by using a similarity approach.*

Keywords— Fuzzy interpolation, indistinguishability operator, fuzzy number, Computer Aided Geometric Design,approximate reasoning, t-norm, fuzzy control.

1 Introduction

The aim of computer aided geometric design (CAGD) is to produce a curve or a surface whose shape is controlled by a set of given points. In this sense, it can be understood as an interpolative procedure where the goal object is not a function but its graphical presentation. Therefore, in the design process it is essential to control the degree on smoothness of the solution.

On the other hand, the user of a CAGD system provides the data "approximately" because, after a first trial, he usually adjusts the result to his own taste or aesthetic point of view. So, it is natural to treat the problem under the setting of approximate reasoning considering the input-output data as fuzzy points of $\mathbb{R}$ or $\mathbb{R}^2$ represented by fuzzy numbers.

In order to build such a system, we focus our attention in the fuzzification and defuzzification procedures. To define and control the shapes on the input-output fuzzy numbers, we assume that there exists a similarity relation in $\mathbb{R}$ that granulates and fuzzifies its elements. With respect to the defuzzification process, we can choose the Takagi-Sugeno model that produces a crisp output or select some of the well known defuzzification methods like the center of gravity. The paper is organized in the following way. After this introduction, section 2 recalls the preliminaries related to indistinguishability operators and fuzzy classes. Sections 3 and 4 presents the general model and a model associated to Takagi-Sugeno technique that tries to overcome some of the problems of the model proposed in [17], such as the effective construction of the fuzzy numbers and the solution for the case of sparse data. In section 5 an example is fully developed and we close with conclusions and future works.

2 Preliminaries

The large number of successful applications of fuzzy sets theory (specially in fuzzy control) has produced numerous concepts based on empirical motivations. In order to give a sound justification to the use of these applied concepts, it is necessary to provide in the setting of a clear model, a well-founded interpretation for them. The aim of this section is to introduce an appropriate background that justifies the different definitions and types of fuzzy real numbers as they are interpreted and used in the applications.

As a starting point it is assumed that the concept of fuzzy real number arises from the fact that there exists a certain equality relation in $\mathbb{R}$, in the sense that numbers laying "close" are indistinguishable to the observer [15]. When the defined equality is a T-indistinguishability operator [7,21], triangular and trapezoidal fuzzy numbers among others, appear in a natural way. The model justifies the use of different types of fuzzy numbers depending on the scale defined in $\mathbb{R}$ and also develops the way of generating indistinguishability operators on $\mathbb{R}$ associated to one or several scales.

Fuzzy numbers arise naturally from the underlying indistinguishalility as the singletons associated to the relation [2,12,13,14]. Therefore, only the fuzzy relations whose singletons fulfill a general definition of fuzzy number are considered. Reciprocally, given a "suitable" fuzzy equality in $\mathbb{R}$, only the fuzzy numbers that are the singletons of this relation will be considered as such. As a natural consequence of this setting, we also study the families of fuzzy numbers that are invariant under translations.

Throughout this paper we use the standard definitions and properties of t-norms. In the sequel Π will represent the product t-norm and $\mathcal{L}$ the Łuckasiewicz t-norm. Some definitions on T-indistiguinshability operators and fuzzy numbers are recalled. For a more detailed exposition on these topics, readers are referred to [1,7,15,16,18].

Definition 2.1. Let T be a t-norm. A *T-indistinguishability operator* E on a set X is a reflexive and symmetric fuzzy relation on X such that

$T(E(x,y),E(y,z)) \leq E(x,z)$ (T-transitive property).

If $T = \text{Min}$, then E is called a *similarity*. For $T = \mathcal{L}$, E is termed a *likeness* and when $T = \Pi$, E is a *probabilistic relation*.

T-indistinguishability operators generalize within the fuzzy framework the concept of equivalence relation and are also called equality relations [2,7]. From a semantical point of view, $E(x,y)$ can be interpreted as the degree of similarity between x and y.

Definition 2.2. Given a T-indistinguishability operator E on a set X and an element x_0 of X, the *singleton* induced by x_0 is the fuzzy set f_{x_0} defined by $f_{x_0}(x) = E(x_0, x)$.

The singleton of an element x of X can be thought as its fuzzy equivalence class with respect to E [12] and its fuzzification

*Partially supported by the project TIN2006-14311

ISBN: 978-989-95079-6-8

taking E into account [8].

Definition 2.3. A *fuzzy real number* is a map $f_a : \mathbb{R} \to [0,1]$ such that there exists a number $a \in \mathbb{R}$ with $f_a(a) = 1$ where f_a is non-decreasing on $(-\infty, a)$ and non-increasing on $(a, +\infty)$.

Definition 2.4. A T-indistinguishability operator E on $\mathbb{R}$ is called *admissible* if and only if its singletons are fuzzy numbers.

The fuzzified of a real number via an admissible indistinguishability operator is therefore a fuzzy number.

The link between these operators and some families of fuzzy numbers will be investigated.

2.1 Admissible T-indistinguishability operators

A scale on $\mathbb{R}$ is defined as a monotonic map $f : \mathbb{R} \to \mathbb{R}$. The possibility to distinguish real numbers in a fuzzy environment is based on the use of a scale. Different scales give different degrees of accuracy and determine different fuzzy equalities on $\mathbb{R}$. Theorem 2.1.2 formalizes this idea.

Theorem 2.1.1.[14] Let m be a pseudometric on a set X and t a generator of an archimedean t-norm T. Then $E = t^{[-1]} \circ m$ is a T-indistinguishability operator on X.

N.B. Different generators of the same t-norm generate different T-indistinguishability operators.

Theorem 2.1.2.[14] Let $f : \mathbb{R} \to \mathbb{R}$ be a monotonic map and t a generator of an archimedean t-norm T

$$E_f(x,y) = t^{[-1]}(|f(x) - f(y)|)$$

is an admissible T-indistinguishability operator.

Two monotonic maps f, g generate the same T-indistinguishability operator if and only if there exists a constant k such that $f = \pm g + k$.

In real situations where different experts produce different scales on $\mathbb{R}$ it is necessary to combine their information. The following theorem gives a way to do it.

Theorem 2.1.5. If $\{E_i\}_{i \in I}$ is a family of admissible T-indistinguishability operators on $\mathbb{R}$, then

$$E = \operatorname*{Inf}_{i \in I} E_i$$

is also an admissible T-indistinguishability operator on $\mathbb{R}$.

Therefore, we can combine admissible likeness or probabilistic relations generated by monotonic functions in order to obtain another equality on $\mathbb{R}$ of the same type.

2.2 Fuzzy numbers invariant under translations

Special families of fuzzy numbers are those who are invariant under translations. Roughly speaking, families formed by fuzzy numbers of the same "shape". In the same sense, we can also consider T-indistinguishability operators, invariant under translations i.e.: the degree of similarity $E(x,y)$ between two numbers x, y depends only on their distance.

In this section, we study the link between the families of fuzzy numbers and T-indistinguishability operators invariant under translations.

Definition 2.2.1. A family $\{f_a\}_{a \in \mathbb{R}}$ of fuzzy numbers associated to an admissible T-indistinguishability operator is *invariant under translations* if and only if

$$f_a(x) = f_b(x - a + b).$$

A T-indistinguishability operator E is *invariant under translations* if and only if

$$E(x + a, y + a) = E(x, y).$$

Let us search for the conditions that such a family $\{f_a\}_{a \in \mathbb{R}}$ must fulfill in order to become the set of the singletons associated to a T- indistinguishability operator E, being T Archimedean.

Theorem 2.2.4.[7] Given an archimedean t-norm T with additive generator t, the families $\{f_a\}_{a \in \mathbb{R}}$ of fuzzy numbers invariant under translations corresponding to singletons of a T-indistinguishability operator on $\mathbb{R}$ can be obtained from f_0 (using definition 2.2.1) defined by

$$f_0 = t^{[-1]} \circ F$$

where F is an even and non decreasing in $\mathbb{R}^+$ subadditive function with $F(0) = 0$.

The associated T-indistinguishability operator E is defined by

$$E(x,y) = t^{[-1]} \circ F(y - x)$$

which is trivially invariant under translations.

In this section, we have introduced a general setting in order to generate fuzzy numbers associated to generalized equalities. The model allows the effective construction of families of fuzzy numbers suitable for applications since the underlying equalities can be deduced from scales proposed by experts and specially adapted to a concrete implementation.

3 Interpolation in CAGD

In CAGD, usually the user provides a set of points and the program produces a curve or surface whose shape is controlled by these points. From a classical point of view an interpolative curve means a curve passing through a given set of points. In our proposal we treat this problem in a more relaxed way in the sense that we obtain the values of the curve from a set of data points but we do not impose the condition that the resulting curve should interpolate in the classical sense. Since it is a design process, it is essential to control the degree of smoothness of the solution.

In this paper, we propose the following problems, paying special attention to the case of plane curves:

Problem 1a) *(Non parametric or functional curves).* Given $n+1$ points of control $\overrightarrow{P_i} = (a_i, b_i) \in \mathbb{R}^2$, $i = 0, 1, \ldots, n$, with $a_0 < a_1 < ... < a_n$, we want to construct the graphic of a function $f : [a_0, a_n] \to \mathbb{R}$ such that $\{a_i\}$ are values of the independent variable and $\{b_i\}$ are values or approximations of $\{f(a_i)\}$.

Problem 1b) *(Parametric curves).* Given $n+1$ points of control $\overrightarrow{P_i} = (a_i, b_i) \in \mathbb{R}^2$, or $\overrightarrow{P_i} = (a_i, b_i, c_i) \in \mathbb{R}^3$, $i = 0, 1, \ldots, n$ we want to construct $n+1$ values $t_0 < t_1 < \ldots <$

ISBN: 978-989-95079-6-8

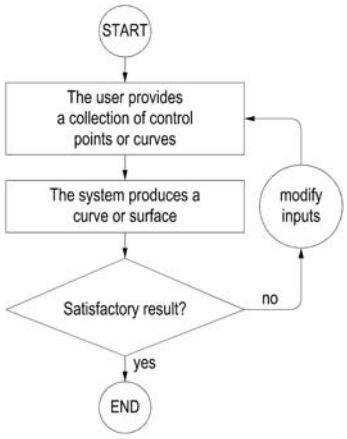


Figure 3.1

t_n of a parameter and a parametric curve $\overrightarrow{f} : [t_0, t_n] \to \mathbb{R}^2$ or $\mathbb{R}^3$ in such a way that the points $\{\overrightarrow{P}_i\}$ are the values or approximations of $\{\overrightarrow{f}(t_i)\}$.

Problem 2a) *(Non parametric or functional surfaces)*. Given $n+1$ arbitrarily distributed points of control $\overrightarrow{P_i} = (a_i, b_i, c_i) \in \mathbb{R}^3$, $i = 0, 1, \ldots, n$, we want to construct a surface passing through or controlled by these points as the graphic of a map $f : A \subset \mathbb{R}^2 \to \mathbb{R}$ (for example, $A = [a_{\min}, a_{\max}] \times [b_{\min}, b_{\max}]$) such that $\{(a_i, b_i)\}$ are the values of the independent variables and $\{c_i\}$ values or approximations of $\{f(a_i, b_i)\}$.

Problem 2b) *(Parametric surfaces)*. Given a net of $(n+1) \times (m+1)$ points $\overrightarrow{P}_{ij} = (a_{ij}, b_{ij}, c_{ij}) \in \mathbb{R}^3 \quad i = 0, \ldots, n$, $j = 0, \ldots, m$, we want to construct $n+1$ values $s_0 < s_1 < \ldots < s_n$ of a parameter s and $m+1$ values $t_0 < t_1 < \ldots < t_m$ of a parameter t and a parametric surface $\overrightarrow{f} : [s_0, s_n] \times [t_0, t_m] \to \mathbb{R}^3$ in such a way that the points $\{\overrightarrow{P}_{ij}\}$ are the values or approximations of $\{\overrightarrow{f}(s_i, t_j)\}$.

There are many standard techniques to treat the preceding problems. For a more complete information, also about methods to construct values of the parameter in parametric cases, the reader is refered to [9,10,11].

The common algorithm of all design methods that interacts between the procedure and the user is shown in the Fig. 3.1.

The algorithm shows the existing vagueness of the data points and its relation with the goal curves. Therefore, it is natural to focus the problem in the setting of approximate reasoning. The model proposed in section 4 uses fuzzy numbers and techniques of Takagi-Sugeno in fuzzy control to design curves and surfaces.

4 The Takagi-Sugeno Model

We propose a system based on fuzzy linguistic rules used in fuzzy control [5,18]. We have adopted the point of view of Takagi and Sugeno [23], in order to avoid the defuzzification process.

The main idea is to use rules of the following type:

If $\overrightarrow{x}$ is **close** to $\overrightarrow{x}_i$ in the domain, then $\overrightarrow{y}$ is **close** to $\overrightarrow{y}_i$ in the image. This idea is developed in the following **model**:

- Use $n+1$ rules

  R_i : If $\overrightarrow{x} \simeq \overrightarrow{x}_i$, then $\overrightarrow{y} \simeq \overrightarrow{y}_i$; $i = 0, 1, \ldots, n$.

- Model the degree of fulfillment of the antecedent of the i^{th} rule with a fuzzy set α_i such that $\alpha_i(\overrightarrow{x}_i) = 1$ and $\alpha_i(\overrightarrow{x})$ is a decreasing function of the distance to $\overrightarrow{x}_i$ [6].

- The consequence function of each rule, $f_i(\overrightarrow{x})$, is the constant $\overrightarrow{y}_i$. (Therefore, the system coincides with the **height defuzzification method**).

- The output of the system is obtained as a mean of the outputs of each rule weighted by the fuzzy sets α_i:

$$\overrightarrow{y}(\overrightarrow{x}) = \frac{\sum_{i=0}^{n} \alpha_i(\overrightarrow{x}) \cdot \overrightarrow{y}_i}{\sum_{i=0}^{n} \alpha_i(\overrightarrow{x})} = \sum_{i=0}^{n} F_i(\overrightarrow{x}) \cdot \overrightarrow{y}_i \quad (2.1.)$$

  where $F_i(\overrightarrow{x}_i) = \frac{\alpha_i(\overrightarrow{x})}{\sum_{j=1}^{n} \alpha_j(\overrightarrow{x})}$.

So, the result is a convex linear combination of the outputs of the rules where the coefficients depend on the data points and the fuzzy sets α_i. This is the general description of our proposal already outlined in [17] that contained the following remarks

4.1 Remarks

4.1.1. If the domain is in $\mathbb{R}$ (the case of a curve), then $\alpha_i = \mu_i$ are **fuzzy numbers**.

4.1.2. If the domain is in $\mathbb{R}^2$ (the case of a surface), then the closeness α_i of a point $\overrightarrow{x} = (x, y)$ to $\overrightarrow{x}_i = (x_i, y_i)$ can be modeled in the following way:
$\overrightarrow{x} = (x, y)$ is **close** to $\overrightarrow{x}_i = (x_i, y_i)$ if and only if x is **close** to x_i **and** y is **close** to y_i. That is, if μ_i and ν_i are fuzzy numbers modeling the closeness of x and y to x_i and y_i respectively, then the fuzzy set α_i can be defined as $\alpha_i(\overrightarrow{x}) = T(\mu_i(x), \nu_i(y))$, where T is a suitable t-norm that plays the semantic role of a conjunction. In order to maintain the degree of smoothness of the resulting surface inherited from the smoothness of the fuzzy numbers, it is convenient to choose a C^∞ t-norm. In our model, we use the **product**.

4.1.3. The fuzzy sets $\{\mu_i\}$ must cover the domain in the sense that for every point $\overrightarrow{x}$ in the domain, there must exists a fuzzy set μ_i with $\mu_i(\overrightarrow{x}) \neq 0$. This is a technical requirement in order to avoid dividing by zero in (2.1) and has the semantical meaning that the image of every point of the domain must be controlled by at least one of the given points $\{\overrightarrow{x}_i\}$.

4.1.4. Curves of type 1b) generated by this model cannot be, simultaneously, interpolative and smooth (i.e. tangent continuous). For interpolative curves $\overrightarrow{y}(t_i) = \overrightarrow{P}_i$, so $\mu_i(t_i) = 1$ and $\mu_j(t_i) = 0 \quad j \neq i$. Due to the monotonicity of the μ_i, $\mu_i(t) \equiv 0 \quad t \notin (t_{i-1}, t_{i+1})$. That means that for every interval $[t_{i-1}, t_i]$ there are only two fuzzy numbers different from zero. Consequently, in this interval the curve is a convex combination of the points $\overrightarrow{P}_{i-1}$ and $\overrightarrow{P}_i$ and we obtain a polygonal curve.

ISBN: 978-989-95079-6-8

4.1.5. It is worth noticing that, whereas in the classical methods used in CAGD only the selection of points is important, here we also have the possibility of choosing the type and size of the fuzzy number associated to each point as a consequence of the underlying indistinguishability, even within the same set of nodes. This aspect gives great versatility to our model.

From these remarks we can detect some problems in order to obtain a general procedure. The first one deals with the effective construction of the fuzzy numbers associated to the crisp data; that is, the fuzzification problem. If we consider that all the given points have the same weight in the construction of the curve or surface, it is natural to accept that the fuzzy points are invariant under translations and we can apply the ideas contained in the subsection 2.2. Otherwise, we can use the scales in order to graduate the importance of the points contained in different intervals of the domain. The fuzzification process of $\mathbb{R}$ is completely meaningful using the underlying equality relations. A second and important problem is mentioned in the remark 4.1.3., since our initial proposal is not applicable when dealing with sparse data. This problem has been extensively treated in the literature[8,10,19,20] from various points of view. Also, we can find many papers where different methods are compared [3,4,6,22]. In order to overcome this problem, we distinguish two different situations:

- a) The data set consists of equidistant x-values. To solve this case, we propose to build a family of parametrized fuzzy equalities in $\mathbb{R}$ whose associated fuzzy numbers cover all the domain for values of the parameter and for every $\vec{x_i}$ of the data set there exists a fuzzy number μ_i such that $\mu_i(\vec{x_j}) = 1$ if $i = j$ and $\mu_i(\vec{x_j}) = 0$ otherwise. Under this hypothesis we can obtain a curve that interpolates the points in the classical sense (see the example in section 5).

- b) The data has sparse non equidistant values. Our proposal consists in constructing a monotonic function f between a subset of the natural numbers and our numerable (finite) data set and using the fuzzy equality $E(x_i, x_j) = t^{-1} \circ F(f^{-1}(x_i) - f^{-1}(x_j))$ in order to define the fuzzy numbers associated to data.

With all this in mind, the solutions proposed to the problems of the introduction are the following:

4.2 Solutions.

Problem 1a) *(Non parametric curves).* In this case, $\overrightarrow{P}_i = (a_i, b_i)$, the domain is some interval in $\mathbb{R}$, $\overrightarrow{x}_i = a_i$, $\overrightarrow{y}_i = b_i$ and $\alpha_i = \mu_i$ are normalized fuzzy numbers. The solution curve is

$$y = f(x) = \frac{\sum_{i=0}^{n} \mu_i(x) \cdot b_i}{\sum_{i=0}^{n} \mu_i(x)} = \sum_{i=0}^{n} F_i(x) \cdot b_i. \qquad (2.2)$$

Problem 1b) *(Parametric curves).* In this case, $\overrightarrow{P}_i = (a_i, b_i) \in \mathbb{R}^2$ or $\overrightarrow{P}_i = (a_i, b_i, c_i) \in \mathbb{R}^3$. The domain is an interval in $\mathbb{R}$, the variable is a parameter t, $\overrightarrow{x}_i = t_i$, $\overrightarrow{y}_i = \overrightarrow{P}_i$ and $\alpha_i = \mu_i$ are normalized fuzzy numbers. The solution curve is

$$\overrightarrow{y} = \overrightarrow{f}(t) = \frac{\sum_{i=0}^{n} \mu_i(t) \cdot \overrightarrow{P}_i}{\sum_{i=0}^{n} \mu_i(t)} = \sum_{i=0}^{n} F_i(t) \cdot \overrightarrow{P}_i. \qquad (2.3)$$

Problem 2a). *(Non parametric surfaces).* In this case, $\overrightarrow{P}_i = (a_i, b_i, c_i) \in \mathbb{R}^3$, the domain is some region on $\mathbb{R}^2$, $\overrightarrow{x}_i = (a_i, b_i)$, $\overrightarrow{y}_i = c_i$ and $\alpha_i(x, y) = \mu_i(x) \cdot \nu_i(y)$ where μ_i, ν_i are normalized fuzzy numbers.

The solution surface is

$$z = f(x, y) = \frac{\sum_{i=0}^{n} \alpha_i(x, y) \cdot c_i}{\sum_{i=0}^{n} \alpha_i(x, y)} = \sum_{i=0}^{n} F_i(x, y) \cdot c_i. \qquad (2.4)$$

Problem 2b). *(Parametric surfaces).*

In this case, $\overrightarrow{P}_{ij} = (a_{ij}, b_{ij}, c_{ij})$, the domain is a rectangle in $\mathbb{R}^2$, the variables are the parameters $s \in [s_0, s_n]$, $t \in [t_0, t_m]$, $\overrightarrow{x}_{ij} = (s_i, t_j)$, $\overrightarrow{y}_{ij} = \overrightarrow{P}_{ij}$ and $\alpha_{ij}(s, t) = \mu_i(s) \cdot \nu_j(t)$ where μ_i, ν_j are normalized fuzzy numbers.

The solution surface is

$$\overrightarrow{P}(s, t) = \frac{\sum_{i=0}^{n} \sum_{j=0}^{m} \alpha_{ij}(s, t) \cdot \overrightarrow{P}_{ij}}{\sum_{i=0}^{n} \sum_{j=0}^{m} \alpha_{ij}(s, t)} = \sum_{i=0}^{n} \sum_{j=0}^{m} F_{ij}(s, t) \overrightarrow{P}_{ij} \qquad (2.5)$$

where

$$F_{ij}(s, t) = \frac{\alpha_{ij}(s, t)}{\sum_{i=0}^{n} \sum_{j=0}^{m} \alpha_{ij}(s, t)}$$

Note: Since in this case, the set of points determines a net, we have used a double index for convenience. If we want to use a single index k, we can, for instance, reassign indices in the following way: $\{i, j\} \to \{k\}$ with $k = i \cdot (m + 1) + j$.

5 An example

In this section we develop an example for fuzzy interpolation of curves and surfaces in its functional and parametric forms. In order to generate the underlying fuzzy equality in $\mathbb{R}$ invariant under translations, we need to select a subadditive, even and non decreasing function in $\mathbb{R}^+$. We have choosen $F = \text{Arctan}\,|x|$. This function fulfills all the conditions of theorem 2.2.1. On the other hand we have selected a parametrized family of strictly decreasing functions $f(x, k) : [0, 1] \to \mathbb{R}^+$ that, jointly with its inverses (Fig. 5.1 and Fig. 5.2), allows us to build a parametrized family of t-norms represented in Fig. 5.3 for $k = 5$.

Fig. 5.4 contains a representation of the associated T-indistinguishability that produces the fuzzification of $\mathbb{R}$ depicted in Fig. 5.5 for a discrete sub-family. By using this fuzzification and the procedure explained in section 4, we produce an interpolating curve for the case of equidistant values in functional form (Fig. 5.6).

Fig. 5.7 and Fig. 5.8 present two solutions of a curve constructed by means of its parametrized representation. As it is

ISBN: 978-989-95079-6-8

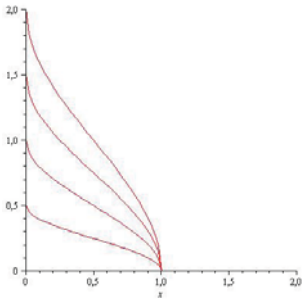


Figure 5.1

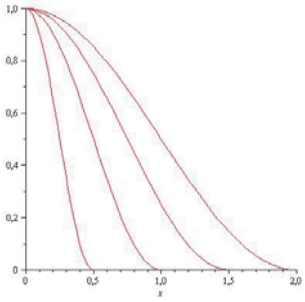


Figure 5.2

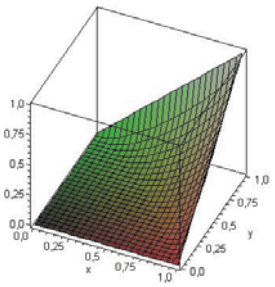


Figure 5.3

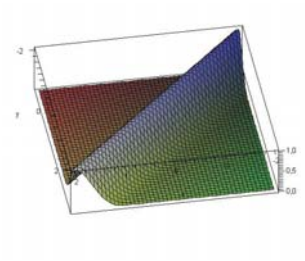


Figure 5.4

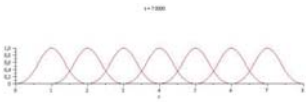


Figure 5.5

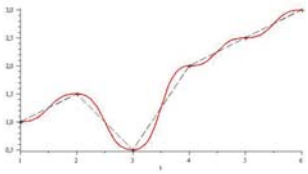


Figure 5.6

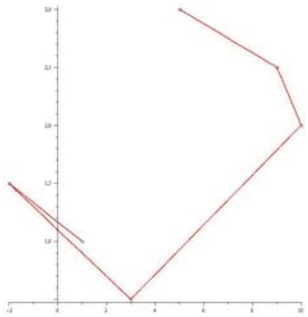


Figure 5.7

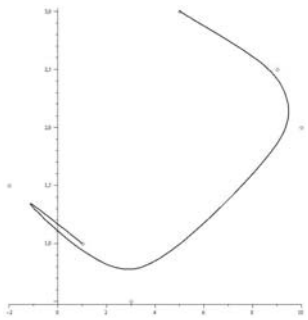


Figure 5.8

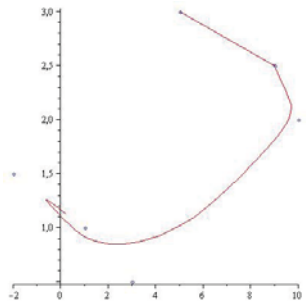


Figure 5.9

ISBN: 978-989-95079-6-8

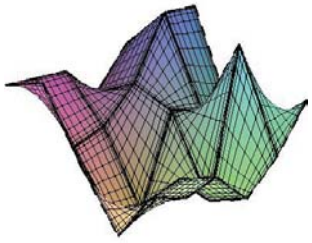


Figure 5.10

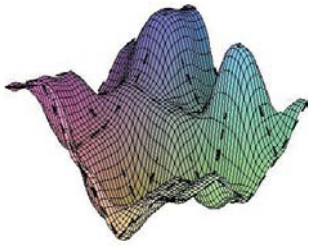


Figure 5.11

pointed out in the remarks, this representation can not be interpolative, in the classical sense, and, at the same time smooth.

Finally, in Fig. 5.9 a parametric curve is represented using a fuzzification of $\mathbb{R}$ defined by a scaled indistinguishability.

In a similar way we can construct surfaces.In this setting,the same incompatibility between interpolation and smoothnes, in the case on parametric representation, applies. Fig. 5.10 shows the case of interpolation for parametric representation. Fig. 5.11 shows the interpolative surface for the functional case.

References

[1] C. Alsina, On a family of connectives for fuzzy sets. *Fuzzy Sets and Systems* 16, (1985) 231-235.

[2] D. Boixader,J. Jacas, J. Recasens, Fuzzy equivalence relations: advanced material. In *Dubois and Prade Eds. Fundamentals of fuzzy sets*.(2000) Kluwer 261-290.

[3] B. Bouchon-Meunier et alt. A comparative view of interpolation methods between sparse fuzzy rules. *Proc. IFSA'01 World Congress*, Vancouver (2001) 2499-2504.

[4] B. Bouchon-Meunier, C. Marsala, M. Rifqi. Interpolative reasoning based on granularity. Proc. FUZZ-IEEE'2000 (2000) 483-487.

[5] J.L.Castro, M. Delgado, Fuzzy Systems with defuzzification are universal approximators, *IEEE Trans. Systems, Man and Cybernetics*, 26 (1996) 149-152.

[6] D. Dubois,H. Prade On Fuzzy interpolation, *Int. J. General Systems* 28 (1999) 103-114.

[7] M. Dermirci, J. Recasens, Fuzzy groups, fuzzy functions and Fuzzy equivalent relations, *Fuzzy Sets and Systems* 144 (2004) 441-450.

[8] F. Esteva et alt. Similarity-based fuzzy interpolation method. *Proc. IPMU 2004*. Perugia (2004) 1443-1449.

[9] G. Farin, *Curves and Surfaces for CAGD.* Academic Press, 1990.

[10] I.D.Faux, M.J.Pratt, *Computational Geometry for Design and Manufacture*, Ellis Horwood Ltd., 1985.

[11] D. Hearn, M.P.Baker, *Computer Graphics*, Prentice Hall, 1994.

[12] U. Höhle, Quotients with respect to similarity relations, *Fuzzy Sets and Systems* 17, (1988) 31-44.

[13] J. Jacas, J. Recasens, On some kind of probabilistic relations. *Proc IPMU Congress*, Palma de Mallorca, (1992) 365-368.

[14] J. Jacas, J. Recasens, Fuzzy Numbers and Equality Relations, *Proc. IEEE International Conference on Fuzzy Systems* San Francisco (1993), 1298-1301..

[15] J. Jacas, J. Recasens, Fixed Points and Generators of Fuzzy Relations, *J. Math. Anal. and Appl.* 186 (1994) 21-29.

[16] J. Jacas, J. Recasens, Fuzzy T-transitive relations: eigenvectors and generators, *Fuzzy Sets and Systems* 72 (1995) 147-154.

[17] J. Jacas, A. Monreal, J. Recasens, A model for CAGD using fuzzy logic, *Int. J. Approximate Reasoning* 72 (1997) 289-308.

[18] F. Klawoon, R. Kruse, Equality Relations as a Basis for Fuzzy Control. *Proc. IEEE International Conference on Fuzzy Systems*, (1992).

[19] L. T. Kòckzy, K. Hirota. Interpolative reasoning with insuficient evidence in sparce rulebases. *Information Sciences* 71(1993) 169-200.

[20] L. T. Kòckzy, K. Hirota. Approximate reasoning by linear rule interpolation and general approximation.*Int. J. Approximate Reasoning* 9(1993) 197-222.

[21] R. Kruse, J. Gebhardt, F. Klawonn, *Foundations of Fuzzy Systems.* John Wiley & Sons, Chichester, 1994.

[22] S. Mizik, P. Baranyi, P. Korondi, H and L. T. Kczy. A comparative of fuzzy interpolation techniques. 4^{th} *Eurofuzzy-Sic'99* (1999) Budapest 544-549

[23] H.Takagi, , M. Sugeno, Fuzzy identification of Systems and its applications to modelling and control, *IEEE Trans. on Systems, Man and Cybernetics*, 15, 116-132, (1985).

ISBN: 978-989-95079-6-8

KM-Fuzzy Approach Space

J. Martínez-Moreno[1] A. Roldán[2] C. Roldán[2]

1. Department of Mathematics. University of Jaén
Jaén, Spain
2. Department of Statistics and Operations Research. University of Jaén
Jaén, Spain
Email: jmmoreno,afroldan,iroldan@ujaen.es

Abstract— *Using the idea of approach space due to Robert Lowen [Approach spaces. The missing link in the topology-uniformity-metric triad], we define the notion of fuzzy approach spaces as a natural generalization of fuzzy metric spaces due to Kramosil and Michalek [Kybernetika , 11, 1975, 336–344] and prove some properties of fuzzy approach spaces.*

Keywords— Approach spaces, Fuzzy approach spaces, Fuzzy metric spaces.

1 Introduction

The notion of fuzzy sets was introduced by Zadeh [1]. Since then, several concepts of fuzzy metric spaces were considered in [2, 3, 4]. In the sequel, we shall adopt the usual terminology, notation and conventions of the theory of fuzzy metric spaces introduced by Kramosil and Michalek [5]. In this section we recall the definitions of a fuzzy metric space, approach space and some elementary properties.

Definition 1.1 *A binary operation* $*\colon [0,1]\times[0,1]\to[0,1]$ *is a continuous t-norm if* $([0,1],*)$ *is a topological monoid with unit 1 such that* $a*b\le c*d$ *whenever* $a\le c$ *and* $b\le d$ *for* $a,b,c,d\in[0,1]$.

Definition 1.2 *A* KM-fuzzy quasi-metric space *(briefly qFM-space) is a triple* $(X,M,*)$ *where* X *is an arbitrary set,* $*$ *is a continuous t-norm and* $M\colon X\times X\times[0,+\infty]\to[0,1]$ *is a (fuzzy) mapping such that for all* $x,y,z\in X$*:*

FM1. $M(x,y,0)=0$*;*

FM2. $x=y$ *if and only if* $M(x,y,t)=1$ *for all* $t>0$*;*

FM3. $M(x,z,t+s)\ge M(x,y,t)*M(y,z,s)$ *for all* $s,t\ge 0$*;*

FM4. $M(x,y,\cdot)\colon[0,+\infty)\to[0,1]$ *is left-continuous.*

If we additionally assume that M satisfies the symmetry condition:

$$M(x,y,t)=M(y,x,t)\ \text{ for all } t>0,$$

then $(X,M,*)$ *is called fuzzy metric space.*

Lemma 1.3 $M(x,y,\cdot)$ *is a nondecreasing function for all* $x,y\in X$.

This paper introduces the concept of a fuzzy approach space. Our definition is inspired in the concept of approach spaces introduced by Lowen [6].

Definition 1.4 *A pair* (X,δ) *where* $\delta: X\times\mathcal{P}(X)\to[0,+\infty]$ *is a function on a set X such that:*

B1. $\delta(x,\{x\})=0$ *for every* $x\in X$.

B2. $\delta(x,\emptyset)=\infty$ *for every* $x\in X$.

B3. $\delta(x,A\cup B)=\min(\delta(x,A),\delta(x,B))$ *for every* $x\in X$ *and* $A,B\in\mathcal{P}(X)$.

B4. $\delta(x,A)\le\delta(x,A^{(r)})+r$ *for every* $p\in X$, $A\in\mathcal{P}(X)$ *and* $r\in[0,\infty]$, *with* $A^{(r)}=\{\,y\in X\ /\ \delta(y,A)\le r\,\}$.

*is called an approach space (**A-space**).*

Lowen ([6], 1997) proved that at least seven different structures can be used to define an approach space, equivalently. Assuming condition B3, B4 can be rewritten as follows (see [7]):

B4′. $(A^{(r)})^{(s)}\subseteq A^{(r+s)}$ for every $A\subseteq X$ and $r,s\in[0,\infty]$.

2 KM-Fuzzy approach spaces

Following, the notions of approach space are generalized to the fuzzy setting and the concept of KM-fuzzy approach spaces is introduced.

Definition 2.1 *A* KM-fuzzy approach space *(briefly FA-space) is a pair* (X,F) *where* X *is an arbitrary set and* $F\colon X\times\mathcal{P}(X)\times[0,+\infty]\to[0,1]$ *is a (fuzzy) mapping satisfying the following conditions:*
For all $x\in X$, $A,B\subset X$ *and* $s,t\in[0,+\infty]$,

FA1. $F(x,A,0)=0$*;*

FA2. $F(x,\emptyset,t)=0$ *for all* $t<+\infty$*;*

FA3. $F(x,\{x\},t)=1$ *for all* $t>0$*;*

FA4. $F(x,A\cup B,t)\ge\max(F(x,A,t),F(x,B,t))$*;*

ISBN: 978-989-95079-6-8

FA5. $F(x,A,t+s) \geq F(x,A^{(r)},t)$, *if* $s > r \geq 0$,
$F(x,A,+\infty) \geq F(x,A^{(+\infty)},t)$,
where $A^{(r)} = \{y \in X : F(y,A,t) = 1, \forall t > r\}$,
and $A^{(+\infty)} = \{y \in X : F(y,A,+\infty) = 1\}$;

FA6. $F(x,A,\cdot)\colon [0,+\infty) \to [0,1]$ *is left-continuous for all* $x \in X$ *and* $A \subset X$.

The map F is symmetric (or the FA-space (X,F) is symmetric) if

$$F(x,\{y\},t) = F(y,\{x\},t) \text{ for all } x,y \in X,\ t > 0.$$

Firstly, we will prove that there are two families of spaces satisfying the above six properties.

Lemma 2.2 (A-space $\hookrightarrow$ FA-space) *Let (X,δ) be an A-space. If we define a fuzzy set F_δ on $X \times \mathcal{P}(X) \times [0,+\infty]$ as follows:*

$$F_\delta(x,A,t) = 0 \qquad \text{if } t \leq \delta(x,A),$$
$$F_\delta(x,A,t) = 1 \qquad \text{if } t > \delta(x,A),$$

then (X,F_δ) is a FA-space.

Proof. FA1, FA2 and FA3 are immediate. In order to prove FA4, if $A,B \subseteq X$ we can suppose $\min(\delta(x,A),\delta(x,B)) = \delta(x,A) \leq \delta(x,B)$ and, then $\max(F(x,A,t),F(x,B,t)) = F(x,A,t)$. Thus $F_\delta(x,A\cup B,t) = 0$ for $t \leq \delta(x,A)$; and $F_\delta(x,A\cup B,t) = 1$ for $t > \delta(x,A)$. This completes the proof of FA4.

For every $r \in [0,+\infty)$, the set $A^{(r)}$ is the same in **A-space** and in **FA-space**, since

$$A^{(r)}_{A-space} = \left\{ x \in X \ / \ \delta(x,A) \leq r \right\}$$

$$= \left\{ x \in X \ / \ F_\delta(x,A,t) = 1, \forall t > r \right\} = A^{(r)}_{FA-space}.$$

Let us suppose that $s > r$. Since $\delta(x,A) \leq \delta(x,A^{(r)}) + r$, thus for $t \leq \delta(x,A) - s$, we have

$$t + r < t + s \leq \delta(x,A) \leq \delta(x,A^{(r)}) + r;$$

i.e., $t < \delta(x,A^{(r)})$ and $F_\delta(x,A,t+s) = F_\delta(x,A^{(r)},t) = 0$ for $t + s \leq \delta(x,A)$.

Otherwise, if $t + s > \delta(x,A)$, then $F_\delta(x,A,t+s) = 1 \geq F_\delta(x,A^{(r)},t)$. This completes the proof.

Theorem 2.3 (FM-space $\hookrightarrow$ FA-space) *Let $(X,M,*)$ be a FM-space. If we define a fuzzy set F on $X \times \mathcal{P}(X) \times [0,+\infty]$ for every $x \in X$, $t \in [0,\infty]$ and $A \subseteq X$ as follows:*

$$F_M(x,A,t) = \sup_{a \in A} M(x,a,t),$$

then (X,F_M) is a FA-space.

Proof. By FM1 and FM2, conditions FA1 and FA3 hold.

The proof of the second condition FA2 follows from the definition of $supA$ for empty sets, $sup\emptyset = 0$. If $x \in X$ and $A,B \subseteq X$ are subsets of X, then

$$\mathbf{Max}\left(\sup_{a \in A} M(x,a,t),\ \sup_{b \in B} M(x,b,t) \right) = \sup_{a \in A \cup B} M(x,a,t).$$

Consequently FA4 is proved.

We will prove now FA5. Let $x \in X$, $A \in \mathcal{P}(X)$ and $r \in [0,\infty]$. If $A = \emptyset$, the condition FA5 is trivial. It is enough to prove that the property FA5 holds true for every $A \neq \emptyset$. Since, for every $x,y,z \in X$, $t,s > 0$

$$M(x,z,t+s) \geq M(x,y,t) * M(y,z,s);$$

in particular,

$$M(x,y,t+s) \geq M(x,a,t) * M(a,y,s)$$

for every $x \in X$, $a \in A^{(r)}$, $y \in A$, $t > 0, s > r \geq 0$.

Taking supremum over $y \in A$, then

$$\begin{aligned} F_M(x,A,t+s) &= \sup_{y \in A} M(x,y,t+s) \\ &\geq M(x,a,t) * \sup_{y \in A} M(a,y,s) \\ &= M(x,a,t) * F_M(a,A,s). \end{aligned}$$

Since, $a \in A^{(r)}$, then $F_M(a,A,s) = 1$ for $s > r$. Thus

$$F_M(x,A,t+s) \geq M(x,a,t).$$

Taking supremum over $a \in A^{(r)}$, we have

$$F_M(x,A,t+s) \geq F_M(x,A^{(r)},t).$$

Note that additionally the following condition is satisfied:

FA7 Let $x,y \in X$ be such that $F_M(x,\{y\},t) = 1$ for every $t > 0$, then $x = y$.

3 Characterization

Now we ask about the maps $F : X \times \mathcal{P}(X) \times [0,+\infty] \to [0,1]$ that could provide to X of a *FA*-space structure. In this cases, it could be interesting to weaken axioms FA4 and FA5.

Proposition 3.1 *Let X be a non-empty set, $F : X \times \mathcal{P}(X) \times [0,+\infty] \to [0,1]$ be a map satisfying axiom FA2 and such that $F(x,A,t) = 1$ for every $t > 0$ and $x \in A \subseteq X$. Then:*
The axiom FA4 is equivalent to

FA4′. $F(x,A\cup B,t) \geq \mathbf{Max}(F(x,A,t),F(x,B,t))$ *for $t > 0$, for every non-empty subsets $A,B \subseteq X$ such that $A \neq B$ and $A \cup B \subset X$, and for each $x \in X \setminus (A \cup B)$.*

The axiom FA5 is equivalent to

FA5′. $F(x,A,t+s) \geq F(x,A^{(r)},t)$ *for $t > 0$, for every $s > r \geq 0$, $\emptyset \neq A \subset X$ and for each $x \in X \setminus A$.*

Proof. We prove that there exists some cases in which FA4 and FA5 are always true.

Property FA4′. If A is empty, then $F(x,A\cup B,t) = F(x,B,t)$ and, by the axiom FA2

$$\mathbf{Max}(F(x,A,t),F(x,B,t)) = \mathbf{Max}(F(x,\emptyset,t),F(x,B,t))$$

$$= \mathbf{Max}(0,F(x,B,t)) = F(x,B,t).$$

A similar conclusion is valid for $B = \emptyset$. We will suppose that A and B are non-empty sets. If $x \in A \cup B$, then, by hypothesis,

$$F(x,A\cup B,t) = 1 \geq \mathbf{Max}(F(x,A,t),F(x,B,t)).$$

ISBN: 978-989-95079-6-8

We can suppose that $x \notin A \cup B$, and thus $A \cup B = X$ is impossible. So, $A \cup B \subset X$. If $A = B$, the inequality is trivial.

Property FA5′. If $r = +\infty$, the inequality $F(x,A,+\infty) \geq F(x,A^{(+\infty)},t)$ is obvious. Therefore we can suppose that $r < +\infty$. If A is empty, then for each $r > 0$,

$$\emptyset^{(r)} = \left\{ x \in X \;/\; F(x,\emptyset,t) = 1, \forall t > r \right\} = \emptyset,$$

and then in FA5 we have the equality. We can suppose that A is non-empty. If $x \in A$, by hypothesis, $F(x,A,t) = 1$ for all $t > 0$. Then

$$F(x,A,t+s) = 1 \geq F(x,A^{(r)},t).$$

Consequently, we can suppose that $x \notin A$.

4 Properties

The *closure* of a subset A of a $FA-space$ (X,F) can be expressed by

$$\overline{A} = A^{(0)} = \{x \in X : F(X,A,t) = 1, \forall t > 0\}.$$

Although the set X has not been endowed with a topology, the following definition can be considered.

Definition 4.1 *A is called **closed** if $\overline{A} \subseteq A$.*

Lemma 4.2 *Let (x,F) be a **FA-space**, $x \in X$ and $A \subseteq X$.*

1. *If $a \in A$, then $F(a,A,t) = 1$ for each $t > 0$.*
2. *If $a \in A$, then $F(x,a,t) \leq F(x,A,t)$ for each $t > 0$.*
3. *If $A \subseteq B \subseteq X$, then $F(x,A,t) \leq F(x,B,t)$ for each $t > 0$.*
4. $\sup_{a \in A} F(x,a,t) \leq F(x,A,t)$ *for each $t > 0$.*
5. *If $A \subseteq B \subseteq X$, then $A^{(r)} \subseteq B^{(r)}$ for each $r \geq 0$.*
6. *Let $0 \leq r \leq s$, then*

$$A \subseteq \overline{A} = A^{(0)} \subseteq A^{(r)} \subseteq A^{(s)} \subseteq \cdots \subseteq X.$$

   In particular, $F(x,A,t) \leq F\left(x,\overline{A},t\right)$.

7. *A is closed if and only if $\overline{A} = A$.*
8. *If $r \geq 0$, then $\emptyset^{(r)} = \emptyset$.*
9. *$x \in A^{(r)}$ if and only if $x \in A^{(r+s)}$ for every $s > 0$.*
10. *The infimum $\inf_{[0,+\infty)}(\{r \geq 0 \,/\, x \in A^{(r)}\})$ is really a minimum, i.e.,*

$$A^{(r)} = \bigcap_{r<s} A^{(s)}.$$

11. *$F(x,A,\cdot)$ is a nondecreasing function.*

Proof.

1. By the axioms FA3 and FA4, if $a \in A$, we have

$$F(a,A,t) = F(a,A \cup \{a\},t)$$

$$\geq \mathbf{Max}(F(a,A,t),F(a,\{a\},t)) = \mathbf{Max}(F(a,A,t),1) = 1.$$

2. By the axiom FA4, if $a \in A$, we have

$$F(x,A,t) = F(x,B \cup \{a\},t)$$

$$\geq \mathbf{Max}(F(x,B,t),F(x,\{a\},t)) \geq F(x,\{a\},t),$$

   where $B = A \backslash \{a\}$.

3. If $A \subseteq B$, then $A \cup B = B$, and thus

$$F(x,B,t) = F(x,A \cup B,t) \geq \mathbf{Max}(F(x,A,t),F(x,B,t))$$

$$\geq F(x,A,t).$$

4. By 2, if $a \in A$, then $F(x,a,t) \leq F(x,A,t)$. Taking supremum

$$\sup_{a \in A} F(x,a,t) \leq F(x,A,t).$$

5. If $A \subseteq B$, by 3

$$F(x,A,t) \leq F(x,B,t).$$

   If $x \in A^{(r)}$, $F(x,A,t) = 1$ for each $t > r$ and then $F(x,B,t) = 1$ for each $t > r$. That is, $x \in B^{(r)}$.

6. If $a \in A$, by 1, $F(a,A,t) = 1$ for each $t > 0$. Thus $a \in \overline{A}$.
   Let $r \leq s$. If $x \in A^{(r)}$, $F(x,A,t) = 1$, for each $t > r$, then, in particular, $F(x,A,t) = 1$ for each $t > s$. Thus $x \in A^{(s)}$.

7. Trivial.

8. By FA2, it is immediate.

9. $x \in A^{(r)}$ if and only if $F(x,A,t) = 1$ for all $t > r$. This condition is equivalent to the following for each $s > r$,

$$F(x,A,t) = 1 \text{ for all } t > s.$$

   Then,

$$x \in A^{(r)} \Leftrightarrow x \in \bigcap_{r<s} A^{(s)} \Leftrightarrow x \in \bigcap_{s>0} A^{(r+s)}.$$

   That is, $x \in A^{(r)}$ if and only if $x \in A^{(r+s)}$ for all $s > 0$.

10. By 3, 6 and FA5

$$F(x,A,t+s) \geq F(x,A^{(r)},t) \geq F(x,A,t),$$

    if $s > 0$.

Next, the previous properties are refined.

Lemma 4.3 *Let (X,F) be a FA-space, $x \in X$, $t \geq 0$ and $A \subseteq X$.*

1. *$F(x,A,t) = F\left(x,\overline{A},t\right)$.*
2. *The closure $\overline{A}$ is a closed set, that is, $\overline{\overline{A}} = \overline{A}$.*
3. *A is closed if and only if for each $x \in X$ satisfying $F(x,A,t) = 1$, with $t > 0$, then $x \in A$.*
4. *$A^{(r)}$ is closed for every $r \geq 0$.*
5. *(X,F) satisfies the axiom FA7 if and only if the points are closed subsets, that is, $\overline{\{x\}} = \{x\}$.*
6. *$(A^{(r)})^{(s)} \subseteq A^{(r+s)}$ for every $r,s \geq 0$.*

ISBN: 978-989-95079-6-8

Proof.

1. Since $A \subseteq \overline{A}$, by 3 of lemma 4.2, $F(x,A,t) \leq F\left(x,\overline{A},t\right)$. Using axiom FA5 with $r=0$, it holds
$$F(x,A,t+s) \geq F(x,A^{(0)},t) = F(x,\overline{A},t)$$
for $s>0$.
Taking infimum on s, we have $F(x,A,t) \geq F(x,\overline{A},t)$.

2. By 6 of lemma 4.2, $\overline{A} \subseteq \overline{\overline{A}}$. Let $x \in \overline{\overline{A}} = A^{(0)^{(0)}}$. This is equivalent to
$$F\left(x,A^{(0)},t\right) = 1$$
for all $t>0$. Using axiom FA5 with $r=0$,
$$F(x,A,t+s) \geq F(x,A^{(0)},t) = 1$$
for $t,s>0$. Then $x \in A^{(0)} = \overline{A}$.

3. Trivial.

4. Let $x \in X$ be such that $F\left(x,A^{(r)},t\right) = 1$ for all $t>0$. By FA5,
$$F(x,A,t+s) \geq F\left(x,A^{(r)},t\right) = 1$$
for all $t>0$, $s>r\geq 0$. In particular $x \in A^{(r)}$. The previous property implies that $A^{(r)}$ is closed.

5. Trivial.

6. Let $r,s \geq 0$ and $x \in (A^{(r)})^{(s)}$, that is,
$$F\left(x,A^{(r)},t\right) = 1$$
for all $t>s$. By FA5,
$$F(x,A,t+s) \geq F\left(x,A^{(r)},t\right) = 1$$
for $t>s>r\geq 0$. Thus $x \in A^{(r+s)}$.

The following proposition is an interesting property related to closures of subsets of FA-spaces.

Proposition 4.4 *Let (X,F) be a FA-space, $r\geq 0$ and $A,B \subseteq X$ subsets of X.*

1. $A^{(r)} \cup B^{(r)} \subseteq (A\cup B)^{(r)}$.
2. *If (X,F) verifies the equality in FA4, then $(A\cup B)^{(r)} = A^{(r)} \cup B^{(r)}$.*

Proof. Since $A \subseteq A\cup B$ and $B \subseteq A\cup B$, by 5 in lemma 4.2,
$$A^{(s)} \subseteq (A\cup B)^{(s)}, \qquad B^{(s)} \subseteq (A\cup B)^{(s)}.$$
Then $A^{(s)} \cup B^{(s)} \subseteq (A\cup B)^{(s)}$.

Suppose that the equality holds in FA4. If $x \in (A\cup B)^{(r)}$, then
$$\mathbf{Max}\left(F(x,A,t),F(x,B,t)\right) = F(x,A\cup B,t) = 1,$$
for all $t>r$.

Then, for each $t>r$,
$$F(x,A,t) = 1 \text{ or } F(x,B,t) = 1.$$
Suppose that $t>r$ is such that $F(x,A,t)=1$. If $s>0$, then
$$F(x,A,t+s) = F(x,A^{(r)},t) \geq F(x,A,t) = 1.$$
Then $F(x,A,t)=1$ for all $t>r$; i.e., $x \in A^{(r)}$.

Consequently, $x \in A^{(r)}$ or $x \in B^{(r)}$and then $x \in A^{(r)} \cup B^{(r)}$.

References

[1] Lotfi A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[2] A. George and P. Veeramani. On some results in fuzzy metric spaces. *Fuzzy Sets Syst*, 64:395–399, 1994.

[3] O. Kaleva and S. Seikkala. On fuzzy metric spaces. *Fuzzy Sets Syst*, 12:215–229, 1984.

[4] D. Mihet. A banach contraction theorem in fuzzy metric spaces. *Fuzzy Sets Syst*, 144:431–439, 2004.

[5] I. Kramosil and J. Michalek. Fuzzy metrics and statistical metric spaces. *Kybernetika*, 11:336–344, 1975.

[6] R. Lowen. *Approach spaces. The missing link in the topology-uniformity-metric triad*. Oxford Mathematical Monographs. Oxford University Press, 1997.

[7] D. Vaughan R. Lowen, M. Sioen. Completing quasi-metric spaces - an alternative approach. *Houston J. Math.*, 29:113–136, 2003.

ISBN: 978-989-95079-6-8

Decision Making in Competitive Location using Fuzzy Sets

Clara M. Campos[1], Lourdes Canós[2], María J. Canós[3],
Vicente Liern[3], José A. Moreno[4], Dolores R. Santos-Peñate[5]

1 Dep. de Economía Financiera y Contabilidad,
University of La Laguna. 38271 La Laguna, Spain
2 Dep. de Organización de Empresas, Economía Financiera y Contabilidad,
Universidad Politécnica de Valencia. 46022 Valencia, Spain
3 Dep. de Economía Financiera y Matemática,
University of Valencia. 46071 Valencia, Spain
4 Dep. de Estadística, I.O. y C. Instituto Universitario de Desarrollo Regional,
University of La Laguna. 38271 La Laguna, Spain
5 Dep. de Métodos cuantitativos en Economía y Gestión,
University of Las Palmas de G.C., 35017 Las Palmas de G.C., Spain

Email: ccampos@ull.es, loucada@omp.upv.es, {maria.j.canos,vliern}@uv.es, jamoreno@ull.es, drsantos@ulpgc.es

Abstract—*This paper deals with the competitive location problems using fuzzy sets. The basic notions on fuzzy optimization and linear programming using fuzzy sets are briefly reviewed. The standard leader-follower location problem and its linear mathematical programming formulation are described. The works appeared in the literature concerning the use of fuzzy sets are analyzed.*

Keywords— Fuzzy Optimization, Competitive Location, Leader-Follower problem.

1 Introduction

Competitive location models represent the competition between two or more firms which provide goods or services to customers. The natural objective of each firm is to maximize its profit, which is often replaced by the maximization of the market share. That is, each firm tries to capture as many customers (or demand) as possible. The problem of locating facilities in a competitive environment has been addressed in several papers in the field of Operations Research, Management Science, Regional Economy and Economic Geography, see e.g. Craig et al. (1984) [9], Drezner (1995) [13], Eiselt and Laporte (1989,1996) [14][15], Friesz et al. (1988) [16], Hakimi (1990) [17], Plastria (2001) [22], Santos-Peñate et al. (2007) [23], Spaulding and Cromley (2007) [25], Till (1992, 2000) [26][27], and the references therein.

In many cases, the real competitive contexts cannot be described with precision. The complex, subjective, and dynamic behavior of real customers produces a high degree of uncertainty related to their decision making process. Consequently the most common scenario corresponds to problems where the information of the parameters and variables is vague or imprecise. Fuzzy sets theory is a rigorous and effective instrument to treat problems where the necessary information is imprecise. In fuzzy optimization two families of problems exist. On the one hand we have problems where the objective function and the constraints are fuzzy and, on the other hand, those where the coefficients are fuzzy. Fuzzy functions are characterized by their membership function. Fuzzy decisions are the intersection of fuzzy sets corresponding to the objective function and to the constraints. Linguistic variables can also be considered (See Zadeh [34]).

In this work we consider the application of fuzzy optimization methodologies to competitive location problems. The paper is organized as follows. In section 2, we describe the fuzzy optimization models. The basic concepts of competitive location are presented in Section 3. Section 4 is devoted to the solution approaches to competitive location problems using fuzzy optimization and fuzzy sets. Finally, Section 5 contains the conclusions.

2 Fuzzy Optimization Problems

Optimization in their most general form involves finding optimal solutions according to stated criteria. This task is usually formulated as optimization problems using objective functions and constraints. The procedures use mathematical properties of the objective and constraints functions. In practice, however, many situations lack the exact information that is needed in the problem, including its objective and constraints, or in other cases, where it is unreasonable to access such specific constraints or clearly defined objective functions. In these situations it is advantageous to model and solve the problem using soft computing and fuzzy techniques.

An *optimization problem* consists of finding the value of the *decision variables* so that an *objective function* is minimized or maximized when the possible values of the variables are subject to a set of *constraints*. The objective function, denoted by f, is defined on a set of solutions denoted by X and the constraints are given by a vector function g in such a way that the problem is formulated as follows:

ISBN: 978-989-95079-6-8

$$\begin{aligned} &\text{Minimize} \\ &\quad f(\vec{x}) \\ &\text{subject to:} \\ &\quad g(\vec{x}) \le \vec{0} \\ &\quad \vec{x} \in X \end{aligned} \tag{1}$$

where $\vec{x} = (x_1, x_2, ..., x_n) \in \mathbb{R}^n$ is the vector of decision variables, f and g are defined on $\mathbb{R}^n$ and X is a subset of $\mathbb{R}^n$. The mathematical formulations are presented in this paper using "minimize" operator because the other case, the maximization case, is analogous.

Among the optimization problems that are included in Mathematical Programming, the models that have received the most attention and have offered the most useful applications in different areas are Linear Programming (LP) models, which is the single objective linear case.

The classic problem of LP is to find the maximum or minimum values of a linear function subject to constraints that are represented by linear equations or inequalities. The most general formulation of this problem is:

$$\begin{aligned} &\text{Minimize} \\ &\quad \vec{c}^{\,t}\vec{x} \\ &\text{subject to:} \\ &\quad A\vec{x} \le \vec{b} \\ &\quad \vec{x} \ge \vec{0} \end{aligned} \tag{2}$$

where $\vec{c} = (c_1, c_2, ..., c_n) \in \mathbb{R}^n$ is the objective vector, $\vec{b} = (b_1, b_2, ..., b_m) \in \mathbb{R}^m$ is the right hand side vector, and $A = [a_{ij}]$ is an (n,m)-matrix where a_{ij} is the coefficient of variable x_j in constraint i. In an economic context where the aim is the minimization of the total cost with a limitation of the resources, $\vec{c}$ is the cost vector, $\vec{b}$ is the vector of resources and A is known as the technological matrix. An equivalent formulation of the linear problem is the following:

$$\begin{aligned} &\text{Minimize} \\ &\quad \sum_{j=1}^{n} c_i x_j \\ &\text{subject to:} \\ &\quad \sum_{j=1}^{n} a_{ij} x_j \le b_i \qquad i = 1, 2, ...m \\ &\quad x_j \ge 0, \qquad\quad j = 1, 2, ..., n \end{aligned} \tag{3}$$

In many real situations not all the constraints and objective functions can be valued in a precise way and in these situations we are dealing with the general problem form of Fuzzy Linear Programming (FLP). FLP is characterized as follows: a_{ij}, b_i and c_j can be expressed as *fuzzy numbers*, x_i as variables whose states are fuzzy numbers, addition and multiplication operates with fuzzy numbers, and the inequalities are among fuzzy numbers. Different FLP models can be considered according to the elements that contain imprecise information, this is the criterion used in the classification proposed in Verdegay (1995) [33] and Cadenas and Verdegay (1999) [1].

If we take as a basis the classification proposed in Verdegay [33], we have models with the fuzzy feasible sets, models with fuzzy goals, models with fuzzy coefficients of objective function, models with coefficients of the technological matrix and fuzzy right hand sides, and totally fuzzy models where all the elements of the problem are fuzzy.

1. Models with fuzzy objective function

These models are those whose objective function is not fully known. In LP problems where the costs are known with imprecision they are represented by an n-dimensional fuzzy vector $\vec{\tilde{c}} = (\tilde{c}_1, \tilde{c}_2, ..., \tilde{c}_n)$, leading to the following model:

$$\begin{aligned} &\text{Minimize} \\ &\quad z = \vec{\tilde{c}}^{\,t} x \\ &\text{subject to} \\ &\quad A\vec{x} \le \vec{b} \\ &\quad \vec{x} \ge \vec{0} \end{aligned} \tag{4}$$

Evidently, z is also a fuzzy number, but $\vec{x}$ can be a vector of fuzzy or non-fuzzy numbers, and each fuzzy cost is described by its corresponding membership function $\mu_j(c)$. Each coefficient $\tilde{c}_{kj}$ of the objective function is a plane fuzzy number of the L-R type with modal interval $\left[\underline{c}_j, \overline{c}_j\right]$ and membership functions g_j and h_j (which can be linear, parabolic, etc.). Delgado et al. (1989) [11] prove that the solution can be obtained with the multi-objective auxiliary model:

$$\begin{aligned} &\text{Minimize} \\ &\quad z = \left[\vec{c}_1^{\,t}\vec{x}, \vec{c}_2^{\,t}\vec{x}, ..., \vec{c}_{2^n}^{\,t}\vec{x}\right] \\ &\text{subject to} \\ &\quad A\vec{x} \le \vec{b} \\ &\quad \vec{x} \ge \vec{0}, \\ &\quad \alpha \in [0,1], c_{kj} \in \left\{g_j^{-1}(1-\alpha), h_j^{-1}(1-\alpha)\right\} \\ &\quad k = 1, ..., 2^n, j = 1, ..., n. \end{aligned} \tag{5}$$

2. Models with feasible fuzzy set (fuzzy constraints)

This is the case where constraints can be satisfied, and consequently the feasible region, can be defined as a fuzzy set; it should be defined by means of a membership function $\mu : \mathbb{R}^n \to [0,1]$. In such a situation, for each constraint i, a desirable quantity b_i is considered, but the possibility that it is greater is accepted until a maximum $b_i + t_i$ (t_i is referred

ISBN: 978-989-95079-6-8

to as a violation *tolerance level*). This model is represented by

$$\begin{aligned} &\text{Minimize} \\ &\quad z = \vec{c}^{\,t}\vec{x} \\ &\text{subject to:} \\ &\quad A\vec{x} \lesssim \vec{b} \\ &\quad \vec{x} \geq \vec{0} \end{aligned} \tag{6}$$

where the symbol $\lesssim$ indicates the imprecision of the constraints and where each fuzzy constraint $\vec{a}_i^{\,t} x \lesssim \vec{b}_i$ is specified by a membership function in the form:

$$\mu_i\left(\vec{a}_i^{\,t}\vec{x}\right) = \begin{cases} 1 & \text{if } \vec{a}_i^{\,t}\vec{x} < b_i \\ f_i\left(\vec{a}_i^{\,t}\vec{x}\right) & \text{if } b_i \leq \vec{a}_i^{\,t}\vec{x} \leq b_i + t_i \\ 0 & \text{if } b_i + t_i < \vec{a}_i^{\,t}\vec{x} \end{cases} \tag{7}$$

where $\vec{a}_i = \left(a_{i1}, a_{i2}, \ldots, a_{in}\right) \in \mathbb{R}^n$ is the vector of coefficients for constraint *i*. Expression (7) means that, for each constraint *i*, given the level of tolerance t_i, each point (*n*-dimensional vector) $\vec{x}$ *is* associated with a number $\mu_i(\vec{x}) \in [0,1]$ which is known as the degree of fulfillment (or verification) of the constraint *i*. The functions f_i are assumed continuous and monotonous non-decreasing functions. In particular, Verdegay (1982) [32], using the representation theorem for fuzzy sets, proves that the solutions for the case of linear functions f_i can be obtained from the auxiliary model:

$$\begin{aligned} &\text{Minimize} \\ &\quad z = \vec{c}^{\,t}\vec{x} \\ &\text{subject to} \\ &\quad A\vec{x} \leq \vec{b} + (1-\alpha)\vec{t}^{\,t} \\ &\quad \vec{x} \geq \vec{0}, \alpha \in [0,1] \end{aligned} \tag{8}$$

where $\vec{t} = \left(t_1, t_2, \ldots, t_m\right)$.

3. Models with fuzzy goals

A fuzzy optimization problem with fuzzy goals is one whose goal set is fuzzy, that is, it allows the objective function value to be slightly below the minimum goal for a maximization problem, and similarly for a minimization problem. The corresponding linear model is expressed in the following way:

$$\begin{aligned} &\widetilde{\text{Minimize}} \\ &\quad z = \vec{c}^{\,t}\vec{x} \\ &\text{subject to} \\ &\quad A\vec{x} \leq \vec{b} \\ &\quad \vec{x} \geq 0 \end{aligned} \tag{9}$$

If t_0 is the maximum quantity that the function objective should be inferior to the minimum goal c_0, then each vector $\vec{x}$ is associated with a number $\mu_0\left(\vec{x}\right)$, which represents the degree that the decision maker considers to be an achieved goal. It is defined according to the following function:

$$\mu_0\left(\vec{x}\right) = \begin{cases} 1 & \text{if } \vec{c}^{\,t}\vec{x} > c_0 \\ f_0\left(\vec{c}^{\,t}\vec{x}\right) & \text{if } c_0 - t_0 \leq \vec{c}^{\,t}\vec{x} \leq c_0 \\ 0 & \text{if } \vec{c}^{\,t}\vec{x} < c_0 - t_0 \end{cases} \tag{10}$$

where f_0 is a continuous monotonous non-decreasing function. The corresponding satisfactory solutions can be obtained by solving the equivalent problem when a level for the α-cuts is chosen.

4. Models with fuzzy coefficients in the technological matrix

Consider a problem of this type:

$$\begin{aligned} &\text{Minimize} \\ &\quad z = \vec{c}^{\,t}\vec{x} \\ &\text{subject to} \\ &\quad \underset{\sim}{A}\vec{x} \lesssim \underset{\sim}{\vec{b}} \\ &\quad \vec{x} \geq \vec{0} \end{aligned} \tag{11}$$

where the values of the technological matrix and the right hand sides are fuzzy numbers. Fuzzy constraints can also be included. Delgado et al. (1987) [10] also include imprecision in the constraints. They consider the fuzzy relations of the constraints with the application of a ranking function *g* to compare the fuzzy terms. This new formulation is expressed by the auxiliary problem:

$$\begin{aligned} &\text{Minimize} \\ &\quad z = \vec{c}^{\,t}\vec{x} \\ &\text{subject to} \\ &\quad \underset{\sim}{\vec{a}_i}\vec{x} \leq_g \underset{\sim}{b_i} + \underset{\sim}{t_i}(1-\alpha),\ i = 1,\ldots m \\ &\quad \vec{x} \geq \vec{0}, \alpha \in [0,1] \end{aligned} \tag{12}$$

3 Competitive Location

In competitive location problems two or more firms compete by mean of their locations for providing services or products to customers. Several scenarios may be considered:

- No firm operates in the market and firm F_1 wants to enter the market with p_1 facilities.
- There are *s* firms F_i operating in the market, each firm F_i has p_i facilities and a new firm F_{s+1} wants to enter the market with p_{s+1} facilities.
- There are *s* firms F_i operating in the market, each firm F_i has p_i facilities and a firm F_k wants to open $\overline{p}_k$ facilities or to close $\underline{p}_k$ facilities.

In particular, the Stackelberg location model is a standard two-stage problem where a firm, the leader, chooses its location points and then a competitor, the follower, enters

ISBN: 978-989-95079-6-8

the market and decides its locations taking into account the leader position. The problem of the follower is to find the optimal locations given the position of the leader, the problem of the leader is to determine the best locations taking into account the reaction of the follower to any possible strategy of a competitor.

Different objectives can be considered:

- The maximization of the market share.
- The minimization of the competitor market share.
- The maximization of the difference between its market share and the competitor market share.

For modelling the customer's behaviour in the problem we can use different rules:

- *Binary choice rule*. A choice rule where the customers patronize their closest facility.
- *Partially binary choice rule*. A choice rule where each customer patronizes the closest facility of each firm.
- *Proportional choice rule.* With this choice rule, customers patronize all the facilities according to a non-increasing function of the travel distance.
- *Threshold choice rule.* A customer choice rule where a threshold-sensitive behaviour is assumed.

Moreover, whether goods are essential or not, demand are said to be inelastic (constant demand) or elastic (demand varies with distance). The customer behaviour is modelled by their optimization problems that result from the application of the choice rule for them.

For binary inelastic demand, the leader, follower and customer decision problems can be formulated as linear optimization models (see Campos-Rodríguez et al. [6]). The problems are stated as follows.

Let $m = |L|$ be the number of possible facility locations and $n = |C|$ be the number of customer locations. Let $d_{ki} = d(c_k,f_i)$ be the distance between the k-th customer location c_k and the i-th facility point f_i. In addition, let h_k be the total demand of the customers located at c_k. Finally, let $K = \{1, 2, ..., n\} = [1..n]$ denote the index set for the customer locations and $I = \{1, 2, ..., m\} = [1..m]$ denote the index set for the facility locations.

A set Z of location points is identified by a binary vector z with size m. This vector is $z = (z_1, z_2,..., z_m) = (z_i, i \in [1..m])$ where $z_i = 1$ if $f_i \in Z$ and $z_i = 0$ otherwise. Thus the corresponding set is given by $Z = \{ f_i \in L: z_i = 1 \}$.

The decision variables in the leader and follower location problems are the m-vectors x and y of 0-1 or binary decision variables corresponding to the sets X and Y.

For the customer problem, let z_{ki} be the 0-1 decision variables indicating whether the customers located at the k customer location c_k prefer the location f_i for the facility. However, in each customer problem, the sets X and Y are data that are represented by corresponding m-vectors of 0-1 values $\bar{x}$ and $\bar{y}$ for the variables x and y.

For the follower location problem we have an m-vector of values $\bar{x}$ and an m-vector of variables y. The binary choice rule oriented to the leader implies that the scalar product $\bar{x} \cdot \bar{y}$ is zero.

Using the coefficients b_{ij}^k and c_{ij}^k given, for any $i,j \in [1..m]$, $k \in [1..n]$, respectively, by:

$$b_{ij}^k = \begin{cases} 1 & d_{ki} \le d_{kj} \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

and

$$c_{ij}^k = \begin{cases} 1 & d_{ki} < d_{kj} \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

The customer decision problem of the customers at location c_k, for every $k \in [1..n]$, is the linear feasible problem in the variables z_{ki}, for any $i \in [1..m]$, given by:

$$\begin{aligned} &\sum_{i=1}^{m} z_{ki} = 1 \\ &z_{ki} \le \bar{x}_i + \bar{y}_i && i \in [1..m] \\ &z_{ki} \le 1 - c_{ji}^k \bar{y}_j - b_{ji}^k \bar{x}_j - b_{ij}^k \bar{x}_i && i, j \in [1..m] \\ &z_{ki} \in \{0,1\} && i \in [1..m] \end{aligned} \tag{15}$$

For the follower problem consider the coefficients c_{ki} given, for any $k \in [1..n]$ and $i \in [1..m]$, by

$$c_{ki} = \begin{cases} 1 & d_{ki} < \min\{d_{kj} : \bar{x}_j = 1\} \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

Then, the follower problem is the linear optimization problem in the variables z_{ki}, $k \in [1..n]$, $i \in [1..m]$, given by:

maximize

$$\sum_{i=1}^{m} \sum_{k=1}^{n} h_k z_{ki}$$

such that:

$$\begin{aligned} &\sum_{i=1}^{m} y_i = r \\ &\sum_{i=1}^{m} z_{ki} \le 1 && k \in [1..n] \\ &z_{ki} - c_{ki} y_i \le 0 && k \in [1..n]; i \in [1..m] \\ &z_{ki}, y_i \in \{0,1\} && k \in [1..n]; i \in [1..m] \end{aligned} \tag{17}$$

Finally, consider the set J of all the possible selection for the follower. The leader problem is the linear optimization problem in the variables w and for any $k \in [1..n]$, $i \in [1..m]$ and $j \in J$, by

ISBN: 978-989-95079-6-8

$$z_{ki}^{j} = \begin{cases} 1 & d_{ki} \le \min\{d_{kj} : j \in J\} \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

given by:

minimize

w

such that:

$$\begin{aligned} & \sum_{i=1}^{m} x_i = p \\ & \sum_{k=1}^{n} h_k - \sum_{i=1}^{m}\sum_{k=1}^{n} h_k z_{ki}^{j} \le w \quad j \in J \\ & \sum_{i=1}^{m} z_{ki}^{j} \le 1; j \in J \qquad k \in [1..n] \\ & z_{ki}^{j} - c_{ki}^{j} x_i \le 0 \qquad k \in [1..n]; i \in [1..m]; j \in J \\ & z_{ki}^{j}, x_i \in \{0,1\} \qquad k \in [1..n]; i \in [1..m]; j \in J \\ & w \ge 0 \end{aligned} \quad (19)$$

See Campos-Rodríguez et al. (2009) [6] for details.

4 Solution Approaches

Several kind of uncertainty have been considered in Location problems (Benati (2000) [2], Benati and Hansen (2000) [1], Devletoglou (1965) [12], and Sikdera et al. (2007) [24]). Fuzzy sets have been used in several location problems (see Canós et al. (1998) [7]), In general, the user preferences have uncertainty and the corresponding utilities are uncertain. Therefore their choice must be described using a membership function and the competitive models become fuzzy. On the other hand, travel times to reach the facilities are also uncertain and they must be expressed as triangular fuzzy number. The distances d_{ki} should be replaced in the corresponding model by a triangular fuzzy distance $\tilde{d}_{ki} = (d_{ki}^{1}, d_{ki}^{2}, d_{ki}^{3})$. The using of the α-cuts will deal to a crisp model and the solution will be a function of this parameter α. There are several works in Literature using fuzzy methods for competitive location problems.

Liang *et al.* [18] analyze the optimum output quantity decision analysis of a duopoly market under a fuzzy decision environment. To efficiently handle the fuzziness of the decision variables, the linguistic values, subjectively represented by the trapezoidal fuzzy numbers, are used to act as the evaluation tool of decision variables such as fixed cost and unit variable cost.

Osumi *et al.* [21] investigate a competitive facility location problem where there are two facilities on a linear market. Customers at a demand point utilize the facility which seems to be the nearest one from them. They do not distinguish the small difference between two distances. This preference is formulated by introducing relative fuzzy difference based on the actual distance between two points. This paper considers the problems to find the optimal location for the follower and for the leader.

Uno and Katagiri [31] study a new optimal location problem, called defensive location problem (DLP). In the DLPs, a decision maker locates defensive facilities in order to prevent her/his enemies from reaching an important site, called a core; the DLPs are formulated as bilevel 0-1 programming problems to find Stackelberg solutions. an interactive fuzzy satisfying method is proposed,

Uno *et al.* [28] y [29] analyse competitive facility location problems and consider the case where the set of customers is divided into several subsets or types by investigating their preferences and criteria. Since the preferences and criteria often include the vagueness of human's judgement they express such parameter by fuzzy numbers.

Uno and Masatoshi [31] propose a multi-objective approach for competitive facility location models with fuzzy numbers. In cases where the objective of each firm that locates its own facilities is only to maximize its reward, the location of all facilities is usually crowded on some regions which have many populations or are a hub for all regions. Such a location forces the firms to compete extremely and the market about facilities is insecure. Therefore, they formulate a multi-objective problem in which the decision maker is an arbitrator and whose objectives are rewards for all firms. By using a solution to the problem, they find a good restriction such that more firms can survive.

Considering the imprecision or vagueness in the customer choice rule, the appropriate model for dealing problem (19) is that with fuzzy coefficients in the technological matrix since the coefficients c_{ki}^{j}. Several relaxed choice rules have been proposed to allow a soft behavior of the customers. These rules will result in corresponding membership functions for the coefficients c_{ki}^{j} that provide a particular linear programming problem with fuzzy technological coefficients. The problem is then solved by the corresponding auxiliary problem.

5 Conclusions and further research

Fuzzy sets constitute an appropriate approach to manage the uncertainty that appears in real competitive location problems. Fuzzy competitive location problems need to be clearly formalized and classified. There are several research works in some competitive location problems with fuzzy elements. However, practical solution procedures have to be tested for real problems Chance-constrained or Possibility theory and the classical use of α-cuts.

Acknowledgements

The authors have supported by projects TIN2008-06873-C04-01, TIN2008-06873-C04-02, ECO2008-05589 of Spanish govern and FEDER, and PI042005/044 of Canary govern.

ISBN: 978-989-95079-6-8

References

[1] S. Benati, P. Hansen, 2000. Competitive Location with Random Utilities: Problem Formulation and Algorithms. *Cahiers du GERAD*

[2] S. Benati 2000. NP-hardness of some competitive location models with probabilistic choice rules. *Studies Locational Analysis* 14, 211-231.

[3] J.M. Cadenas, J.L. Verdegay, 1999. Modelos de optimización con datos imprecisos. *Ed. Universidad de Murcia*.

[4] J.M. Cadenas, J.L. Verdegay, 2006. A Primer on Fuzzy Optimization Models and Methods. *Iranian Journal of Fuzzy Systems* 3(1):1-21.

[5] J.M. Cadenas, J.L. Verdegay, 1997. Using fuzzy numbers in linear programming. *IEEE Transactions on Systems, Man, and Cybernetics*, Part B 27(6): 1016-1022.

[6] C. Campos-Rodríguez, J.A. Moreno-Pérez, D.R. Santos Peñate, 2009. An exact procedure and LP formulation for the leader-follower location problem. To appear in TOP

[7] M.J. Canos, C. Ivorra, V. Liern, 1998. Fuzzy methods in location problems. *X Meeting of the EURO Working Group on Locational Analysis*, February 10-13, Murcia (Spain)

[8] A. Charnes, W.W. Cooper, 1959. Chance-Constrained Programming. *Management Science,* 6:73-79.

[9] S. Craig, Ghosh A., McLafferty S. 1984. Models of the retail location process: A review. *Journal of Retailing* 60(1), 5-36.

[10] M. Delgado, J.L. Verdegay, M.A. Vila, 1987. Imprecise costs in mathematical programming problems. *Control and Cybernetics* 16(2):113–121.

[11] M. Delgado, J.L. Verdegay, M.A. Vila, 1989. A general model for fuzzy linear programming. *Fuzzy Sets and Systems* 29:21-29.

[12] N.E. Devletoglou, 1965. A dissenting view of duopoly and spatial competition. *Economica* 32(126), 140-160.

[13] T. Drezner, 1995. Competitive facility location in the plane. In *Facility Location. A survey of applications and methods*, Z. Drezner (Ed.), Springer, Berlin, 285-300.

[14] H.A. Eiselt, G. Laporte, 1989. Competitive spatial models. *European Journal of Operational Research* 39, 231-242.

[15] H.A. Eiselt, G. Laporte, 1996. Sequential location problems. *European Journal of Operational Research* 96, 217-231.

[16] T.L. Friesz, T. Miller, R.L. Tobin, 1988. Competitive network facility location models: A survey. *Papers of the Regional Science Association* 65, 47-57.

[17] S.L. Hakimi,. 1990. Location with spatial interactions: Competitive locations and games. In *Discrete Location Theory*, Mirchandani, P.B. and R.L. Francis (eds.), John Wiley \& Sons, New York, 439-478.

[18] G.-S. Liang, L.-Y. Lin, C.-F. Liu, 2008. The optimum output quantity of a duopoly market under a fuzzy decision environment *Computers & Mathematics with Applications* 56(5):1176-1187

[19] B. Liu, 2001. Fuzzy random Chance-Constrained Programming. *IEEE Transactions on Fuzzy systems* 9(5):713-720

[20] B. Liu, 2009. *Theory and practice of uncertain programming* (2nd ed.). Springer-Verlag.

[21] S. Osumi, S. Shiode, 2003. Competitive facility location problem with fuzzy relative distance *Journal of Nonlinear Convex Analysis*. 4(1):141-149.

[22] F. Plastria, 2001. Static competitive facility location: An overview of optimisation approaches. *European Journal of Operational Research* 129:461-470.

[23] D.R. Santos-Peñate, R. Suárez-Vega, P. Dorta-González 2007. The leader-follower location model. *Networks and Spatial Economics* 7:45-61.

[24] I.U. Sikdera, A. Gangopadhyay, 2007. Managing uncertainty in location services using rough set and evidence theory. *Expert Systems with Applications* 32(2):386-396

[25] B.D. Spaulding, R.G. Cromley, 2007. Integrating the maximum capture problem into a GIS framework. *Journal of Geographical Systems. Geographical Information, Analysis, Theory and Decision* 9(3): 267-288.

[26] J-C. Till, 1992. Competitive strategies for multi-establishment firms. *Economic Geography* 68(3): 290-309.

[27] J-C. Till, 2000. Network competition and branch differentiation with consumer heterogeneity. *The Annals of Regional Science* 34:451-468.

[28] T. Uno, H. Ishii, S. Saito, S. Osumi, 2001. An introduction to competitive facility location problems with vagueness. *Knowledge-Based Intelligent Information Engineering System and Allied Technologies* 69:1375-1379.

[29] T. Uno, H. Ishii, S. Saito, S. Osumi, 2004. Competitive facility location problem: An algorithm for the problem concerning the existence of multi-type customers, *Central European Journal of Operations Research*, 12(1):79-85.

[30] T. Uno, H. Katagiri, 2008. Single- and multi-objective defensive location problems on a network. *European Journal of Operational Research*, 188(1):76-84

[31] T. Uno, S. Masatoshui, 2003. Multi-Objective Approach for Competitive Facility Location Model with Fuzzy Numbers *Faji Shisutemu Shinpojiumu Koen Ronbunshu* 19:711-714

[32] J.L. Verdegay, 1982. Fuzzy mathematical programming *In: Fuzzy Information and Decision Processes,* Gupta MM, Sanchez E (eds). North-Holland.

[33] J.L. Verdegay, 1995. Fuzzy optimization: models, methods and perspectives. *6th IFSA-95 World Congress*. p. 39-71.

[34] L.A. Zadeh, 1965. Fuzzy sets. *Information and Control*, 8:338-353.

ISBN: 978-989-95079-6-8

Using Syntactic Possibilistic Fusion for Modeling Optimal Pessimistic Qualitative Decision

Salem Benferhat[1] Hadja Faiza Khellaf-Haned[2] Aicha Mokhtari[2] Ismahane Zeddigha[2]

1.CRIL, Université d'Artois
Rue Jean Souvraz, S.P.18 62307 Lens, Cedex France
2.LRIA, Université des Sciences et de la Technologie Houari Boumediene
BP 32 EL ALIA 16111 Bab Ezzouar Alger Algérie
Email: benferhat@cril.univ-artois.fr, hanedfaiza@yahoo.com, {aissani_mokhtari,izeddigha}@yahoo.fr

Abstract— *This paper describes the use of syntactical data fusion to computing possibilistic qualitative decisions. More precisely qualitative possibilistic decisions can be viewed as a data fusion problem of two particular possibility distributions (or possibilistic knowledge bases): the first one representing the beliefs of an agent and the second one representing the qualitative utility. The proposed algorithm computes a pessimistic optimal decisions based on data fusion techniques. We show that the computation of optimal decisions is equivalent to computing an inconsistency degree of possibilistic bases representing the fusion of agent's beliefs and agent's preferences.*

Keywords— Data Fusion, Decision theory, Pessimistic Criteria, Possibilistic Logic.

1 Introduction

The problem of decision on uncertainty is crucial for many applications in artificial intelligence. A decision theory must provide some criteria for representing behaviors of an agent in order to take the optimal decision amongst a set of decisions. Possibility theory represents one of the theory which deals with uncertain information. This paper presents a computation of pessimistic decisions based on syntactic possibilistic fusion operations[1]. A qualitative possibilistic decision model [2] allows a gradual expression of both agent's preferences and knowledge. The preferences and the available knowledge about the world are expressed in ordinal way. In [2], the authors have proposed two qualitative criteria for ordinal decision approaches under uncertainty: the pessimistic and the optimistic decisions criteria. The first one being more cautious than the second one for computing optimal decisions. A method for computing optimal decisions, based on ATMS (Assumption-based Truth Maintenance System), has been proposed in [3]. Using the pessimistic criteria, the procedure is translated into an ATMS problem [4, 5].

Recently, in a companion paper [6] a method has been proposed for computing optimal optimistic decisions using possibilistic fusion modes. This paper is also developed in the same spirit. Indeed, qualitative possibilistic decisions can be viewed as a data fusion problem of two particular possibility distributions: the first one representing the beliefs of an agent and the second one representing the qualitative utility (or agent's preferences). The handling of possibilistic decision raises new issues and it requires additional steps such that the computation of negated possibilistic bases.

The rest of this paper is organized as follow. Section 2 gives a brief backgrounds on possibilistic logic, on qualitative decision problems under uncertainty based on possibilistic logic where agent's beliefs and the preferences are expressed by means of possibilistic bases. In Section 3, we provide a preliminary step for our work. Section 4 presents an efficient and unified way of computing pessimistic qualitative decisions based on syntactic counterpart of data fusion problem. The paper ends with some conclusions in Section 5.

2 Backgrounds

2.1 Possibilistic Logic

This section gives a brief refresher on possibilistic logic. See [7] for more details on possibilistic logic. Let $\mathcal{L}$ be a finite propositional language and Ω be the set of all propositional interpretations. $\phi, \psi, \ldots$ denote propositional formulas. For an interpretation ω and a propositional formula $\phi, \omega \models \phi$ means that ω is a model (in the sense of propositional logic) of ϕ. A possibility distribution [7, 8] π is a mapping from a set of interpretations Ω into the unit interval [0,1]. $\pi(\omega)$ represents the degree of compatibility (or consistency) of the interpretation ω with available pieces of information. Given a possibility distribution π, two dual measures are defined on the set of propositional formulas:

- The possibility (or consistency) measure of a formula ϕ, defined by:

$$\Pi(\phi) = max\{\pi(\omega) : \omega \models \phi \text{ and } \omega \in \Omega\} \quad (1)$$

  which evaluates the extent to which ϕ is consistent with available beliefs expressed by π.

- The necessity (or certainty) measure of a formula ϕ, defined by:

$$N(\phi) = 1 - \Pi(\neg\phi) \quad (2)$$

  which evaluates the extent to which ϕ is entailed by available beliefs.

A possibilistic knowledge base Σ is a set of weighted formulas:

$$\Sigma = \{(\phi_i, \alpha_i) : i = 1, ..., n\},$$

where ϕ_i is a propositional formula and $\alpha_i \in]0, 1]$ represents the certainty level of ϕ_i. Each piece of information (ϕ_i, α_i) of a possibilistic knowledge base can be viewed as a constraint that restricts possibility degrees associated with interpretations [7].

If an interpretation ω satisfies ϕ_i then its possibility degree is equal to 1 (ω is completely compatible or consistent with

ISBN: 978-989-95079-6-8

the belief ϕ_i), otherwise it is equal to $1 - \alpha_i$ (the more ϕ_i is certain, the less ω is possible). In particular, if $\alpha_i = 1$, then any interpretation falsifying ϕ_i, is such that its possibility degree is equal to 0, namely is impossible.

More formally, the possibility distribution associated with a weighted formula (ϕ_i, α_i) is [7] $\forall \omega \in \Omega$:

$$\pi_{(\phi_i,\alpha_i)}(\omega) = \begin{cases} 1 - \alpha_i & if\ \omega \not\models \phi_i \\ 1 & \text{otherwise} \end{cases} \quad (3)$$

More generally, the possibility distribution associated with a qualitative possibilistic knowledge base Σ is: $\forall \omega \in \Omega$:

$$\pi_\Sigma(\omega) = min\{\pi_{(\phi_i,\alpha_i)}(\omega) : (\phi_i, \alpha_i) \in \Sigma\}. \quad (4)$$

2.2 Possibilistic Qualitative Decision

Several works have been proposed for dealing with qualitative decision theory under uncertainty. Some approaches consider only all-or-nothing notions of utility and plausibility [9], others use in addition a preference ordering on consequences [10]. However, in many applications the knowledge bases may be pervaded with uncertainty, and the goals may not have equal priority. Dubois, Le Berre, Prade and Sabbadin [3] have enriched logical view of the decision problem by assigning levels of certainty to formulas in the knowledge bases, and levels of priority to the goals. They have proposed two syntactic approaches based on possibilistic logic, applied on two stratified logical bases that model gradual knowledge and preferences. The first one being more cautious than the second, for computing optimal decisions [3].

It has been shown in [3] that the semantics underlying the two syntactic approaches are in agreement with the two qualitative utility functions proposed in [2].

Let K be a stratified base which represents level of certainty of the knowledge, $K = \{(\phi_i, \alpha_i) : i = 1, ..., n\}$ where $\alpha_i \in S$ such that $(\alpha_i > 0)$ denotes the degree of certainty, and the ϕ_i 's are formulas in $\mathcal{L}$ where decision literals may appear.

Let P be a stratified base expressing preferences, $P = \{(\psi_i, \beta_i) : i = 1, ..., n\}$ where $\beta_i \in S$ such that $(\beta_i > 0)$ is a degree of priority, and the ψ_i's are formulas in $\mathcal{L}$ where decision literals may also appear.

Let $K_{\geq\alpha}$ denotes the set of formulas with certainty at least equal to α . Let $P_{\geq\beta}$ denotes the set of formulas with priority at least equal to β. We also recall that $K_{>\alpha}$ (with $\alpha < 1$) denotes the set of formulas with certainty strictly greater than α and $P_{>\beta}$ (with $\beta < 1$) denotes the set of formulas with priority strictly greater than β.

The certainty degrees and the priority degrees are assumed to be commensurate and assessed to the same linearly order scale S [2]. The top element of S will be denoted by 1, and the bottom element will be denoted by 0.

For any set A of formulas, $A^\wedge$ denotes the logical conjunction of the formulas in A. Let $D = \{l_i\}$ be a set of decision variables, where l_i are distinguished variables of the language $\mathcal{L}$. Let $d \subseteq D$, then the decision $d^\wedge$ is the logical conjunction of literals in the chosen subset. The variables that are not in D are state variables.

Each set of decision d induces a possibility distribution π_{K_d} in the following way [7]:

$$\pi_{K_d}(\omega) = \begin{cases} 1 \text{ if } \forall(\phi_i, \alpha_i) \in K, \omega \models \phi_i \text{ and } \omega \models d^\wedge \\ min_{(\phi_i,\alpha_i)\in K/\omega\models\neg\phi_i}(1 - \alpha_i) \text{ if } \omega \models d^\wedge \\ 0 \text{ if } \omega \not\models d^\wedge \end{cases}$$

Where α_i represents the degrees of necessity of the formulas in the corresponding layers of K_d.

In other hand, the utility function μ_P associated to the preferences base P is built over Ω in a similar way:

$$\mu_P(\omega) = \begin{cases} 1 \text{ if } \forall(\psi_j, \beta_j) \in P, \omega \models \psi_j \\ min_{(\psi_j,\beta_j)\in P/\omega\models\neg\psi_j}(1 - \beta_j) \end{cases}$$

where β_j represents a degree of priority of a formulas in P. Making a decision amounts to choosing a subset d of the decision set D. The objective is to rank-order decisions on the basis of K and P. The pessimistic utility function is expressed in terms of the possibility distribution π_{K_d} and the utility function μ [3]:

$$u_*(d) = min_{\omega\in\Omega} max(1 - \pi_{K_d}(\omega), \mu(\omega)) \quad (5)$$

In the pessimistic case, the decision d must satisfy [1]:

$$K^\wedge_\alpha \wedge d^\wedge \vdash P^\wedge_{>(1-\alpha)} \quad (6)$$

The decision d associated with the most certain part of K entails the satisfaction of the goals, even those with low priority.

3 Syntactic Counterparts of Negated Preference Bases

This section presents a first result of this paper needed for the development of our algorithm. It consists in a characterization of a syntactic counterpart of negated possibilistic base defined by: $\forall \omega \in \Omega, \pi_{n_\Sigma}(\omega) = 1 - \pi_\Sigma(\omega)$.

Let $\Sigma = \{(\phi_i, \alpha_i) : i = 1, ..n\}$ be a possibilistic base. We assume that: $\alpha_0 = 0 < \alpha_1 < ... < \alpha_n$. The following definition gives the possibilistic knowledge base associated with the negation of Σ.

Definition 1. The negated base of $\Sigma = \{(\phi_i, \alpha_i) : i = 1, ..., n\}$ is a possibilistic base, denoted by n_Σ, and defined by:

$$n_\Sigma = \{(d_i, 1 - \alpha_i) : i = 1, ..., n\} \cup \{(\bot, 1 - \alpha_n)\}$$

where $d_i = \neg\phi_i \vee \neg\phi_{i+1} \vee ... \vee \neg\phi_n$.

Example 1. Let $\Sigma = \{(\neg a \vee b, 0.3), (b \vee \neg c, 0.6)\}$. Then, $n_\Sigma = \{((a \wedge \neg b) \vee (\neg b \wedge c), 0.7), (\neg b \wedge c, 0.4)\}$

The following proposition shows that n_Σ is indeed encodes the negation of Σ:

Proposition 1. Let $\Sigma = \{(\phi_i, \alpha_i) : i = 1, ..., n\}$ be a preference base, and n_Σ its negated base obtained using Definition 1. Let π_Σ and π_{n_Σ} be the utility distributions associated with Σ and n_Σ respectively. Then:

$$\forall \omega \in \Omega, \pi_\Sigma(\omega) = 1 - \pi_{n_\Sigma}(\omega).$$

Proof. Recall first that when $\pi_\Sigma(\omega) = \alpha$ (resp $\pi_{n_\Sigma} = \beta$) then ω falsifies formulas of Σ (resp of n_Σ) having a weight equal to α (resp β), but satisfies all formulas of Σ (resp n_Σ) having a weight strictly larger than α (resp β). We distinguish 3 cases:

ISBN: 978-989-95079-6-8

1. $\pi_\Sigma(\omega) = 1$, which means that $\forall(\phi_i, \alpha_i) \in \Sigma, \omega \models \phi_i$. This means that ω falsifies all formulas of n_Σ, and in particular the highest formula α_1, namely $\omega \not\models \alpha_1$. Hence, $\pi_{n_\Sigma}(\omega) = 0 = 1 - \pi_\Sigma(\omega)$.

2. $\pi_\Sigma(\omega) = \alpha_n$. This means that ω satisfies all formulas in Σ except α_n. This also means that ω satisfies all formulas d_i (since d_i is a disjunction of $\neg\phi_n$ and others terms). The only formula falsified by $\omega \in n_\Sigma$ is $(\bot, 1 - \alpha_n)$. Hence, $\pi_{n_\Sigma}(\omega) = 1 - (1 - \alpha_n) = \alpha_n = \pi_\Sigma(\omega)$.

3. $\pi_\Sigma(\omega) = \alpha_j$ with $j < n$. This means that $\omega \not\models \phi_j$. Hence, $\forall i \leq j, \omega \models \phi_j$ (since d_j contains $\neg\phi_j$). $\mu_P(\omega) = \alpha_j$ also means that $\forall k > j, \omega \models \phi_k$, namely $\omega \models \phi_n \wedge ... \wedge \phi_{j+1}$. Hence, $\omega \not\models \neg\phi_n \vee ... \vee \phi_{j+1}$. Namely, $\omega \not\models d_{j+1}$. Therefore: $\pi_{n_\Sigma}(\omega) = 1 - (1 - \alpha_{(j+1)-1}) = \alpha_j = \pi_\Sigma(\omega)$.

The obtained base n_Σ must be put in a clausal form to get C_{n_Σ}.

Example 2. Let n_Σ be the base obtained in example 1. The clausal form of n_Σ is then: $C_{n_\Sigma} = \{(c, 0.4), (a \vee \neg b, 0.7), (a \vee c, 0.7), (\neg b, 0, 7), (\neg b \vee c, 0.7)\}$

4 Computation of Pessimistic Decision

In this section, we propose an algorithm for computing qualitative optimal decision in the pessimistic case. The knowledge base K can be the result of merging several knowledge bases $K_1, K_2, ..., K_n$. The possibilistic distribution associated to K is then obtained by merging possibility distributions associated with $K_1, , K_2, ..., K_n$ using some merging operator /oplus, namely: $\pi_K = \oplus(\pi_{K1}, ..., \pi_{Kn})$. When $\oplus = min$, it has been shown that $K = K_1 \cup ...K_n$. If $\oplus = max$, then $K = \{(\phi_1 \vee ...\phi_n, min(\alpha_1, ..., \alpha_n)) : (\phi_1, \alpha_1) \in K_1, ..., (\phi_n, \alpha_n) \in K_n\}$. For syntactic counterparts of more general operator, see [1]. Similarly, P can be also the result of merging several preference bases.
Now, once K and P are fixed, we propose a syntactic computation of optimal decisions. We recall that an optimal pessimistic decision d maximizing $u_*(d)$ is such that:

$$u_*(d) = min_{\omega\in\Omega} max(1 - \pi_{K_d}(\omega), \mu(\omega)) \quad (7)$$

From equation (6), a first way to syntactically compute optimal decisions is first to compute the counterpart of $1 - \pi_{K_d}$, then compute the counterpart of $max(1 - \pi_{K_d}(\omega), \mu(\omega))$ and lastly compute $u_*(d)$. This approach is not interesting since first in general Σ is more important than P and computation of $max(1 - \pi_{Kd}(\omega), \mu(\omega))$ is more expensive than the computation of the minimum. We propose to explore another possibility by noting that equation (6) is equivalent to:

$$u_*(d) = 1 - max_{\omega\in\Omega} min(\pi_{K_d}(\omega), 1 - \mu(\omega)) \quad (8)$$

Besides, the syntactic counterpart of $min(\pi_1(\omega), \pi_2(\omega))$ is the possibilistic base $\Sigma_{min} = \Sigma_1 \cup \Sigma_2$ [1].
Thus, combining these results, the corresponding base Σ_{min} associated to $min(\pi_{K_d}(\omega), 1 - \mu(\omega))$ is the possibilistic base $K \cup n_P \cup \{(d, 1)\}$, such that n_P is the possibilistic base corresponding to the utility function $1 - \mu(\omega)$. We recall that:

$$u_*(d) = 1 - max_{\omega\in\Omega}\{min(\pi_{K_{d(\omega)}}, 1 - \mu(\omega))\} \quad (9)$$

Lemma 1. Let n_P be the negated based, obtained using Definition 1, of the preferences base P . Then, the syntactic counterparts of $min(\pi_{K_d(\omega)}, 1 - \mu(\omega))$ is $K \cup n_P$.

On the other hand, the inconsistency degree $Inc(K \cup C_{n_P})$ of a possibilistic base $K \cup C_{n_P}$, where C_{n_P} represents the conjunctive form of the base n_P, is defined as follow [7]:

$$Inc(K \cup C_{n_P}) = 1 - max\{\pi_{K_d}(\omega)\}$$

Lemma 2. The pessimistic utility function associated to decision d is :

$$u_*(d) = Inc(\Sigma_{min})$$

Where $Inc(\Sigma_{min})$ represents the inconsistency degree of the base $K \cup C_{n_P} \cup \{(d, 1)\}$.
Then, the computation of optimal pessimistic decisions is obtained using the Algorithm 1.
The computation of inconsistency degree is performed by a call to the function $incons(B \cup \{(\neg\phi, 1)\}, Inc, bool)$. This function, given by the Algorithm 2, has three parameters: a stratified knowledge base, an integer representing current inconsistency degree and a boolean variable.
The base $B^*_{\geq\alpha_r}$ is defined as the classical projection of the $\alpha_cut\ B_{\geq\alpha_r}$. $B^*_{\geq\alpha_r} = \{\phi : (\phi, \beta) \in B, \beta \geq \alpha_r\}$.
The function incons is on the extension of dichotomous algorithm for computing inconsistency proposed in [11] on which some adaptations have been made in order to compute optimal decision. Indeed, one the inconsistency degree of $K \cup C_{n_P} \cup \{(d_i, 1)\}$ can be greater than the inconsistency degree of $K \cup C_{n_P} \cup \{(d_{j\in[1,i-1]}, 1)\}$, the algorithm stops.
In terms of complexity, the proposed algorithm is based on

Algorithm 1 Computation of optimal pessimistic decision

```
Input: K knowledge base
       C_{n_P} conjunctive form of n_p
       N number of decision variables
       D = {d_{i∈[1,n]}} set of decisions
Output: decision
Begin
i ← 1
max ← 0
inc ← 1
for i = 1 to N do
  incons(K ∪ C_{n_P} ∪ {(d_i, 1)}, inc, bool)
  if (bool=true) then
    if (inc > max) then
      max ← inc
      decision ← {d_i}
    else if inc = max then
      decision ← decision ∪ {d_i}
    end if
  end if
end for
RETURN (decision)
End
```

ISBN: 978-989-95079-6-8

an inconsistency degree computation which is known to be NP-hard and requires in the worst case $[log_2 m]$ satisfiability checks, where m is the number of different valuations involved in $K \cup C_{n_P} \cup \{(d_i, 1)\}$ using any prouver for the propositional satisfiability problem SAT.

Algorithm 2 Function $incons(B \cup \{(\neg\phi, 1)\}, inc, bool)$

Input B:stratified base
ϕ:weighted formula
n: number of strates in base B
Output inc: inconsistency degree
bool:boolean
Begin
$l \leftarrow 0$ {initially pointed on the last strate of the base}
$u \leftarrow m$ {initially pointed on first strate of the base}
$bool \leftarrow true$
while $((l < u)$ and $(bool = true))$ **do**
  $r \leftarrow [(l + u)/2]$ {pointer uses for dichotomy}
  if $(B^*_{\geq \alpha_r} \wedge \neg\phi$ consistent) **then**
    $u \leftarrow r - 1$ {check the inconsistency in the most big base}
  else
    if $(inc \geq \alpha_r)$ **then**
      $l \leftarrow r$
    else
      $bool \leftarrow false$
    end if
  end if
end while
if ($\alpha_r \geq inc$) **then**
  $bool \leftarrow false$
end if
if (bool = true) **then**
  $inc \leftarrow \alpha_r$
end if
RETURN(inc,bool)
End

Example 3. Let us exemplify the algorithm on an example initially proposed in [12] and reused in [2]. The example is about taking an umbrella or not, knowing that the sky is cloudy.
The literals of the language are:

- It rains: r
- Being wet: w
- Taking an umbrella: um
- The sky is cloudy: c

The stratified knowledge base is $K = \{(\neg um \vee \neg w, 1), (\neg r \vee um \vee w, 1), (r \vee \neg w, 1), (c, 1), (\neg c \vee r, 0.6)\}$.
The stratified preference base is $P = \{(\neg um, 0.2), (\neg w, 1)\}$. We do not like to take an umbrella, but it is more important to be dry.
The set of decisions is $D = \{um, \neg um\}$, i.e taking an umbrella or not.

The preliminary step consists to compute the conjunctive normal form C_{n_P} of the negated form n_P of the preference base P. Then, the algorithm applies as follow:

- initially, $i \leftarrow 1$, $inc \leftarrow 1$, max $\leftarrow 0$ and $d_1 \leftarrow um$. We have to call to the function $incons(K \cup C_{n_P} \cup \{(um, 1)\}, inc, bool)$. This function return inc=0.1 and bool=true. In this case, $inc > max$, so $decision \leftarrow \{um\}$ and $max \leftarrow inc$.
- In the next step, $i \leftarrow 2$ and $d_2 \leftarrow \neg um$. The call of the function $incons(K \cup C_{nP} \cup \{(\neg um, 1)\}, inc, bool)$ returns $inc = 0.1$ and bool is true. As inc=max, the set of optimal decisions do not change.

Thus, the best pessimistic solution of the umbrella problem is to take an umbrella.

5 Conclusion

The main contribution of this paper is a proposition of a new approach to compute a qualitative possibilistic pessimistic decision problem. This problem is viewed as the one of computing inconsistency degrees of particular bases in the framework of possibilistic logic. The application exploits the syntactic counterparts of data fusion techniques.
Our approach is an alternative to the ATMS-based solution, proposed in [3] and it avoids the use of the ATMS to compute the pessimistic optimal qualitative decision which is known to be not efficient when it deals with an important number of variables. Then, in the pessimistic case, the knowledge base and the preferences base are fused and the decision problem is translated into the computation inconsistency degree of a specific base. This process leads to use a logical technique more adapted than the ATMS one.

References

[1] S. Benferhat, D. Dubois, and H. Prade. From semantic to syntactic approaches to information combination in possibilistic logic. *Aggregation and Fusion of Imperfect Information, Physica Verlag*, pages 141–151, 1997.

[2] D. Dubois and H. Prade. Possibility theory as a basis for qualitative decision theory. In *14th International Joint Conference on Artificial Intelligence (IJCAI'95), Montreal*, pages 1924–1930, 1995.

[3] D. Dubois, D. Le Berre, H. Prade, and R. Sabbadin. Logical representation and computation of optimal decisions in a qualitative setting. In *AAAI-98*, pages 588–593, 1998.

[4] J. De Kleer. An assumption-based truth maintenance systems. *Artificial Intelligence Jounral*, 28:127–162, 1986.

[5] J. De Kleer. Extending the assumptions-based truth maintenance systems. *Artificial Intelligence Journal*, 28:163–196, 1986.

[6] S.Benferhat, F.Khellaf-Haned, A.Mokhtari, and I.Zeddigha. On the use of syntactic possibilistic fusion for computing optimistic qualitative decisions. In *Proceedings of JCIS/FTT Salt Lake City, USA*, 2007.

[7] D. Dubois, J. Lang, and H. Prade. Possibilistic logic. *In Handbook of Logic in Artificial Intelligence and Logic Programming*, D. Gabbay et al., eds, 3, Oxford University Press:pages 439–513, 1994.

ISBN: 978-989-95079-6-8

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

[9] B. Bonet and H. Geffner. Arguing for decisions: a qualitative model of decision making. In *Proceedings of the 12th conference on uncertainty in Artificial Intelligence (UAI'96)*, pages 98–105. Portland, Oregon: Morgan Kaufman, 1996.

[10] R. Brafman and M. Tennenholtz. On the axiomatization of qualitative decision criteria. In *Proceedings of the 14th National Conference on Artificial Intelligence (AAAI'97)*, pages 76–81, 1997.

[11] J. Lang. Possibilistic logic: Complexity and algorithms. *In Handbook of Defeasible Reasoning and Uncertainty Management Systems*, 5:179–220, 2000.

[12] C. Boutilier. Toward a Logic for Qualitative Decision Theory. In *Proceedings of the 4th International Conference on Principles of Knowledge Representation, (KR'94)*, pages 75–86, 1994.

ISBN: 978-989-95079-6-8

A Fuzzy-Semiotic Framework for Modeling Imprecision in the Assessment of Depression

Mila Kwiatkowska[1] Krzysztof Kielan, MD[2] Krzysztof Michalik[3]

1 Department of Computing Science, Thompson Rivers University
Kamloops, British Columbia, Canada
2 Department of Psychiatry, Diana, Princess of Wales Hospital, Grimsby, UK
3 Department of Knowledge Engineering, University of Economics, Katowice, Poland
Email: mkwiatkowska@tru.ca, krystof.kielan@nelctp.nhs.uk, k.michalik@aitech.pl

Abstract—*This paper presents a new framework to model an assessment process for a complex and multidimensional syndrome such as depression. Since the measurements of depression are inherently imprecise, we explicitly model the context of the assessment process, and we analyze various aspects of imprecision (syntactic, semantic, and pragmatic). The framework is based on fuzzy logic and semiotics. The fuzzy-logic approach allows for the representation of quantitative imprecision of the measurements and the semiotic approach allows for the representation of the qualitative imprecision of the concepts. We have applied this fuzzy-semiotic framework to two types of clinical measurements: the rating by the experts and the filling out of self-administered questionnaires. The proposed framework provides a conceptual foundation for the construction of a medical decision support tool.*

Keywords— Assessment of depression, fuzzy logic, imprecision, modelling, medical decision, semiotics.

1 Introduction

Depression is a common mental disorder characterized by sad moods, loss of interest or pleasure, feelings of guilt or low self-worth, disturbed sleep or appetite, low energy, and poor concentration. Depression represents a major public health problem, and it has a high prevalence worldwide and high societal costs [1]. The World Health Organization (WHO) reports that "depression is affecting about 121 million people worldwide;" furthermore, WHO ranks depression as one of the most burdensome diseases in terms of disability costs [1, 2]. The assessment and treatment of depression are lengthy and complex, and often the patients wait a long time to gain access to psychiatrists. We believe that specialists as well as general practitioners may be assisted by a decision-support tool, which may reduce the waiting time and facilitate earlier medical assistance. The medical decision support tool is not intended to make therapeutic and diagnostic decisions, and it should never replace proper psychiatric examination or harm the crucial relationship and trust between a doctor and a patient. The effectiveness of such a support system depends profoundly on the underlying models of the measurement process. Thus, we present first a framework based on fuzzy-logic and semiotic approaches to model the complex and multi-dimensional concept of depression and its measurement methods. Second, we then use this framework to model two general types of measurements: (1) rating by an expert using widely accepted rating scales such as the Hamilton Rating Scale for Depression (HRSD) and (2) self-administered questionnaires filled out by patients such as the Beck Depression Inventory (BDI).

This paper is structured as follows. Section 2 surveys various classifications of depression and methods used to measure the severity of depression. Section 3 presents the definition of imprecision in medical data and defines three characteristics. Section 4 presents a fuzzy-semiotic framework for explicit representation of imprecision in medical data. Section 5 and 6 describe the implementation of the framework for the assessment of depression in the context of treatment evaluation and in the context of screening. Finally, the last section presents conclusions and suggests directions for future research.

2 Depression and its measurement process

Depression is a term used to cover a wide range of states, from feeling sad or helpless, through minor depression to major depression (MD). There are many approaches to the definitions, classifications, diagnostic criteria, and measurements of depression. These diverse approaches reflect the fact that depression has a complex etiology and presents itself with a variety of symptoms, which differ in different patients. In this paper, we focus on modeling the measurement process; thus we concentrate on a classification of symptoms, their frequency, and duration. However, the measurement process is determined by the definition of depression (its conceptualization). Thus, we cannot model the measurement process without specifying first the various conceptualization approaches.

2.1 Conceptualization of Depression

Several conceptual approaches to depression exist, and many authors view depression as a syndrome rather than a single diagnostic entity. In this paper, we focus on major characteristics: (1) depression can be defined by a set of presenting symptoms, which display specific severity, frequency, and duration, (2) depression can be viewed as a dimensional concept in which symptoms may be grouped or clustered into specific dimensions, and (3) depression may be conceptualized as a state or trait.

The symptomatic approach to depression identifies more than 10 symptoms, which have varied definitions and which

ISBN: 978-989-95079-6-8

are used in different ways by the diagnostic criteria. The symptoms are generally grouped into three classes: *affective* (crying, sadness, apathy), *cognitive* (thoughts of hopelessness, helplessness, suicide, worthlessness, guilt), and *somatic* (sleep disturbance, changes in energy level, changes in appetite, and elimination). Not all symptoms are present at the same time, and the severity of symptoms differs. Moreover, the symptoms may vary in their "directions." For example, two subtypes of depression are distinguished: depression with vegetative symptoms (e.g., appetite loss, weight loss, insomnia) and depression with reverse vegetative symptoms (e.g., appetite increase, weight gain, hypersomnia). The second subtype, according to the epidemiological studies, is characteristic of one-fourth to one-third or of all people with major depression, and it is more common among women.

2.2 Operationalization of Depression

In our discussion, we refer to two general diagnostic criteria: WHO's International Classification of Disease (ICD-10 in Great Britain) and the Diagnostic and Statistical Manual of Mental Diseases Fourth Edition (DSM-IV in North America). Both criteria are based on a symptomatic approach and have many similarities and differences [3]. They differ in the set of symptoms; however, both have eight symptoms in common: depressed mood, loss of interest, decrease in energy or increased fatigue, sleep disturbance, appetite disturbance, recurrent thoughts of death, inability to concentrate or indecisiveness, psychomotor agitation or retardation. The ICD-10 and DSM-IV criteria have a significant overlap in diagnosis, yet in some cases the patient may meet the diagnostic criteria in one system but not in the other. For example, the ICD-10 criteria do not take into account bereavement, while DSM-IV excludes a diagnosis of major depression if the symptoms may be linked with the bereavement process. Both diagnostic criteria have been modified several times and remain subjects of ongoing discussions. These modifications clearly indicate the difficulties in defining such a heterogeneous syndrome.

2.3 Assessment of Depression

Assessment of depression must be placed in a broader context of clinical decision making. Thus, for example, Nezu et al. [4] define the following questions to guide the selection and interpretation of the appropriate measurement process: (1) What are the goals of assessment? (2) Who is to be assessed? (3) What is the value of a given measure and who is the source of the information? The goals, for example, are screening, diagnosis and classification, clinical hypothesis testing, treatment planning, prediction of behavior, or outcome evaluation. The assessed groups are heterogeneous in terms of age, comorbidity, cultural background, and gender. The measures can be divided along the lines of *idiographic* and *nomothetic* philosophies of measurement. The idiographic approach assumes that each person is unique and thereby requires an individualized method of measurement. Thus, the measuring process requires an assessment by a clinician who follows the general structure of a rating scale, but who can vary the specific questions according to individual needs of the patient. The nomothetic approach has the goal of generalization, and the measurements strictly follow standardized questionnaires. Moreover, various measures are characterized by varied reliability, validity, and practical utility.

3 Imprecision

Imprecision is an intrinsic part of many medical concepts and their measurements. Concepts such as *quality of life*, *state of health,* and *depression* are difficult to define, measure, and quantify. Moreover, a certain level of imprecision is characteristic of all medical data. Imprecision is an inherent part of conceptualization, operationalization, and the measurements themselves. Moreover, imprecision is distinct from incompleteness (absence of value), inaccuracy (value is not close to the "true" value), inconsistency (dissimilar values from several sources), and uncertainty (probability or belief that the value is the "right" value. In our discussion, we focus on three characteristics:

1) Imprecision has two aspects: *qualitative* and *quantitative*. The qualitative imprecision is a result of a vagueness of the concept (e.g., quality of life, status of health) and the inability to precisely measure the concept (e.g., inherent imprecision of self-administered depression questionnaires). The quantitative imprecision is a result of a lack of precision in a measurement. We view these aspects of imprecision as pragmatic (vagueness of the concept), semantic (lack of precise measures), and syntactic (lack of precision in a measurement).

2) Imprecision is highly *contextual* and *interpretative*, i.e., a statement "I feel sad some of the time" may be sufficiently precise in a specific situation or a more precise statement such as "I feel sad 2-3 times a week for more than 2 hours without any particular reason" may be needed. Thus, imprecision is a quality of a specific value used in reference to a concept in a specific context. Often, imprecise values are sufficient, since the precision may be impossible to attain, impractical, too expensive, or unneeded.

3) Imprecision should be viewed in *terms of degrees.* Each concept, its representation, and its interpretation have a certain degree of imprecision or precision. Imprecision in a sense has continuous values. The values can be ordered in increasing or decreasing order of precision. For example, the severity of depressive mood can be described in terms of its frequency and duration with an increasing precision: "occasional sadness" "sadness 2-3 times a day" and "sadness 2-3 times a day more than for 2 hours."

4 A fuzzy-semiotic framework

In this section, we present a conceptual framework for modeling imprecision and its qualitative and quantitative aspects. Our framework is based on a complementary combination of the semiotic approach and the fuzzy-logic approach. The semiotic approach provides a model for *qualitative* imprecision, whereas the fuzzy-logic approach

ISBN: 978-989-95079-6-8

provides explicit representation for *quantitative* imprecision. Our framework addresses also the contextual and interpretative characteristics of imprecision.

4.1 Semiotic Approach

We based the framework on Peircean semiotics, which defines "sign" as any entity carrying some information and used in a communication process. Peirce divided semiotics into three categories [5]: syntax (the study of relations between signs), semantics (the study of relations between signs and the referred objects), and pragmatics (the study of relations between the signs and the agents who use the signs to refer to objects in the world). This triadic distinction is represented by Peirce's semiotic triangle: the *object*, *representamen* (the form which the sign takes), and *interpretant*. Fig. 1 shows the semiotic representation for the assessment of depression. In this model, the "interpretant" refers to the purpose of the assessment (e.g., screening, diagnosis, treatment evaluation), the agents (e.g., patients, clinicians), the perspectives (e.g., health care costs, accessibility, ethics), the biases (e.g., specific subgroups of agents), and the views (e.g., variations in the diagnostic criteria).

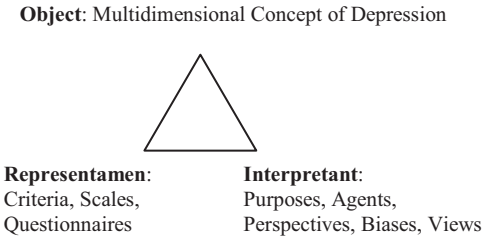


Figure 1: Semiotic triangle.

The assessment involves three aspects: conceptualization (what to measure), operationalization (how to measure), and utilization (how to use the measure). These three aspects constitute a triplet: $<O,M,U>$, where O represents a set of objects, M a set of measures, and U a set of utilization parameters (interpretants). For example, the representation of depression has a set of objects such as depressive episode, cognitive symptoms, affective symptoms, somatic symptoms, a set of diagnostic criteria and assessment instruments, and a set of clinical guidelines for the assessment process in a particular context.

4.2 Fuzzy- Logic Approach

We based the framework on a fuzzy-logic approach (1) to assess the severity of depressive episode as "absent," "mild," "moderate," and "severe" and (2) to assess the severity of particular dimensions of depression. Since the assessment process depends on the clinical context, the accepted definition of depression, and the specific diagnostic criteria, we apply the fuzzy-logic framework to the treatment evaluation and screening.

5 Modeling the assessment of depression in the context of treatment evaluation

In this section, we describe the treatment evaluation process, which typically is used in psychiatric clinics to assess clinical situations and to evaluate the effectiveness of pharmaceutical treatment. We have applied the fuzzy-semiotic framework to model the assessment of depression in the context of treatment evaluation. The evaluation protocol is based on our earlier work on a fuzzy-logic based system to support a depressive episode therapy [6]. In the following subsections, we discuss diagnostic criteria, an assessment instrument, a measure for treatment effectiveness, and an implementation of the fuzzy-semiotic framework.

5.1 Diagnostic Criteria

The treatment evaluation requires at least two consultations: pre- and post-treatment. During the first consultation, the clinician evaluates the patients according to diagnostic criteria accepted by the clinic, such as DSM-IV or ICD-10. In this specific application, the clinician uses the Research Diagnostic Criteria (RDC) for ICD-10 to assess the severity of a depressive episode as *mild*, *moderate*, or *severe*, and to classify it as an episode with or without psychotic symptoms and with or without somatic symptoms. During the second consultation, the clinician repeats the assessment, compares the results with the results from the first consultation, and evaluates the clinical situation as a *recovery*, *partial improvement*, *lack of improvement*, and *deterioration*. The RDC for ICD-10 have 16 items, which are rated by the clinician using the Hamilton Rating Scale for Depression (HRSD) and eight additional questions for the items not covered by HRSD.

5.2 Assessment Instrument

The Hamilton Rating Scale for Depression (HRDS) is an assessment instrument, which has been the most frequently used clinical rating scale since its inception in 1960 [7]. The HRDS is completed by a clinician, and it used to indicate the severity of depression in patients already diagnosed with a depressive disorder. The HRDS has 21 items; 17 items are usually used for scoring. The items are measured on a three–point scale (0, 1, 2) or a five-point scale (0, 1, 2, 3, 4). The items are based on symptoms. The items on the three-point scale are quantified as 0 = "symptom absent," 1 = "slight or doubtful," and 2 = "clearly present." The items on the five-point scale are quantified in terms of increasing intensity: 0 = "symptom absent," 1 = "doubtful or trivial," 2 = "mild," 3 = "moderate," and 4 = "severe." For example, the symptom *depressed mood* is quantified as 0 = "absent," 1 = "feeling of sadness, gloomy attitude, pessimism about future," 2 = "occasional weeping," 3 = "frequent weeping," 4 = "extreme symptoms." Typically, the scores from 17 HRDS's items are added together, and the final score ranges from 0 to 52 points. Originally, Hamilton did not specify cutoff points; however, generally the scores lower than 7 indicate an absence of depression, scores 7 to 17 indicate mild depression, scores 18 to 24 indicate moderate depression, and scores of 25 and above indicate severe depression. Generally, the HRDS has high reliability and validity. On the other hand, the literature on the HRSD includes many papers criticizing the scale on a number of grounds. One of the important issues for our explicit model of imprecision is a critical assumption about the type of a measurement. The problem is related to the quantification of the concept: is the

ISBN: 978-989-95079-6-8

HRDS an *ordinal measurement* scale or is it an *interval measurement* scale? Thus, although from a theoretical perspective, the HRDS is an ordinal scale, from the practical perspective, the HRDS is perceived as an interval measurement, which means that one unit on the scale represents the same magnitude of change across the entire scale. For example, the symptom *depressed mood* rated from 0 to 4 should have equal distances between the four points. The distance of two units between "absence of depressed moods" (rated as 0) and "occasional weeping" (rated as 2) should be the same as the distance between "occasional weeping" (rated as 2) and "extreme symptoms, when the patient reports only sad feelings in verbal and non-verbal communication" (rated as 4).

We have observed that (1) the HRDS represents an ordinal scale, and the assumption of the equity of distances between units introduces a large measurement imprecision, and (2) the range of values on three-point scale and five-point scale is not sufficient for scaling small changes in symptoms during the treatment. To address these two problems, we have introduced an 11-point rating scale – a range of values from a discrete set: {0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1}. The value 0 means "normal state" (not precisely defined), and the value 1 means extreme pathology (also not specified), the medium values are based on the HRDS range of scores 0–4 (1 = 0.2, 2 = 0.4, 3 = 0.6, and 4 = 0.8). Each item has the same rating range. The 16 items on the RDC for ICD-10 and an example of values for patient p_1 are presented in Fig. 2.

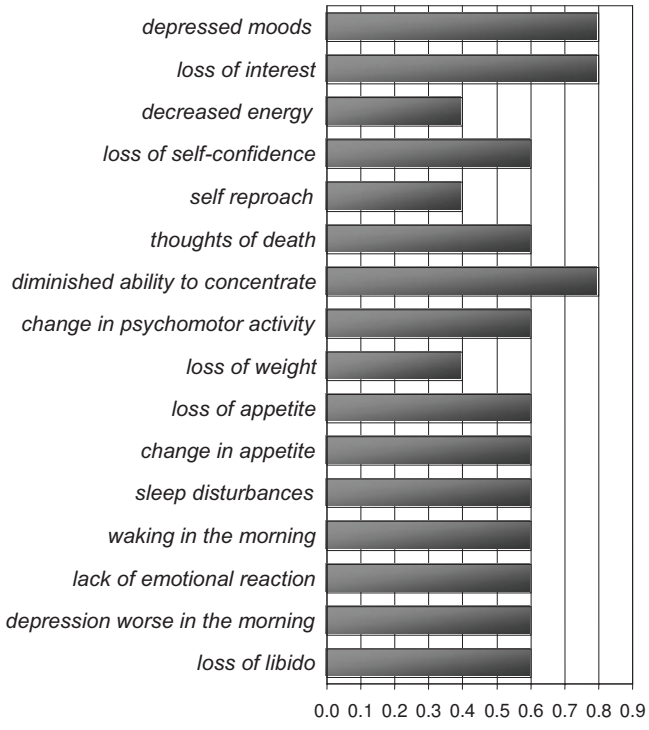


Figure 2: Assessment of depressive episode for patient p_1 using the RDC for ICD-10.

5.3 Measure of the Treatment Effectiveness

The treatment evaluation is based on the comparison between the pre-treatment and post-treatment assessment outcomes. In the assessment process, a clinician (a rater) evaluates the intensity of depression for each of the 16 items, and then calculates the relative intensity of depression in percent. The *Relative Depression Intensity*, *RDI,* is calculated for a specific patient, p_1, rated by a specific rater, r_1, at given point in time, t_1, using (1). The RDC criteria for ICD-10 are represented as C_1,..., C_{16} (n = 16). The maximum value for each C_i is 1.

$$\mathrm{RDI}\ (\mathrm{p}_1, \mathrm{r}_1, \mathrm{t}_1) = \frac{\sum_{i=1}^{n} C_i(p_1, r_1, t_1)}{\sum_{i=1}^{n} \mathrm{C}_\mathrm{i}^{\max}} 100\% \qquad (1)$$

The RDI values range from 0% to 100% and are interpreted using linguistic descriptors as follows: "absent" corresponds to interval <0, 10), "mild" corresponds to <10, 40), "moderate" corresponds to <40, 70), and "severe" corresponds to <70, 100>.

The efficiency of therapy is measured using the *Therapy Effectiveness Index*, *TEI*. The TEI is based on the RDI assessed during the initial consultation at time t_1, and the RDI assessed during a later consultation at time t_2, where $t_1 < t_2$, using (2).

$$\mathrm{TEI} = \frac{RDI(p_1, r_1, t_1) - RDI(p_1, r_1, t_2)}{RDI(p_1, r_1, t_1)} 100\% \qquad (2)$$

The values of the TEI relate to four clinical situations: "no improvement," TEI = 0; "recovery," TEI = 1, "partial improvement," 0 < TEI < 1; and "deterioration," TEI < 0.

5.4 Fuzzy-Semiotic Model for the Treatment Evaluation

From the semiotic perspective, the assessment process involves a representation and an interpretation of the assessment outcomes. The outcomes of the assessment are interpreted with reference to specific patients, raters (clinicians), and specific timing. Thus, we conceptually model the assessment process as a tuple shown in (3).

$$\mathrm{Assessment} = < P, R, \{(O,V)\}_{i=1}^{n}, T > \qquad (3)$$

Where P represents a finite set of references to the patients, R represents a finite set of references to the raters, O represents a finite set of outcomes, V represents a finite set of values for the outcomes, and T represents a finite set of time points. The outcomes, O, consist of the criteria, C, and the measurements, M. The measurements can be *quantitative* and *qualitative*.

The quantitative measurements are absolute scores (calculated as the sum of all items from a diagnostic scale) and relative scores or indexes (calculated as a ratio between the patient's score and the possible maximum score value). For example, the relative intensity of depression, the RDI (shown earlier in Fig. 2) is 60 %.

The qualitative measurements are linguistic descriptions such as "symptoms do not meet criteria for any depressive episode," "mild," "moderate," and "severe." Additionally, the linguistic descriptions may include a phrase concerning somatic and psychotic symptoms. Thus, a qualitative

ISBN: 978-989-95079-6-8

outcome may be specified, for example as "severe with exclusion of psychotic symptoms."

Using our framework, an assessment of patient, p_1, by a clinician, r_1, at time, t_1, with outcomes o_1 and o_2 is represented as (p_1, r_1, (o_1, 60) (o_2, "moderate"), t_1).

We present the fuzzy-logic aspect of the framework based on the standardized rating guidelines for the administration of the HRDS, GRID-HAMD-17 [8]. The GRID-HAMD-17 overcomes several shortfalls in the original HRDS, particularly, the high level of imprecision in measures of frequency and intensity. The GRID-HAMD-17 provides specific instructions for the evaluation of the *frequency* and *intensity* for the 12 items and *intensity* only for 5 items (frequency is not applicable for these items). The frequency is represented by four linguistic terms: "absent or clinically insignificant," "occasional," "much of the time," and "almost all of the time." The GRID-HAMD-17 guidelines specify the mapping between the linguistic terms and the frequency of symptoms measured in days/week. The mapping is defined as follows: "absent" = not occurring, "occasional" = less than 3 day/week, "much of the time" = 3–5 days/week, "almost all the time" = 6–7 days/week. The same definition of frequency is used for all applicable items. The intensity is represented by five linguistic terms: "absent," "mild," "moderate," "severe," and "very severe." The terms have specific qualitative mappings for each item.

We present the fuzzy-logic approach, using an example of the first item on the GRID-HAMD-17, "depressed moods." We have constructed two linguistic variables: *Frequency* and *Intensity of Depressed Moods*. The variable *Frequency* has four terms: *absent*, *occasional*, *much of the time*, *almost all the time*. The membership functions, MFs, for the symptom frequency are shown in Fig. 3. They have been created based on the frequency measured by days per week.

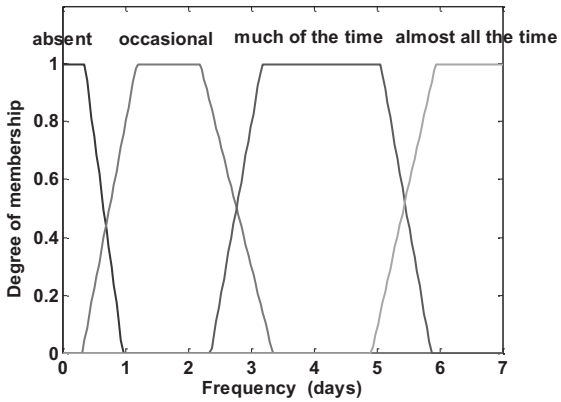


Figure 3: MFs for the frequency of a symptom.

The MFs for the intensity of a depressed mood are shown in Fig. 4. The MFs are based on a continuous scale from 0 to 1, which corresponds to the clinician's rates on a discrete scale from 0 to 1 with an increment of 0.1. The intensity of the symptom is rated by the clinician based on the GRID-HAMD-17 specification. For example, the depressed mood with a "severe" intensity is described as "intense sadness; hopelessness about most aspects of life, feeling of complete helplessness or worthlessness."

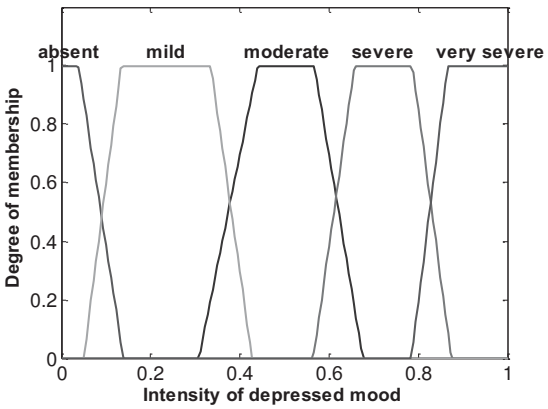


Figure 4: MFs for the depressed mood intensity.

6 Modeling the assessment of depression in the context of screening

In this section, we describe the screening process using the Beck Depression Inventory (BDI). The BDI was originally developed by Beck in 1961 [9], and since then, it has been used to measure the intensity of depression in psychiatric patients, to screen for depression in family practice, and to screen for depression in research studies. BDI has 21 items evaluating 21 symptoms of depression. Each item is rated on a four-point intensity scale (0 to 3). The currently used BDI version refers to the last 7 days and is typically self-administered [10]. The scores are added together and given a total score between 0 and 63. The common guidelines for the interpretation of scores specify that scores less than 10 indicate "absent" depression, 10 to 18 indicate "mild to moderate", 19 to 29 indicate "moderate to severe," and scores of 30 and above indicate "severe" depression.

Using our framework, an assessment of patient, p_1, by self-reporting, $r_1 = p_1$ at time, t_1, with outcomes o_1 and o_2 is represented as (p_1, r_1, (o_1, 35) (o_2, "severe"), t_1). Using the fuzzy-logic approach, we have constructed a linguistic variable *Severity of Depression* with four terms: absent, mild, moderate, and severe. The MFs for the severity of depression are shown in Fig. 5.

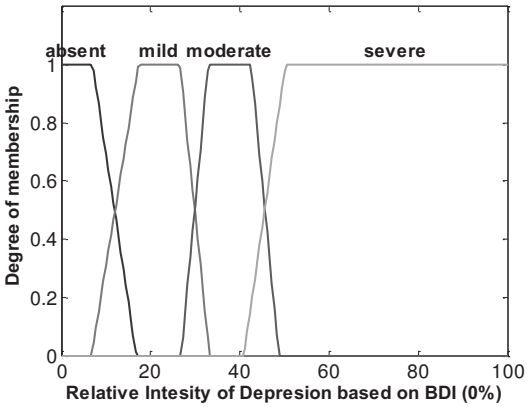


Figure 5: MFs for the severity of depression based on BDI.

We have observed that the assessment of depression is highly contextual and depends on (1) the conceptual basis of the measures, (2) the use of specific guidelines or approaches to the measuring instruments, and (3) the interpretation of the

ISBN: 978-989-95079-6-8

outcomes. To demonstrate the inherent imprecision of interpretation, we have compared the guidelines for the RDC for ICD-10 (described in Section 5), HAMD-17, and BDI. We have converted the HAMD-17 scale and BDI to indexes, by dividing the HAMD-17 score by its maximum score of 52, and dividing the BDI score by its maximum score of 63. Table 1 shows the mappings between qualitative descriptors and corresponding ranges of scores for the RDC-based evaluation, HAMD-17, and BDI. We have assumed that the scales are interval measurements, and they measure intensity of similar conceptual constructs of depression. Interestingly, we have noticed that the patient p_1, scoring 60% = "moderate" on RDI, will be classified as "severe" on HAMD-17 and BDI.

Table 1: Depression severity for three assessment measures.

Severity	RDC ICD-10(%)	Indexed HAMD-17 %	Indexed BDI (%)
Absent	< 10	< 14	< 16
Mild	10 -39.9	14 – 33.9	16 – 29.9
Moderate	40 -69.9	34 – 46.9	30 – 46.9
Severe	> 69.99	> 46.9	> 46.9

7 Conclusions

In this paper, we have examined three essential characteristics of imprecision: (1) qualitative and quantitative aspects, (2) contextual and interpretative nature, and (3) graduality of imprecision. We have demonstrated that imprecision is an intrinsic part of the assessment of such a complex syndrome as depression. Furthermore, we have presented a fuzzy-semiotic framework for the explicit representation of qualitative, quantitative, contextual, and interpretative aspects of imprecision. We have used a semiotic approach to represent the concept of depression, its symptomatic representations, and the clinical utilization of the measures. Furthermore, we have applied a fuzzy-logic approach to explicitly represent the imprecision of various measures.

We are planning to expand and further formalize the proposed framework and to build a comprehensive data model for the medical concept of "depression" and its assessment in treatment evaluation and screening. We will apply this model in a clinical decision support system for the diagnosis and treatment of depression, as well as in a support system for the treatment of sleep disorders.

Acknowledgment

The work presented in this paper is partially supported by NSERC Grant # 327545-06.

References

[1] R.C. Kessler, Epidemiology of depression [In] I.H. Gotlib and C.L. Hammen (Eds) *Handbook of depression*, New York: Guilford Press, 2002.

[2] World Health Organization, Mental Health: Depression http://www.who.int/mental_health/management/depression/definition/en, accessed on January 19, 2009.

[3] A.M. Gruenberg, R.D. Goldstein, and H.A. Pincus, Classification of Depression: Research and Diagnostic Criteria: DSM-IV and ICD-10 [In] *Biology of Depression. From Novel Insights to Therapeutic Strategies*. J. Licincio, Ma-Li Wong (Eds), WIILEY-VCH, Weinheim, 2005.

[4] A.M. Nezu, C.M. Nezu, K.S. McClure et al., Assessment of Depression [In] I.H. Gotlib and C.L. Hammen (Eds) *Handbook of depression*, New York: Guilford Press, 2002.

[5] T.A. Sebeok, *Signs: An introduction to semiotics,* University of Toronto Press, 1999.

[6] K. Kielan, The Salomon advisory system supports a depressive episode therapy. *Polish Journal of Pathology*, 54, 3, 215-218, 2003.

[7] M. Hamilton, A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry*, 23 (1), 56-61, 1960.

[8] A. Kalali, J.B.W. Williams, K.A. Kobak, et al., The new GRID HAM-D: pilot testing and international field trials. *International Journal of Neuropsychopharmacology*, 5, S147–S148, 2002.

[9] A.T. Beck, C.H Ward, M. Mendelson, et al., An Inventory for measuring depression, *Archives of General Psychiatry*, 4, 561-571, 1961.

[10] I. McDowell and C. Newell, *Measuring Health: A Guide to Rating Scales and Questionnaires*, New York: Oxford University Press, 1996.

An Axiomatical Approach to *T*–Multi-indistinguishabilities

Javier Martín Gaspar Mayor

Department of Mathematics and Computer Sciences
University of the Balearic Islands
Palma (Mallorca), Spain
Email: javier.martin@uib.es, gmayor@uib.es

Abstract— *Axiomatic definitions for T–multi-indistinguishabilities and multidistances are introduced, with the objective to measure the degree of similarity not only between two elements but also between all of the elements of a finite list. The consistency of these definitions with respect to classic T–indistinguishabilities and distances allows us to translate the close relationship existing between these two concepts.*

Keywords— additive generator, distance, multidistance, pseudo-distance, T–indistinguishability, T–multi-indistinguishability, t-norm.

1 Introduction

Fuzzy equivalence relations, i.e. reflexive, symmetric and Min–transitive relations, were introduced in 1971 by L.A. Zadeh [1] with the name of similarities. Later, the Minimum used in their definition was replaced by a t–norm T [2] and they were called T–indistinguishabilities. Since then, fuzzy equivalences have received many other denominations, depending on the author, the context or sometimes on the particular t-norm used to define them: likeness relations, fuzzy equalities, T–equivalence relations (the one preferred in [3]), proximity relations, etc. Here we will use the term T–indistinguishability.

T–indistinguishabilities are then the fuzzy concept corresponding to classic equivalence relation, and they have become from the first beginning as a useful tool in situations with a certain degree of uncertainty where the transitivity of the equivalence relations is too strong. They have been widely used in many applications: Fuzzy Control, Cluster Analysis, Pattern Recognition, Approximate Reasoning, etc. They are also an important branch of Fuzzy Sets Theory and Fuzzy Logic.

For a given set, T–indistinguishabilities measure the degree of similarity between two of its elements. In this paper we propose an extension of this concept in order to be used to measure the similarity between all of the elements of a finite list. This extension to multiple arguments will be called T–multi-indistinguishabilitiy, and we will deal also with its counterpart in the field of metric spaces: multidistance.

The paper is organized as follows. First, in Section 2, well known definitions and properties of t–norms are recalled. In Sections 3 and 4 axiomatical definitions for T–multi-indistinguishabilities and related pseudo-multidistances and multidistances are proposed, respectively. After, the close relationship existing between T–indistinguishabilities and pseudodistances, by means both of an additive generator of the t–conorm and strong negations, is recalled and then translated to our multidimensional definition in Section 5. Finally, an example of how to construct T–multi-indistinguishabilities from multidistances by using this relationship is detailed in Section 6.

2 Definitions and basic properties of triangular norms

Triangular norms are involved in the definition of T–indistinguishabilities, and so a quick overview on definitions, examples and properties is convenient (see [3] for a more complete study).

Definition 1 *A* triangular norm*, or a* t–norm *in short, is a binary operation $T\colon [0,1]^2 \to [0,1]$ satisfying the following properties, for all $x, y, z \in [0,1]$:*

(i) commutativity: $T(x,y) = T(y,x)$,

(ii) associativity: $T(x, T(y,z)) = T(T(x,y), z)$,

(iii) monotonicity: $y \leqslant z \Rightarrow T(x,y) \leqslant T(x,z)$,

(iv) neutral element: $T(x,1) = x$.

A t-conorm *is a binary operation satisfying (i), (ii),(iii), with* 0 *as a neutral element.*

Definition 2 *A triangular norm is called* Archimedean *if for each $(x,y) \in (0,1)^2$ there exists an $n \in \mathbb{N}$ such that $T(\overbrace{x, \ldots, x}^{n}) < y$. Analogous definition for Archimedean t–conorm.*

Basic t-norms are:

- minimum: $T_M(x,y) = \min\{x,y\}$,
- product: $T_P(x,y) = xy$,
- Łukasiewicz t-norm:

$$T_L(x,y) = \max\{x+y-1, 0\},$$

- drastic t-norm:

$$T_D(x,y) = \begin{cases} x & \text{if } y = 1, \\ y & \text{if } x = 1, \\ 0 & \text{otherwise.} \end{cases}$$

In this work we will deal with continuous Arquimedean t–norms, whose characterization is given in the following result.

ISBN: 978-989-95079-6-8

Proposition 1 *A t–norm T is a continuous Archimedean t–norm if and only if it has a continuous additive generator, i.e., there exists a continuous, strictly decreasing function $t\colon [0,1] \to [0,\infty]$ with $t(1) = 0$, which is uniquely determined up to a positive multiplicative constant, such that*

$$T(x,y) = t^{(-1)}(t(x) + t(y)),$$

for all $(x,y) \in [0,1]^2$.

Remember that the *pseudo-inverse* $t^{(-1)}$ of an additive generator t of T is defined as:

$$t^{(-1)}(x) = \sup\{z \in [0,1] : t(z) > x\}.$$

3 *T*–indistinguishabilities and *T*–multi-indistinguishabilities

Let $\mathbb{X}$ be a non empty set. T–indistinguishabilities are defined as follows (see for example [3]).

Definition 3 *Let T be a t–norm. We say that a function $E\colon \mathbb{X}^2 \to [0,1]$ is a* T–indistinguishability *on $\mathbb{X}$ if and only if for all $x_1, x_2, y \in \mathbb{X}$:*

(Ti1) $E(x_1,x_2) = 1$ *when* $x_1 = x_2$,

(Ti2) $E(x_1,x_2) = E(x_2,x_1)$,

(Ti3) $E(x_1,x_2) \geqslant T(E(x_1,y),E(x_2,y))$.

And finally, a T–indistinguishability E separates points *if, in addition,*

(Ti4) $E(x_1,x_2) = 1 \Rightarrow x_1 = x_2$.

Special cases are the *similarity relations* of Zadeh , the *equality relations* of Menger and the *likeness relations* of Ruspini, corresponding to the minimum, product and Łukasiewicz t-norms, respectively.

The elements (lists) of the cartesian product $\mathbb{X}^n$ will be denoted by $(x_1,\ldots,x_n)$, or $\mathbf{x}$.

A *partition* of $(x_1,\ldots,x_n)$ is a set of lists $\{\mathbf{x_i} = (x_{i_1},\ldots,x_{i_{n_i}}) \in \mathbb{X}^{n_i}; i = 1,\ldots,k\}$ such that the sets of subindexes $\{\{i_1,\ldots,i_{n_i}\}, i = 1,\ldots,k\}$ constitute a partition of $\{1,\ldots,n\}$.

Example 1 *We consider the list $\mathbf{x} = (1,0,0,3,1,1)$ and the partition $\{\{1,4,5\},\{2,3,6\}\}$ of $\{1,2,3,4,5,6\}$. The corresponding partition of the list $\mathbf{x}$ is $\{\mathbf{x_1} = (1,3,1), \mathbf{x_2} = (0,0,1)\}$.*

We introduce now our proposal for an axiomatic definition of T–multi-indistinguishability.

Definition 4 *Let T be a t–norm. We say that a function $\mathcal{E}\colon \bigcup_{n\geqslant 1}\mathbb{X}^n \to [0,1]$ is a* T–multi-indistinguishability *on $\mathbb{X}$ if and only if the following properties hold for all n, for all $x_1,\ldots,x_n,y \in \mathbb{X}$.*

(Tmi1) $\mathcal{E}(x_1,\ldots,x_n) = 1$ *when* $x_1 = \ldots = x_n$,

(Tmi2) $\mathcal{E}(x_1,\ldots,x_n) = \mathcal{E}(x_{\pi(1)},\ldots,x_{\pi(n)})$ *for any permutation π of* $1,\ldots,n$,

(Tmi3) $\mathcal{E}(x_1,\ldots,x_n) \geqslant T(\mathcal{E}(x_1,y),\ldots,\mathcal{E}(x_n,y))$.

We say that $\mathcal{E}$ is a strong *T–multi-indistinguishability if it fulfills (Tmi1), (Tmi2) and:*

(Tmi3') $\mathcal{E}(x_1,\ldots,x_n) \geqslant T(\mathcal{E}(\mathbf{x_1},\mathbf{y}),\ldots,\mathcal{E}(\mathbf{x_k},\mathbf{y}))$ *for any partition $\{\mathbf{x_1},\ldots,\mathbf{x_k}\}$ of $\boldsymbol{x} = (x_1,\ldots,x_n)$, for all $\mathbf{y} \in \bigcup_{n\geqslant 1}\mathbb{X}^n$.*

A T–multi-indistinguishability $\mathcal{E}$ separates points *if, in addition,*

(Tmi4) $\mathcal{E}(x_1,\ldots,x_n) = 1 \Rightarrow x_i = x_j$ *for all* $i,j = 1,\ldots n$.

To conclude this definition, we will say that $\mathcal{E}\colon \bigcup_{n=1}^{k}\mathbb{X}^n \to [0,1]$ is a k–T–multi-indistinguishability *if it satisfies the corresponding properties in $\bigcup_{n=1}^{k}\mathbb{X}^n$. In this sense, classic T–indistinguishabilities are 2–T–multi-indistinguishabilities.*

Remark 1 *We denote by $\mathcal{E}(\mathbf{x},\mathbf{y})$ the T–multi-indistinguishability between the elements of the joint list $\mathbf{x},\mathbf{y}$:*

$$\mathcal{E}(\mathbf{x},\mathbf{y}) = \mathcal{E}(x_1,\ldots,x_n,y_1,\ldots,y_m).$$

However, this does not means that $\mathcal{E}$ is a T–indistinguishability between lists of $\bigcup_{n\geqslant 1}\mathbb{X}^n$ because (Ti1) is not fulfilled. But T–indistinguishabilities on this set of lists can be constructed from $\mathcal{E}$ and a T–indistinguishability E on $[0,1]$ as follows:

$$E_{\mathcal{E}}(\mathbf{x},\mathbf{y}) = E(\mathcal{E}(\mathbf{x}),\mathcal{E}(\mathbf{y})).$$

In this case, $E_{\mathcal{E}}$ does not separate points even if E and $\mathcal{E}$ do, because two different lists $\mathbf{x},\mathbf{y}$ with the same degree of indistinguishability verify $E_{\mathcal{E}}(\mathbf{x},\mathbf{y}) = 1$.

4 Distances and multidistances

The definition of classic distances and metric spaces [4] goes back to 1906. We recall it here.

Definition 5 *A function $d\colon \mathbb{X}^2 \to [0,\infty]$ is a* pseudo–distance *on a set $\mathbb{X}$ if for all $x_1,x_2,y \in \mathbb{X}$:*

(d1) $d(x_1,x_2) = 0$ *when* $x_1 = x_2$,

(d2) $d(x_1,x_2) = d(x_2,x_1)$,

(d3) $d(x_1,x_2) \leqslant d(x_1,y) + d(x_2,y)$.

If the following property also holds:

(d4) $d(x_1,x_2) = 0 \Rightarrow x_1 = x_2$,

then we say that d is a distance.

The related properties we propose [5] to define pseudo–multidistances and multidistances are the following.

Definition 6 *A function $D\colon \bigcup_{n\geqslant 1}\mathbb{X}^n \to [0,\infty]$ is a* pseudo-multidistance *on a set $\mathbb{X}$ if it fulfills the following properties for all n, for all $x_1,\ldots,x_n,y \in \mathbb{X}$:*

(md1) $D(x_1,\ldots,x_n) = 0$ *when* $x_1 = \ldots = x_n$,

(md2) $D(x_1,\ldots,x_n) = D(x_{\pi(1)},\ldots,x_{\pi(n)})$ *for any permutation π of* $1,\ldots,n$,

ISBN: 978-989-95079-6-8

(md3) $D(x_1,\dots,x_n) \leqslant D(x_1,y)+\dots+D(x_n,y)$.

We say that D is a strong *pseudo–multidistance if it fulfills (md1),(md2) and:*

(md3') $D(x_1,\dots,x_n) \leqslant D(\mathbf{x_1},\mathbf{y})+\dots+D(\mathbf{x_k},\mathbf{y})$ *for any partition* $\{\mathbf{x_1},\dots,\mathbf{x_k}\}$ *of* $\mathbf{x}=(x_1,\dots,x_n)$*, for all* $\mathbf{y}\in\bigcup_{n\geqslant 1}\mathbb{X}^n$.

If in addition the following property holds:

(md4) $D(x_1,\dots,x_n)=0 \Rightarrow x_i=x_j$ *for all* $i,j=1,\dots n$,

we will say that D is a multidistance.

Finally, $D\colon \bigcup_{n=1}^{k}\mathbb{X}^n \to [0,\infty]$ *is a* k–pseudo-multidistance*, or* k–multidistance*, if it satisfies the corresponding properties in* $\bigcup_{n=1}^{k}\mathbb{X}^n$*. In this sense, classic distances are 2–distances.*

Remark 2 *Again (see remark 1 after Definition 4), expressions like $D(\mathbf{x},\mathbf{y})$ in condition (Tmi3') have to be interpreted as:*

$$D(\mathbf{x},\mathbf{y})=D(x_1,\dots,x_n,y_1,\dots,y_m),$$

but D is not an ordinary pseudodistance on $\bigcup_{n\geqslant 1}\mathbb{X}^n$ because (md1) does not hold. However, pseudodistances d_D in this set can be constructed by means of D and a pseudodistance d on $\mathbb{R}$ in this way:

$$d_D(\mathbf{x},\mathbf{y})=d(D(\mathbf{x}),D(\mathbf{y})).$$

This distance function d_D does not fulfill (md4): different lists $\mathbf{x},\mathbf{y}$ such that $D(\mathbf{x})=D(\mathbf{y})$ verify $d_D(\mathbf{x},\mathbf{y})=0$, and then it is not a multidistance.

Example 2 *The drastic distance* $d\colon \mathbb{X}^2\to\{0,1\}$*, given by:*

$$d(x,y)=\begin{cases}1 & \text{if } x\neq y,\\ 0 & \text{if } x=y,\end{cases}$$

can be extended in several ways. For example, with the multidistances $D_1,D_2,D_3\colon \bigcup_{n\geqslant 1}\mathbb{X}^n\to[0,\infty]$ *defined as follows:*

- $D_1(x_1,\dots,x_n)=\begin{cases}0 & \text{if } x_i=x_j\ \forall i,j,\\ 1 & \text{otherwise},\end{cases}$
- $D_2(x_1,\dots,x_n)=|\{x_1,\dots,x_n\}|-1$,
- $D_3(x_1,\dots,x_n)=n-m$*, where m is the number of times that appears the most repeated element of the list.*

It can be proved that D_1,D_2 are strong multidistances. But D_3 is not strong: for instance, if we take the partition $\{(0,0),(1,1)\}$ of $(0,0,1,1)$,

$$D_3(0,0,1,1)=2\not\leqslant D_3(0,0,0)+D_3(1,1,0)=1.$$

Example 3 *For any metric space $(\mathbb{X},d)$, the function D_M defined by:*

$$D_M(x_1,\dots,x_n)=\max\{d(x_i,x_j);\ i<j\},$$

is a strong multidistance on $\mathbb{X}$.

Another interesting examples and some properties of multidistances can be found in [5].

There exist another definitions for multidistances, or multimetrics. For example the one in [6], in the field of Probability and Statistics, which is more general because it allows to measure the distance between functions. Reduced to lists, multidistances there correspond with strong multidistances here. On the other hand, symmetry is not an axiom in [6], and hence it can not be avoided, preventing the definition of *asymmetric* multidistances.

5 *T*–multi-indistinguishabilities and pseudo multidistances

One way to construct T–indistinguishabilities is to start from pseudodistances, because there is a close relationship between these two concepts. Here we recall two important conexions that relate them [7] and, in both cases, we do the translation to the multidimensional case.

5.1 Relationship via an additive generator of the t–norm

The first relationship between T–indistinguishabilities and pseudo-distance is established by means of an additive generator of the t–norm T.

Proposition 2 *Let T be a continuous Archimedean t–norm and t an additive generator of T.*

(i) If $d\colon \mathbb{X}^2\to[0,\infty]$ *is a pseudo-distance on a set $\mathbb{X}$, then the function* $E_d\colon \mathbb{X}^2\to[0,1]$ *defined by* $E_d=t^{(-1)}\circ d$ *is a T–indistinguishability on $\mathbb{X}$. Furthermore, E_d separates points if and only if d is a distance.*

(ii) If $E\colon \mathbb{X}^2\to[0,1]$ *is a T–indistinguishability on $\mathbb{X}$, then the function* $d_E\colon \mathbb{X}^2\to[0,\infty]$ *given by* $d_E=t\circ E$ *is a pseudo-distance; d_E is a distance if and only if E separates points.*

The literal translation of this result to our multidimensional proposal is as follows.

Proposition 3 *Let T be a continuous Archimedean t–norm with continuous additive generator t.*

(i) If $D\colon \bigcup_{n\geqslant 1}\mathbb{X}^n\to[0,\infty]$ *is a pseudo-multidistance on a set $\mathbb{X}$, then the function* $\mathcal{E}_D\colon \bigcup_{n\geqslant 1}\mathbb{X}^n\to[0,1]$ *defined by* $\mathcal{E}_D=t^{(-1)}\circ D$ *is a T–multi-indistinguishability on $\mathbb{X}$. Furthermore, $\mathcal{E}_D$ separates points if and only if D is a multidistance.*

(ii) If $\mathcal{E}\colon \bigcup_{n\geqslant 1}\mathbb{X}^n\to[0,1]$ *is a T–multi-indistinguishability on $\mathbb{X}$, then the function* $D_{\mathcal{E}}\colon \bigcup_{n\geqslant 1}\mathbb{X}^n\to[0,\infty]$ *given by* $D_{\mathcal{E}}=t\circ\mathcal{E}$ *is a pseudo-multidistance; $D_{\mathcal{E}}$ is a multidistance if and only if $\mathcal{E}$ separates points.*

Proof

For (i), condition (Tmi1) is fulfilled:

$$\mathcal{E}_D(x,\dots,x)=t^{(-1)}(D(x,\dots,x))=t^{(-1)}(0)=1,$$

ISBN: 978-989-95079-6-8

(Tmi2) is immediate, due to the symmetry of D, and the inequality (Tmi3) also holds:

$$\begin{aligned}\mathcal{E}_D(x_1,\ldots,x_n) &= t^{(-1)}(D(x_1,\ldots,x_n))\\ &\geqslant t^{(-1)}(\textstyle\sum_{i=1}^n D(x_i,y))\\ &= t^{(-1)}(\textstyle\sum_{i=1}^n t(\mathcal{E}_D(x_i,y)))\\ &= T(\mathcal{E}_D(x_1,y),\ldots,\mathcal{E}_D(x_n,y)).\end{aligned}$$

Finally,

$$\begin{aligned}\mathcal{E}_D(x_1,\ldots,x_n)=1 &\Leftrightarrow t^{(-1)}(D(x_1,\ldots,x_n))=1\\ &\Leftrightarrow D(x_1,\ldots,x_n)=0,\end{aligned}$$

which suffices to verify that conditions (Tmi4) and (md4) are equivalent.

For (ii), the reasoning is the same. From:

$$M_{\mathcal{E}}(x,\ldots,x)=t(\mathcal{E}(x,\ldots,x))=t(1)=0,$$

we have (md1). The symmetry of $\mathcal{E}$ ensures (md2). Condition (md3) is also satisfied:

$$\begin{aligned}M_{\mathcal{E}}(x_1,\ldots,x_n) &= t(\mathcal{E}(x_1,\ldots,x_n))\\ &\leqslant t(T(\mathcal{E}(x_1,y),\ldots,\mathcal{E}(x_n,y))\\ &= \textstyle\sum_{i=1}^n t(\mathcal{E}(x_i,y))\\ &= \textstyle\sum_{i=1}^n M_{\mathcal{E}}(x_i,y),\end{aligned}$$

and (Tmi4) and (md4) are equivalent on account of this:

$$\begin{aligned}D_{\mathcal{E}}(x_1,\ldots,x_n)=0 &\Leftrightarrow t(\mathcal{E}(x_1,\ldots,x_n))=0\\ &\Leftrightarrow \mathcal{E}(x_1,\ldots,x_n)=1.\end{aligned}$$

■

Note that we have used this property of the additive generator: for all n,

$$T(x_1,\ldots,x_n)=t^{(-1)}(t(x_1)+\ldots+t(x_n)).$$

It always holds in the case of continuous t–conorms. But there exist classes of non–continuous t–conorms not fulfilling it, such as non Archimedean t–norms. More details can be found in [8].

5.2 Relationship via strong negations

Strong negations [9] are decreasing functions $n\colon [0,1]\to[0,1]$, such that $n\circ n=\mathrm{Id}_{[0,1]}$.

If T is a t-norm, then the binary operation $T^*\colon [0,1]^2\to[0,1]$ defined by:

$$T^*(x,y)=n\circ T(n(x),n(y)),$$

is a t–conorm, called the *dual* of T with respect to n.

It is said that distance functions (distances, multidistances, etc.) are *normalized* if $\mathrm{Ran}D\subseteq[0,1]$. In this case, the sum in inequalities (d3), (md3) and (md3') in Definitions 5 and 6 can be replaced by a t-conorm S. For example, for a distance d or a multidistance D we have, respectively,

(Sd3) $d(x_1,x_2)\leqslant S(d(x_1,y),d(x_2,y))$,

(Smd3) $D(x_1,\ldots,x_n)\leqslant S(D(x_1,y),\ldots,D(x_n,y))$.

So, d is an *S–distance* [7] and D is an *S–multidistance*.

In the case that S is the Maximum t–conorm, we have *ultradistances* or *multi-ultradistances*.

The next result states the close relationship between T–indistinguishabilities and S–distances. It admits more general formulations but at this point the following suffices.

Proposition 4 *Let T be a t–norm and T^* its dual t–conorm with respect to a strong negation n.*

(i) If E is a T–indistinguishability on a set $\mathbb{X}$, then $d_E=n\circ E$ is a normalized T^*–pseudodistance; d_E is a distance if and only if E separates points.

(ii) If d is a normalized T^*–pseudodistance on $\mathbb{X}$, then $E_d=n\circ d$ is a T–indistinguishability; E_d separates points if and only if d is a distance.

There exists a complete analogy with the multidimensional case.

Proposition 5 *Let T be a continuous t–norm and T^* its dual t–conorm with respect to a strong negation n.*

(i) If $\mathcal{E}$ is a T–multi-indistinguishability on a set $\mathbb{X}$, then $D_{\mathcal{E}}=n\circ\mathcal{E}$ is a normalized T^*–pseudo- multidistance on the same set; $D_{\mathcal{E}}$ is a multidistance if and only if $\mathcal{E}$ separates points.

(ii) If D is a normalized T^*–pseudo-multidistance on $\mathbb{X}$, then $\mathcal{E}_D=n\circ D$ is a T–multi-indistinguishability on $\mathbb{X}$; $\mathcal{E}_D$ separates points if and only if D is a multidistance.

Proof

A straightforward verification is only needed. If D and $\mathcal{E}$ are related by means of a strong negation,

$$\begin{aligned}\mathcal{E}(x,\ldots,x)=1 &\Leftrightarrow n\circ\mathcal{E}(x,\ldots,x)=n(1)\\ &\Leftrightarrow D(x,\ldots,x)=0,\end{aligned}$$

and thus (Tmi1) and (md1) are equivalent. If one of them is symmetric, the other one also is. Also (Tmi3) and (md3) are equivalent: for all $y\in\mathbb{X}$,

$$\begin{aligned}&\mathcal{E}(x_1,\ldots,x_n)\geqslant T(\mathcal{E}(x_1,y),\ldots,\mathcal{E}(x_n,y))\\ &\Leftrightarrow n\circ\mathcal{E}(x_1,\ldots,x_n)\leqslant n\circ T(\mathcal{E}(x_1,y),\ldots,\mathcal{E}(x_n,y))\\ &\Leftrightarrow D(x_1,\ldots,x_n)\leqslant n\circ T(n\circ D(x_1,y),\ldots,n\circ D(x_n,y))\\ &\Leftrightarrow D(x_1,\ldots,x_n)\leqslant T^*(D(x_1,y),\ldots,D(x_n,y)).\end{aligned}$$

Finally, for (Tmi4) and (md4):

$$\begin{aligned}\mathcal{E}(x_1,\ldots,x_n)=1 &\Leftrightarrow n\circ\mathcal{E}(x_1,\ldots,x_n)=1\\ &\Leftrightarrow D(x_1,\ldots,x_n)=0\end{aligned}$$

■

Obviously, if $\mathcal{E}$ is a T–multi-indistinguishability then it is a T'–multi-indistinguishability for all t–norm T' such that $T'\leqslant T$. If D is a normalized S–multidistance then it is an S'–multidistance for all t–conorm S' such that $S'\geqslant S$.

In particular, if E is a Min–multi-indistinguisability (E is a multisimilarity) then it is a T–multi-indistinguishability for all t–norm T. If D is a normalized Max–multidistance (D is an ultramultidistance) then it is an S–multidistance for all t–conorm S.

ISBN: 978-989-95079-6-8

It is interesting to observe that the ways to normalize an ordinary distance d with the formulas $d' = \frac{d}{1+d}$ and $d'' = \min\{1, d\}$ also work for multidistances, as we can see in the following result.

Proposition 6 *If D is a multidistance on $\mathbb{X}$ with values in $[0,\infty]$, then the functions D', D'' defined by*

$$D' = \frac{D}{1+D} \text{ and } D'' = \min\{1, D\}$$

are multidistances on $\mathbb{X}$ with RanD',RanD'' $\subseteq [0,1]$.

Proof
Conditions (md1), (md2) and (md4) are immediate. Let us see that condition (md3) also holds.

For $D' = \frac{D}{1+D} = 1 - \frac{1}{1+D}$ we must prove $D'(\mathbf{x}) \leqslant \sum_{i=1}^n D'(x_i, y)$ for all y, that is,

$$\begin{aligned} 1 - \tfrac{1}{1+D(\mathbf{x})} &\leqslant \textstyle\sum_{i=1}^n (1 - \frac{1}{1+D(x_i,y)} \\ &= n - \textstyle\sum_{i=1}^n \frac{1}{1+D(x_i,y)}. \end{aligned}$$

From $D(\mathbf{x}) \leqslant \sum_{i=1}^n D(x_i, y)$ we have

$$1 - \tfrac{1}{1+D(\mathbf{x})} \leqslant 1 - \tfrac{1}{1+\sum_{i=1}^n D(x_i,y)}.$$

But

$$1 - \tfrac{1}{1+\sum_{i=1}^n D(x_i,y)} \leqslant n - \textstyle\sum_{i=1}^n \frac{1}{1+D(x_i,y)},$$

because (suppose that $D(x_1, y) \leqslant D(x_i, y)$ for all i)

$$\begin{aligned} &\textstyle\sum_{i=1}^n \frac{1}{1+D(x_i,y)} - \frac{1}{1+\sum_{i=1}^n D(x_i,y)} \\ &\quad\leqslant \textstyle\sum_{i=1}^n \frac{1}{1+D(x_i,y)} - \frac{1}{1+D(x_1,y)} \\ &\quad= \textstyle\sum_{i=2}^n \frac{1}{1+D(x_i,y)} \\ &\quad\leqslant n-1. \end{aligned}$$

And for the function $D'' = \min\{D, 1\}$, the inequality $D(\mathbf{x}) \leqslant \sum_{i=1}^n D(x_i, y)$ implies that

$$\min(1, D(\mathbf{x})) \leqslant \min(1, \textstyle\sum_{i=1}^n D(x_1, y)).$$

But

$$\min(1, \textstyle\sum_{i=1}^n D(x_1, y)) \leqslant \sum_{i=1}^n \min(1, D(x_1, y)),$$

that is, $D''(\mathbf{x}) \leqslant \sum_{i=1}^n D''(x_i, y)$ for all $y \in \mathbb{X}$. ∎

6 An example: *averaged* multidistances and related *T*–multi-indistinguishabilities

Let $(\mathbb{X}, d)$ be a metric space. Let us consider the function $G\colon \bigcup_{n\geqslant 1} \mathbb{X}^n \to [0,\infty]$ defined by

$$\begin{cases} G(x_1) = 0, \\ G(x_1, \ldots, x_n) = \sum_{i<j} d(x_i, x_j), \end{cases}$$

that is, the sum of all of the pairwise ordinary distances. It is not a multidistance: it fulfills (md1), (md2) and (md4) but not (md3). For example, in the set $\mathbb{R}^2$ equipped with the Euclidean distance we take the vertexes A, B, C of a equilateral triangle of side of length 1. With these points and taking $y = A$, condition (md3) is not fulfilled:

$$G(A, B, C) = 3 \not\leqslant d(A, A) + d(B, A) + d(C, A) = 2.$$

In order to obtain multidistances based on G, we have to multiply it by a factor λ which depends on the length of the list: $\lambda = \lambda(n) \in \mathbb{R}^+$, $n \geqslant 1$.

Proposition 7 *The function $D_\lambda\colon \bigcup_{n\geqslant 1} \mathbb{X}^n \to [0,\infty]$ defined by*

$$\begin{cases} D_\lambda(x_1) = 0, \\ D_\lambda(x_1, \ldots, x_n) = \lambda(n) \sum_{i<j} d(x_i, x_j), \end{cases}$$

is a multidistance if and only if:

(i) $\lambda(2) = 1$,

(ii) $0 < \lambda(n) \leqslant \frac{1}{n-1}$ *for all* $n > 2$.

Proof
The conditions are necessary: the first is needed to fulfill $D_\lambda|_{\mathbb{X}^2} = d$ and the first inequality of the second condition, $0 < \lambda(n)$, is needed to fulfill $D_\lambda \geqslant$ (md3).

Let us see that the second inequality of (ii), $\lambda(n) \leqslant \frac{1}{n-1}$, is also necessary with a counterexample. Let us suppose that there exists n_0 such that $\lambda(n_0) > \frac{1}{n_0-1}$. We can consider the drastic distance, two different points $x, y \in \mathbb{X}$ and the list $(\overbrace{x, \ldots, x}^{n_0-1}, y) \in \mathbb{X}^{n_0}$. We have:

$$D_\lambda(\overbrace{x, \ldots, x}^{n_0-1}, y) = \lambda(n_0) \cdot (n_0 - 1) > 1,$$

but the sum of distances from the components of $(\overbrace{x, \ldots, x}^{n_0-1}, y)$ to x is 1, and therefore (md3) does not hold.

Conditions (i) and (ii) are also sufficient: condition (i) ensures that $D_\lambda \geqslant 0$. The first inequality of condition (ii) ensures (md1) and condition (md2) is fulfilled from the definition of D_λ.

Finally, let us see that condition (md3) follows from $\lambda(n) \leqslant \frac{1}{n-1}$: for any $y \in \mathbb{X}$,

$$\begin{aligned} D_\lambda(x_1, \ldots, x_n) &= \lambda(n) \textstyle\sum_{i<j} d(x_i, x_j) \\ &\leqslant \lambda(n) \textstyle\sum_{i<j} (d(x_i, y) + d(x_j, y)) \\ &= \lambda(n)(n-1) \textstyle\sum_i d(x_i, y) \\ &\leqslant \textstyle\sum_i d(x_i, y). \end{aligned}$$

∎

These multidistances are in some sense like averaged sums of the pairwise distance values for all pairs of elements of the list. In fact, if $\lambda(n) = \frac{1}{\binom{n}{2}} \leqslant \frac{1}{n-1}$, D_λ is the arithmetic mean of these distances.

Let us find the T_L and T_P–indistinguishabilities corresponding to the multidistance $D = D_\lambda$.

The generator of the Łukasiewicz t-norm is $t(x) = 1 - x$, and its pseudo-inverse is:

$$t^{(-1)} = \begin{cases} 1 - x & \text{if } x \in [0, 1], \\ 0 & \text{if } x > 1, \end{cases}$$

Thus, the T_L–multi-indistinguishability is given by $\mathcal{E}_D = t^{(-1)} \circ D$, that is:

$$\mathcal{E}_D(x_1, \ldots, x_n) = \max\{1 - \lambda(n) \sum_{i<j} d(x_i, x_j), 0\}.$$

For the product t–norm T_P, with additive generator $t = -\ln x$ and pseudo-inverse $t^{(-1)}(x) = e^{-x}$, the T_P–multi-indistinguishability is:

$$\mathcal{E}_D(x_1, \ldots, x_n) = e^{-\lambda(n) \sum_{i<j} d(x_i, x_j)}.$$

ISBN: 978-989-95079-6-8

To finalize, we are going to translate the multidistance D into T–multi-indistinguishabilities through the strong negation $n(x) = 1 - x$. First, the normalizations of D given in Proposition 6:

$$D' = \frac{D}{1+D} = 1 - \frac{1}{1+D},$$

$$D'' = \min\{D, 1\},$$

and then the negations:

$$n \circ D' = \frac{1}{1+D}.$$

$$n \circ D'' = 1 - \min\{D, 1\} = \max\{1 - D, 0\}.$$

Thus, the T–multi-indistinguishabilities are

$$\mathcal{E}'_D(x_1, \ldots, x_n) = \frac{1}{1 + \lambda(n) \sum_{i<j} d(x_i, x_j)},$$

$$\mathcal{E}''_D(x_1, \ldots, x_n) = \max\{1 - \lambda(n) \sum_{i<j} d(x_i, x_j), 0\}.$$

7 Conclusions

The concept of T–indistinguishability relation, which measures the degree of similarity between two elements of a set, has been extended with the aim that it also measures the similarity between all of the elements of a finite list. Thus, we have stated an axiomatical definition for T–multi-indistinguishabilities, and a parallel definition for multidistances and pseudo-multidistances. The close relationship between them has been studied and an example of how to obtain T–multi-indistinguishabilities from multidistances has been explained. But this is only the basis; a deeper development from both theoretical and practical points of view must be done.

Acknowledgment

The authors acknowledge the support of the Govern Balear grant PCTIB2005GC1-07 and the Spanish DGI grant MTM2006-08322.

References

[1] L. A. Zadeh. Similarity relations and fuzzy orderings. *Information Sciences*, 3:177–200, 1971.

[2] E. Trillas and L. Valverde. An inquiry into indistinguishability operators. In H.J. Skala, S. Termini, and E. Trillas, editors, *Aspects of Vagueness*, pages 231–256. Reidel, Dordrecht, 1983.

[3] E.P. Klement, R. Mesiar, and E. Pap. *Triangular norms*. Kluwer Academic Publishers, The Netherlands, 2000.

[4] M. Fréchet. Sur quelques points du calcul fonctionnel. *Rendiconti del Circolo Matematico di Palermo*, 22:1–74, 1906.

[5] J. Martín and G. Mayor. How separated Palma, Inca and Manacor are? In *Proceedings of AGOP 2009, Fifth International Summer School on Aggregation Operators*.

[6] D.H. Wolpert. Metrics for more than two points at once. In Y. Bar-Yam, editor, *Proceedings of the International Conference on Complex Systems ICCS 2004*. Perseus books, New York, in press.

[7] L. Valverde. On the structure of F–indistinguishability operators. *Fuzzy Sets and Systems*, 17(3):313–328, 1985.

[8] A. Mesiarová-Zemánková. Ranks of additive generators. *Fuzzy Sets and Systems*, in press.

[9] E. Trillas. Sobre funciones de negación en la teoría de conjuntos difusos. *Stochastica*, III(1):47–59, 1979.

ISBN: 978-989-95079-6-8

Mining Linguistic Information for Configuring a Visual Surface Inspection System

D. Martín, D.M. Guinea, M.C. García-Alegre, D. Guinea

Industrial Automation Institute (IAI), Spanish Council for Scientific Research (CSIC)
28500 Arganda del Rey, Madrid, Spain
Email: {dmartin, dmguinea, maria, domingo}@iai.csic.es

Abstract— *The configuration of a surface inspection vision system as a complex task, requires mining associations among attributes due to the variability of the surface and the environment in real-time production process. The surface inspection task has to change to deal with different elements such as, wood, stainless steel or paper inspection and, in the case of stainless steel, with reflectance and thickness variations.*

This work is an approach to mine linguistic information based on the fuzzy transform method, capable of finding associations among features. We are interested in the linguistic form of the observed associations, derived from numerical data, as the configuration task needs such linguistic knowledge. This linguistic knowledge is handled by the dynamic cognitive architecture (ARDIS), here propose, to deal with the system knowledge, represented by means of IF-THEN rules, required to configure a specific inspection system.

Keywords— Cognitive Architecture, Fuzzy Transform, Knowledge-based Vision Systems, Configuration Task, Surface Defects, Visual Inspection.

1 Introduction

The configuration of a surface inspection system requires mining associations among attributes, to avoid the data analyses performed by a human expert, due to the variability of the surface in the real-time production process. The surface inspection task changes, for different materials such as wood, stainless steel or paper and, in the case of stainless steel, the reflectance, thickness or stainless steel type varies.

This work proposes the use of the fuzzy transform method [1,2], capable of finding associations from a set of numerical features. The linguistic expressions of the observed associations, derived from numerical data is an input for the configuration of the dynamic knowledge-based architecture (ARDIS) [3], which uses IF-THEN rules to configure a specific inspection system.

Visual systems for inspecting surface defects have always been present in the laminated materials industry [4]. Nevertheless, these systems still show several drawbacks, such as the vision system reusing, as they are designed for a particular surface inspection application and do not offer the possibility of changing either the objectives or the inspection goals. They are designed for a specific application and are not ready to hold unexpected variations. This fact entails a high cost as each new visual inspection task has to be redesigned, from the superficial defect analysis to the overall inspection system by a human expert. The solution here proposed points to extract linguistic knowledge on surface visual inspection, which is easily differentiated in types, by means of associations. This would allow changing or reconfiguring the components related to the inspection process that vary, in the production line, due to changes on either camera, surface or defect type.

To this aim a visual inspection architecture, namely ARDIS, is proposed to surface dynamic inspection in laminated materials based on the configuration of a specific visual system to obtain a good quality control of the manufacturing surface. The configuration task for surface inspection is analysed at the knowledge level and the task is decomposed into simple subtasks to reach the inference level, the most basic tasks. This task is solved as a Configuration-Design task following the CommonKADS methodology [5] leading to the knowledge-based dynamic architecture (ARDIS), which can account with all the knowledge involved in the process. The surface inspection generic knowledge is differentiated among environment, image quality, real-time and computer vision techniques to be integrated in ARDIS.

The injection of surface inspection knowledge from the human expert is an essential issue in the configuration process of any visual inspection system. For instance, the change of the surface thickness or lighting in a production line implies a variation that is not considered in current inspection systems. In this case, the distance between the camera and the surface varies and the camera gets out of focus, generating blurred images. Thus, expert knowledge is necessary to solve the problem as human experts do. But often, we lack knowledge from the expert, so mining linguistic information from numerical data will aid the architecture to re-configure, just in time, the inspection system.

In ARDIS, the surface inspection task, at the domain knowledge level, is analysed and decomposed into simple subtasks to reach the inference level that allows differentiating the Environment (E), Image Quality (IQ), Real-Time (RT) and Computer Vision Techniques (CVT) requirements.

This work is a first approach, to obtain the associations among E, IQ, RT and CVT, to configure the parameters of the inspection system to carry out a specific surface inspection task. In stainless steel inspection, real-time parameters could affect the more the CVT parameters.

ISBN: 978-989-95079-6-8

To reach the proposed goal, numerical data from E, IQ, RT and CVT are obtained and evaluated over a set of stainless steel and wood images. Results are shown where CVT is selected as a dependent attribute and E, IQ, RT are independent attributes.

2 ARDIS a cognitive architecture

The cognitive architecture, ARDIS, is proposed to deal with the complex task of visual inspection of laminated surfaces in industrial environments displaying a high variability in lighting, reflectance and real-time defect detection conditions. The generic expert knowledge on surface visual inspection and the deep knowledge on the inspection restrictions and defects characteristics that has the line inspector, is used in the design of a visual inspection system.
The architecture has to offer an adaptive behaviour for detecting different defects or inspecting different type of surfaces. To this aim the cognitive architecture is designed so as to be able to configure in real-time a visual inspection system, adapted for each specific type of laminated surface and defect.
The proposed architecture allows using generic knowledge on surface inspection to configure a specific system of visual surface inspection for different laminated materials such as, metal or wood. This knowledge, in the static roles of the architecture, is codified by means of crisp and fuzzy rules.
The configured visual inspection system is composed of a set of components, design elements, and their parameters. Each configured component depends on the type of surface and defect to be inspected. An environment component can be the lighting device or the camera, which could be composed of several sub-components, each one described by a set of parameters.
The method Propose-&-Revise used to solve the visual surface inspection task is displayed in Figure 1, showing the diagram of subtasks and inferences. This main method shows how the ARDIS architecture operates in three steps: (1) the initialization process, (2) the extension of the initial inspection skeleton and, (3) the revision process of the overall inspection system.

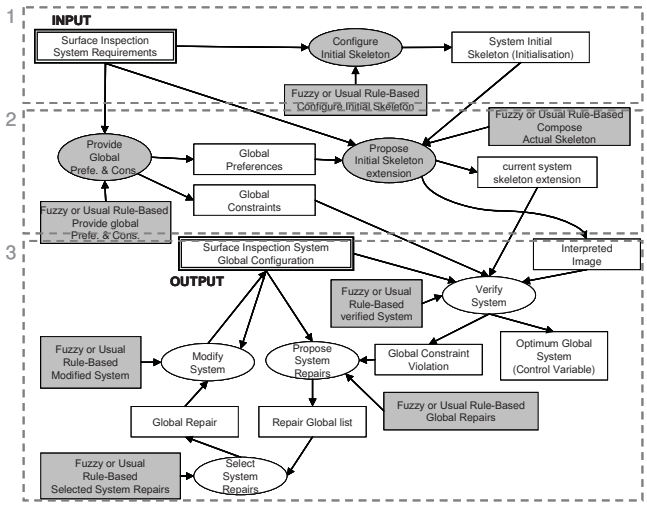


Figure 1: Subtask and inference diagram of the method Propose-&-Revise to solve the surface inspection task.

2.1 The domain generic knowledge on surface inspection

The domain generic knowledge on surface inspection is composed of different types of generic knowledge on real-time restrictions, image quality control restrictions, environment conditions and computer vision techniques. This differentiation of the type of knowledge makes possible to distinguish the knowledge used in each inference, making easier the specification of components and their interactions. Thus, the ARDIS architecture has been provided with all the necessary functionalities required to configure a specific complete system of visual inspection where knowledge is partitioned to be ease reused.
The relationships between the knowledge types are described in Figure 2. Here, the environment knowledge is related to real-time knowledge, such as image acquisition rate or lighting components that can influence real-time components in one or another way. So, during the configuration step of an inspection system, if an environment component is configured this has to be taken into account in the configuration of the real-time components.

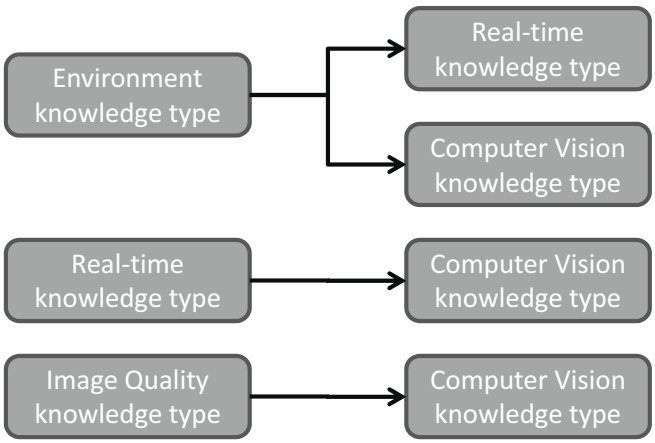


Figure 2: Influence diagram in the surface inspection process

2.2 Dynamic configuration using ARDIS architecture

The dynamic configuration process relies on (i) the cognitive architecture ARDIS, (ii) the domain generic knowledge on surface inspection and (iii) the line inspector knowledge who selects the requirements of the inspection system. This dynamic configuration process allows the inspection of each specific surface in a production line. Consequently, it is possible to inspect stainless steel or wood laminates in the same production line. Following the control structure of the ARDIS architecture, requirements and generic knowledge are integrated into crisp and fuzzy rules. This process gives rise to the configuration of the components of the surface inspection system. The Table 1 shows the set of rules used to configure the environment components of the inspection system.

ISBN: 978-989-95079-6-8

Table 1: Knowledge Base for configuring the surface inspection system.

KNOWLEDGE BASE USED IN THE INFERENCES OF THE SUBTASK Propose Environment Skeleton Extension
IF ProductionLineSpeed is high THEN ExpositionTime is short
IF IlluminationSystem is ExteriorLaserIllumination THEN CameraSensorGain is automatic
IF IlluminationSystem is ExteriorLaserIllumination THEN Camera-IlluminationRelativePosition is 45degrees
IF IlluminationType is GreenLaserSource THEN ImageChannel is Green

3 Stainless steel and wood images

The work departs from a set of images from inspections of residual oxide scale on cold stainless steel strip and wood. The first visual system is configured to detect 50 microns size defects and the second one to inspect larger defects in a wood surface. The images are obtained with a robust laser technique for diffuse illumination which allows stainless steel or wood industrial inspection with the same system. This visual technique for surface inspection consists on a smart vision system based on a green laser diode diffuse illumination, thus images display a green predominant colour.

The Figure 3 shows a set of images of stainless steel to inspect micro oxide defects, and Figure 4 displays a new laminated material, wood, where knot defects and wood irregularities have to be inspected in the same production line. The wood and stainless steel images have been acquired with an experimental system based on green laser illumination which allows inspecting defects ranging micro to five millimetres size. The acquisition visual system utilizes magnification 1.

4 Defining parameters from images

The analysis of the images has been accomplished with a HP xw4600 Workstation and "Intel(R) Core(TM)2 Quad CPU Q6600 @ 2.40GHz" processor. The numeric parameters have been calculated using the image processing toolbox of MATLAB [6].

Parameters of environment (E), image quality (IQ), real-time (RT) and computer vision techniques (CVT) are differentiated and obtained from the original image acquired by the visual inspection system. The E, IQ, RT and CVT parameters are all derived from an image, either stainless steel or a wood, and the evaluation is performed under the same conditions.

Several features are obtained from images but no associations among them. At this point, some additional knowledge is required to establish the priority on the selected parameter (E, IQ, RT) for the configuration of the inspection system.

The numerical data and F-transform will aid to mine the influence of the E, IQ and RT parameters on the CVT parameters. This knowledge will be, later on, used to set the priority in the configuration process. Then, fixing a CVT segmentation technique, we will try to find the influence between "E, IQ and RT" and the CVT technique. Thus, a segmentation technique could be more influenced by real-time inspection than a non-uniform lighting in the environment.

The linguistic information obtained for configuring a visual surface inspection system has the following structure: IF SegmentationTechnique is CVT and SurfaceInspection is StainlessSteel THEN ParameterPriority is Real-Time.

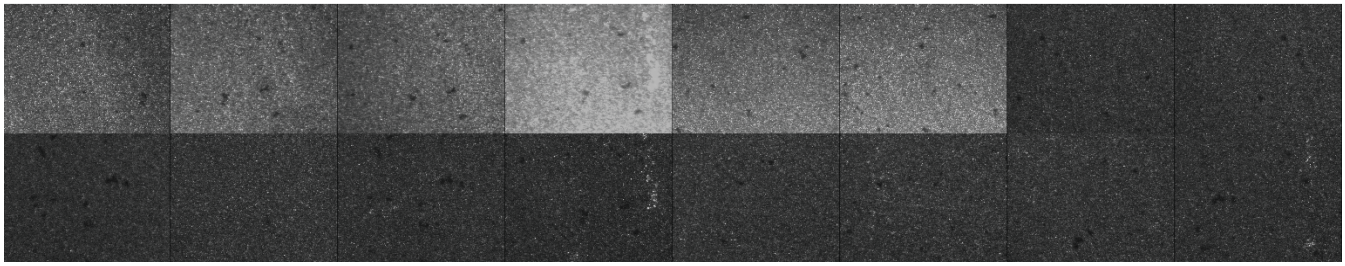


Figure 3: Images of stainless steel with micro oxide defects

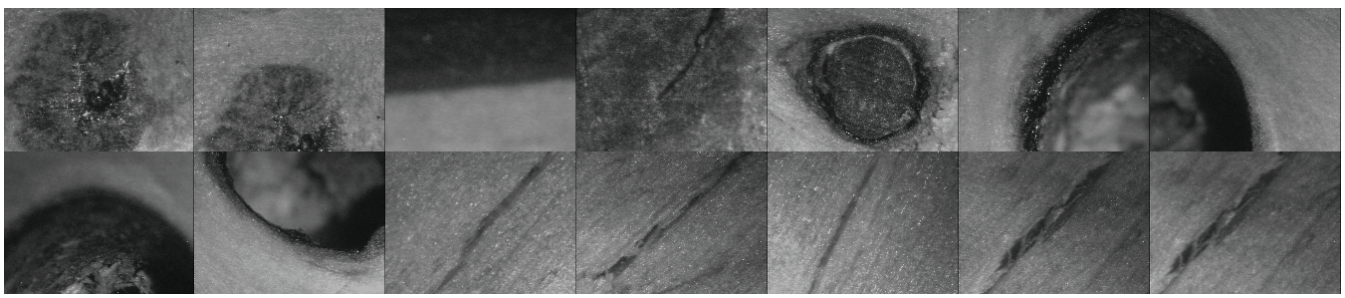


Figure 4: Images of wood with knot defect and other wood irregularities

ISBN: 978-989-95079-6-8

The parameters chosen to evaluate the E, IQ, RT and CVT properties over the acquired image, selecting the Mean-Shift algorithm [7] as the segmentation technique, are:

1) Evaluation of the Environment (E): the lighting non-uniformity of the image is used to measure the influence of the environment in the image. This environment evaluation will be an independent attribute or parameter and the corresponding contexts are $E_{StainlessSteel}$= [0.0736 8.1698], E_{wood}= [8.0138 50.2337]. Figure 5 shows data obtained from stainless steel and wood images.

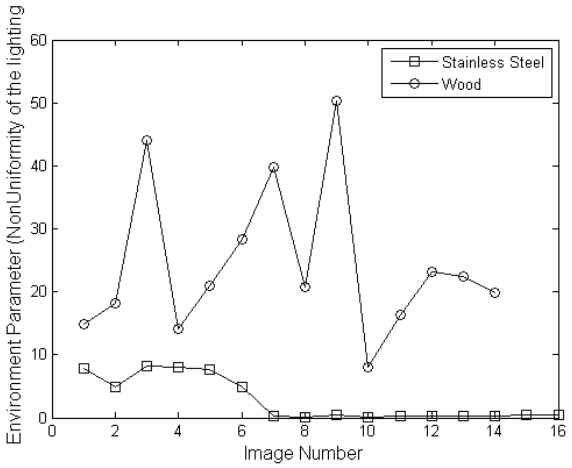


Figure 5: Evolution of the lighting in Stainless Steel and Wood images.

2) Evaluation of the Real-Time (RT): the computation time of the Mean-Shift algorithm [7] is used to estimate a Real-Time parameter in the image. The Real-Time evaluation will be an independent attribute or parameter and the corresponding contexts are $RT_{StainlessSteel}$ = [8.3594 12.3281], RT_{Wood}= [10.1719 13.7813]. Figure 6 shows data obtained on a set of stainless steel and wood images.

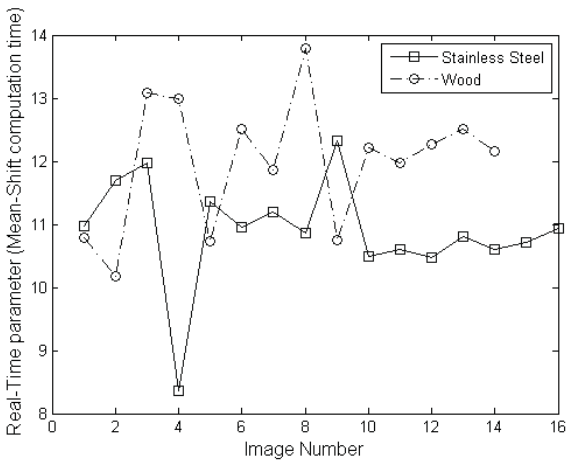


Figure 6: Evolution of the Mean-Shift computation time in Stainless Steel and Wood images.

3) Evaluation of Image Quality (IQ): gray-level co-occurrence matrix from images, was used to calculate the contrast property from the gray-level co-occurence matrix which returns a measure of the intensity contrast between a pixel and its neighbour over the whole image. The Image Quality evaluation will be an independent parameter and the corresponding contexts are $IQ_{StainlessSteel}$ = [5.0702 35.2404], IQ_{Wood} = [0.2268 10.5327]. Figure 7 shows data obtained from stainless steel and wood images.

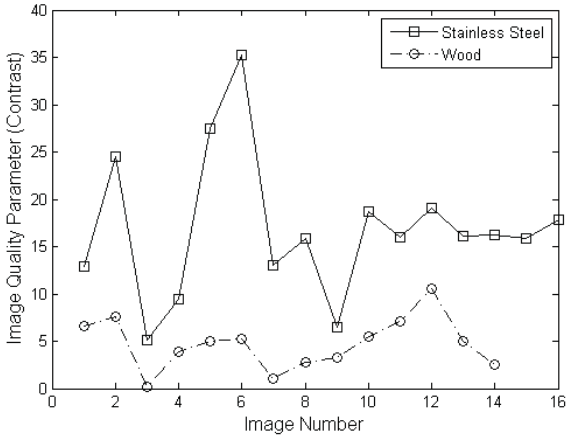


Figure 7: Evolution of the contrast in Stainless Steel and Wood images.

4) Evaluation of the segmentation technique (CVT): the Mean-Shift technique is used as the segmentation technique and the evaluation is based on the number of Mean-Shift regions located by the algorithm. This CVT evaluation will be the dependent parameter and the corresponding contexts are $CVT_{StainlessSteel}$ = [919 2968], CVT_{Wood} = [8 243]. Figure 8 shows data obtained from stainless steel and wood images.

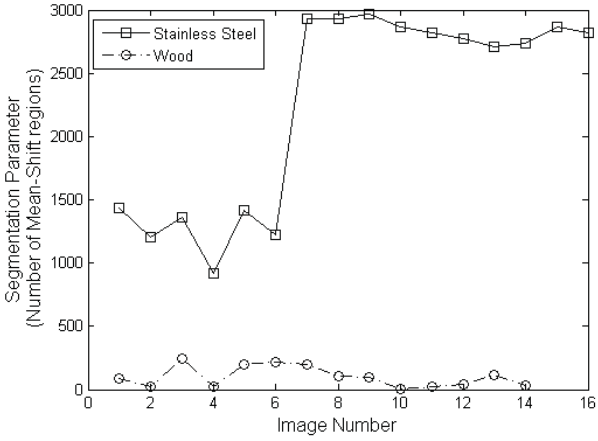


Figure 8: Evolution of the number of Mean-Shift regions in Stainless Steel and Wood images.

ISBN: 978-989-95079-6-8

5 Membership functions for data analysis

The membership functions map points in the input space to a membership value (or degree of membership) between 0 and 1.

If X is the input space and its elements are denoted by x, then a fuzzy set A in X is defined as a set of ordered pairs.

$$A = \{x, \mu_A(x) \mid x \in X\}$$

Where $\mu_A(x)$ is the membership degree of x in A.

Gaussian functions are selected to represent membership functions, due to their smoothness and concise notation. These curves have the advantage of being smooth and nonzero at all points.

The membership functions selected for the environment (E), real-time (RT), image quality (IQ) and computer vision techniques (CVT) parameters are shown in the Figures 9, 10 and the vectors of nodes $\langle l_1,...,l_k \rangle$ which define the membership functions of the induced fuzzy sets on the E, IQ, RT and CVT parameters for Stainless Steel and Wood, are displayed in Table 2.

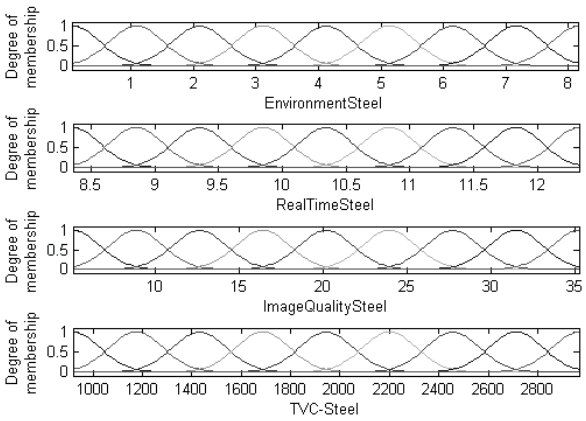


Figure 9: Membership functions with 9 nodes for E, IQ, RT and CVT parameters on Stainless Steel images.

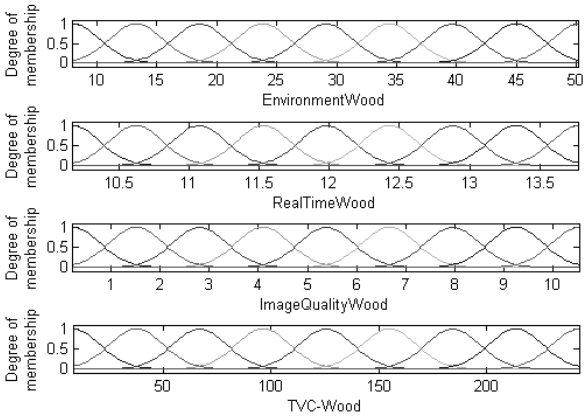


Figure 10: Membership functions with 9 nodes for E, IQ, RT and CVT parameters on Wood images.

Table 2: Vectors of nodes of the membership functions.

	Stainless Steel	*Wood*
Environment **(E)**	[0.0736, 1.086, 2.098, 3.11, 4.122, 5.134, 6.146, 7.158, 8.17]	[8.014, 13.29, 18.57, 23.85, 29.12, 34.4, 39.68, 44.96, 50.23]
Real-Time **(RT)**	[8.359, 8.855, 9.352, 9.848, 10.34, 10.84, 11.34, 11.83, 12.33]	[10.17, 10.62, 11.07, 11.53, 11.98, 12.43, 12.88, 13.33, 13.78]
Image Quality **(IQ)**	[5.07, 8.841, 12.61, 16.38, 20.16, 23.93, 27.7, 31.47, 35.24]	[0.2268, 1.515, 2.803, 4.092, 5.38, 6.668, 7.956, 9.244, 10.53]
Computer Vision Techniques **(CVT)**	[919, 1175, 1431, 1687, 1944, 2200, 2456, 2712, 2968]	[8, 37.38, 66.75, 96.13, 125.5, 154.9, 184.3, 213.6, 243]

6 Mining associations between environment, real-time, image quality and computer vision techniques parameters: A case study

The proposed method is based on F-transform to search dependence between the CVT segmentation technique and the E, IQ, RT parameters. These associations allow to determine the priority of the parameters E, IQ, RT for configuring a visual surface inspection system.

Following the applications of F-transform to data analysis [2], we look for dependences among some attributes having a general expression of the type: $X_z = H(X_1,...,X_k)$, where X_z is the dependent attribute, such as the CVT segmentation technique and $X_1,\ldots,X_k$ are the independent attributes as E, IQ, RT parameters. The evaluation of these attributes is carried out on a set of 16 images (objects $\{o_1,\ldots,o_j,\ldots,o_{16}\}$) for Stainless Steel inspection and 14 images (objects $\{o_1,\ldots,o_j,\ldots,o_{14}\}$) for Wood inspection.

This first approach selects a computer vision technique for image segmentation, such as the Mean-Shift algorithm and afterwards, the F-transform is used, as well as the "*degree of support*" (r) and "*confidence*" (γ) parameters [2], for mining associations in the data, finding hypothesis of potential dependences among the selected segmentation technique and the E, RT, IQ parameters. The *degree of support* (r) is defined by (1):

$$r = \frac{|\{o_j \mid Fn[l_1...l_k](o_j) > 0\}|}{m} \quad (1)$$

Where m is the number of objects, Fn $[l_1\ldots l_k](o_j)$ is the membership function of an induced fuzzy set on the set of objects $\{o_1,...,o_m\}$ and $|\cdot|$ denotes a number of elements of the given set.

The *degree of confidence* (γ) using the F-transform and inverse F-transform is formulated in (2):

$$\gamma = \sqrt{\frac{\sum_{j=1}^{m}(f_F(o_j) - F_{l_1...l_k})^2 \cdot Fn[l_1...l_k](o_j)}{\sum_{j=1}^{m}(f_{jz} - F_{l_1...l_k})^2 \cdot Fn[l_1...l_k](o_j)}} \quad (2)$$

Where, $f_F(o_j)$ is an auxiliary function on the set of objects $\{o_1,...,o_m\}$ induced by the inverse F-transform with components $F_{l_1 \ldots l_k}$ of F-transform, and f_{jz} is the value of the dependent attribute X_z measured on the object o_j.

Finally, each association is supported by the mean of the "*degrees of support*" (r_{mean}) and "*confidence*" (γ_{mean}) parameters, as each r and γ corresponding to the *k*-tuple of fuzzy sets $\langle Fn[l_1],\cdots,Fn[l_k]\rangle$ is analysed, therefore the mean value represents a meaningful value of r and γ.

The results are presented under the format (3):

$$(X_1 \; is \; \langle Fn[l_1],\cdots,Fn[l_k]\rangle) \underset{mean\; r,\gamma}{\overset{F}{\sim}} X_z(segmentation) \quad (3)$$

Then applying the F-transform, three associations for Stainless Steel and Wood are obtained, using the segmentation algorithm (Mean-Shift):

$$(E_{\substack{Stainless\\Steel}} \; is \; \langle Fn[l_1],\cdots,Fn[l_k]\rangle) \underset{\substack{mean\\r=0.3194\\\gamma=0.00021982}}{\overset{F}{\sim}} CVT_{\substack{Stainless\\Steel}}(Mshift)$$

$$(RT_{\substack{Stainless\\Steel}} \; is \; \langle Fn[l_1],\cdots,Fn[l_k]\rangle) \underset{\substack{mean\\r=0.3889\\\gamma=0.00040048}}{\overset{F}{\sim}} CVT_{\substack{Stainless\\Steel}}(Mshift)$$

$$(IQ_{\substack{Stainless\\Steel}} \; is \; \langle Fn[l_1],\cdots,Fn[l_k]\rangle) \underset{\substack{mean\\r=0.3750\\\gamma=0.0015}}{\overset{F}{\sim}} CVT_{\substack{Stainless\\Steel}}(Mshift)$$

$$(E_{Wood} \; is \; \langle Fn[l_1],\cdots,Fn[l_k]\rangle) \underset{\substack{mean\\r=0.3968\\\gamma=0.0547}}{\overset{F}{\sim}} CVT_{Wood}(Mshift)$$

$$(RT_{Wood} \; is \; \langle Fn[l_1],\cdots,Fn[l_k]\rangle) \underset{\substack{mean\\r=0.4048\\\gamma=0.0084}}{\overset{F}{\sim}} CVT_{Wood}(Mshift)$$

$$(IQ_{Wood} \; is \; \langle Fn[l_1],\cdots,Fn[l_k]\rangle) \underset{\substack{mean\\r=0.3810\\\gamma=0.0142}}{\overset{F}{\sim}} CVT_{Wood}(Mshift)$$

The linguistic approach of these associations for configuring a visual surface inspection system can be formulated as:

Stainless Steel:
"IF $CVT_{StainlessSteel}$ is Mean-Shift THEN ConfigurationPriority is IQ"
with *degree of support* r=0.3750 and *confidence* γ=0.0015.

The three Stainless Steel associations, highlight the fact that Mean-Shift technique has a stronger influence on image quality than do the others parameters, so the priority of the Stainless Steel configuration process is the image quality parameters. This association is quite obvious as a consequence of the influence of small defects and huge number of regions, Figure 8, detected by the Mean-Shift algorithm and the high frequency of the images.

Wood:
"IF CVT_{Wood} is Mean-Shift THEN ConfigurationPriority is E"
with the *degree of support* r= 0.3968 and *confidence* γ= 0.0547

The three Wood associations points that Mean-Shift has a larger influence on environment than the others parameters, so the priority of the Wood configuration process is the environment parameters. This association is also obvious due to the influence of huge defects and small number of regions, Figure 8, detected by the Mean-Shift algorithm and the low frequency of the images.

7 Conclusions

This work proposes a first approach to derive associations among environment, real-time, image quality and computer vision techniques, based on the F-transform and inverse F-transform.
This is the first step to tune the parameters of a visual inspection system to accomplish a specific surface inspection task, either on Stainless steel or Wood.
The mean of r and γ, is used to summarize the results instead of using individual r and γ values. The derived associations are quite obvious and suggest hypothesis that correspond to expert knowledge.
Future work will be performed on applications that use the linguistic form of the observed associations from numerical data as the configuration task for surface inspection that requires such linguistic knowledge.

Acknowledgment

Authors express their gratitude to ACERINOX S.A. (Spanish stainless steel manufacturing company) for fruitful cooperation and financial support and CICYT- DPI - 2006-14497 for insightful knowledge on cognitive architectures. Authors also thank Eugenio Villanueva for his valuable aid and comments in technical aspects of this work.

References

[1] I. Perfilieva. Fuzzy transforms: Theory and applications. Fuzzy Sets and Systems, 157:993-1023, 2006.

[2] I. Perfilieva, V. Novák and A. Dvořák. Fuzzy transform in the analysis of data. International Journal of Approximate Reasoning, 48:36-46, 2008.

[3] D. Martín. Arquitectura dinámica para inspección visual de superficies en tiempo real. PhD thesis, Madrid, Spain: UNED, 2008.

[4] T. S. Newman and A. K. Jain, 'A survey of automated visual inspection', Computer Vision and Image Understanding, 61:231–262, 1995.

[5] G. Schreiber, H. Akkermans, A. Anjewierden, R. de Hoog, N. Shadbolt, W. V. de Velde, and B.Wielinga, Knowledge Engineering and Manage-ment: The CommonKADS Methodology, The MIT Press, 2000.

[6] Matlab, The MathWorks Inc. Natick, MA, United States, http://www.mathworks.com/

[7] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(5): 603-619, May 2002.

ISBN: 978-989-95079-6-8

A construction of an L-fuzzy valued measure of L-fuzzy sets

Vechislav Ruzha, Svetlana Asmuss

Department of Mathematics, University of Latvia
Zellu street 8, Riga, LV-1002, Latvia
Institute of Mathematics and Computer Science of University of Latvia
Rainis blvd. 29, Riga, LV-1459, Latvia
Email: vecislavs.ruza@ge.com, svetlana.asmuss@lu.lv

Abstract— *We develope a construction of an L-fuzzy valued measure by extending a measure defined on a σ-algebra of crisp sets to an L-fuzzy valued T_M-countably additive measure defined on a T_M-clan of L-fuzzy sets in the case when operations for L-fuzzy sets and L-fuzzy real numbers are based on the minimum t-norm T_M.*

Keywords— L-fuzzy set, L-fuzzy real number, L-fuzzy valued measure.

1 Introduction

There are many works devoted to measures of fuzzy sets. The most important concepts and results concerning this topic are considered by D. Butnariu and E.P. Klement [2]. Our interest is in developing a theory of measure and integral where not only sets are fuzzy, but also measure and integral take fuzzy real values. To do this we need a concept of a fuzzy real number. There are several (quite different) ways to define a fuzzy real number. For our purposes we use fuzzy real numbers as they were first defined by B. Hutton [3], and then studied thoroughly in a series of papers (see e.g.[4],[5]). The preference of using this approach for defining fuzzy real numbers is motivated by our intention to develop the investigations from [6],[7].

In our previous work [1] we suggested the construction of a T-measure of L-fuzzy sets by extension of a measure defined on a σ-algebra of crisp sets to a T-measure on a T-tribe in the case of the minimum t-norm T_M and $L=[0,1]$. The main purpose of the present paper is to show that the construction still can be generalized in the case when L is complete and completely distributive lattice $L(\wedge,\vee,0_L,1_L)$. It was achieved in the case of the minimum t-norm T_M: $T_M(x,y)=x\wedge y$.

2 Preliminaries

Let X be a nonempty set. The class of all L-fuzzy subsets of X (in the sequel L-fuzzy sets) is denoted L^X. The operations for L-fuzzy sets $A,B\in L^X$ such as intersection, union and difference are defined by using a minimum triangular norm T_M, its corresponding triangular conorm S_M and an involution N:
$(A\cap B)(x)=A(x)\wedge B(x)$, $(A\cup B)(x)=A(x)\vee B(x)$,
$(A\setminus B)(x)=A(x)\wedge N(B(x))$.

2.1 Classes of L-fuzzy sets

In order to define an L-fuzzy valued function of L-fuzzy sets we consider such classes of L-fuzzy sets as T_M-semirings (defined by analogy with classical case, see e.g. [8]) and T_M-clans (see e.g. [2]).

Definition 2.1. *A subclass $\wp\subset L^X$ is called a T_M-semiring on X if the following properties are satisfied:*

(i) $\emptyset\in\wp$;

(ii) for all $A,B\in\wp$ *we have* $A\wedge B\in\wp$;

(iii) for all $A,B\in\wp$ *there exist such T_M-disjoint L-fuzzy sets* $A_1,A_2,...,A_n\in\wp$ *that* $A\setminus B=\bigvee_{i=1}^{n}A_i$.

A finite family of L-fuzzy sets $A_1,A_2,\ldots,A_n$ is said to be *T_M-disjoint* (see e.g.[2]) iff for each $k\in\{1,\ldots,n\}$

$$(\bigvee_{j=1,j\neq k}^{n}A_j)\wedge A_k=\emptyset.$$

A countable family of L-fuzzy sets is said to be T_M-disjoint iff every finite subfamily of this family is T_M-disjoint.

Definition 2.2. *A subclass $\mathcal{A}\subset L^X$ is called a T_M-clan on X iff the following properties are satisfied:*

(i) $\emptyset\in\mathcal{A}$;

(ii) for all $A\in\mathcal{A}$ *we have* $N(A)\in\mathcal{A}$;

(iii) for all $A,B\in\mathcal{A}$ *we have* $A\wedge B\in\mathcal{A}$.

2.2 L-fuzzy real numbers

For our purposes we use the L-fuzzy real numbers as they were first defined by B. Hutton [3].

Definition 2.3. *An L-fuzzy real number is a function $z:\mathbb{R}\to L$ such that*

(i) z is non-increasing;

(ii) $\bigwedge_x z(x)=0_L$, $\bigvee_x z(x)=1_L$;

(iii) z is left semi-continuous, i.e. $\bigwedge_{t<x} z(t)=z(x)$.

The set of all L-fuzzy real numbers is called *the L-fuzzy real line* and it is denoted by $\mathbb{R}(L)$. An L-fuzzy number z is called *non-negative* if $z(0)=1_L$. The set of all non-negative L-fuzzy real numbers we denote by $\mathbb{R}_+(L)$.

The ordinary real line $\mathbb{R}$ can be identified with the subspace

$$\{z_a\mid a\in\mathbb{R}\}\subset\mathbb{R}(L)$$

ISBN: 978-989-95079-6-8

by assigning to a real number $a \in \mathbb{R}$ the fuzzy real number z_a defined by

$$z_a(x) = \begin{cases} 1_L, \text{ if } x \leq a, \\ 0_L, \text{ if } x > a. \end{cases}$$

The operation of L-fuzzy addition $\oplus$ defined by

$$(z_1 \oplus z_2)(x) = \bigvee_t \left(\{ z_1(t) \wedge z_2(x-t) \} \right)$$

whenever $z_1, z_2 \in \mathbb{R}(L)$, is a jointly continuous extension of addition $+$ from the real line $\mathbb{R}$ to the L-fuzzy real line $\mathbb{R}(L)$.

Given a sequence of non-negative L-fuzzy real numbers $(z_n)_{n\in\mathbb{N}} \subset \mathbb{R}_+(L)$ we consider the countable sum

$$\bigoplus_{n\in\mathbb{N}} z_n = \bigvee_{n\in\mathbb{N}} (z_1 \oplus z_2 \oplus ... \oplus z_n).$$

2.3 L-fuzzy valued functions

Let $\aleph$ be a class of L-fuzzy sets. Within this section some basic properties of an L-fuzzy valued function $\eta : \aleph \to \mathbb{R}_+(L)$ are considered.

Definition 2.4. *An L-fuzzy valued function η is called T_M-additive iff*

for all $A, B \in \aleph$ such that $A \wedge B = \emptyset, A \vee B \in \aleph$ it holds $\eta(A \vee B) = \eta(A) \oplus \eta(B)$.

Definition 2.5. *An L-fuzzy valued function η is called T_M-countably additive iff*

for all $(A_n)_{n\in\mathbb{N}} \subset \aleph$ such that
$(\forall i, j \in \mathbb{N} : i \neq j \Rightarrow A_i \wedge A_j = \emptyset)$ and $\bigvee_{n=1}^{\infty} A_n \in \aleph$
it holds $\eta(\bigvee_{n=1}^{\infty} A_n) = \bigoplus_{n\in\mathbb{N}} \eta(A_n)$.

Definition 2.6. *Let $\wp$ be a T_M-semiring. A function $m : \wp \to \mathbb{R}_+(L)$ is called an L-fuzzy valued elementary T_M-measure if it satisfies the following conditions:*

(i) $m(\emptyset) = z_0$;

(ii) m is T_M-additive.

Definition 2.7. *Let $\mathcal{A}$ be a T_M-clan. A function $\mu : \mathcal{A} \to \mathbb{R}_+(L)$ is called an L-fuzzy valued T_M-countably additive measure if it satisfies the following conditions:*

(i) $\mu(\emptyset) = z_0$;

(ii) μ is T_M-countably additive.

3 A construction of an L-fuzzy valued elementary T_M-measure

Let Φ be a σ-algebra of "crisp" subsets of X and ν is a finite measure $\nu : \Phi \to [0, +\infty[$. Our aim is to construct an L-fuzzy valued T_M-countably additive measure $\tilde{\mu}$ on a T_M-clan by extension a crisp measure ν. To achieve this we generalize the well known construction of "classic" measure theory (see e.g. [8]) to the L-fuzzy case.

3.1 T_M-semiring $\tilde{\wp}$ of L-fuzzy sets

To realise the construction we use the special type of L-fuzzy sets with respect to σ-algebra Φ of crisp sets on X. For $M \in \Phi$, $\alpha \in L$ we define L-fuzzy set $A(M, \alpha)$:

$$(A(M,\alpha))(x) = \begin{cases} \alpha, x \in M, \\ 0_L, x \notin M. \end{cases}$$

For convenience we denote this special type of L-fuzzy sets by $A(M, \alpha)$ (the choice of a letter in the notation is not crucial: $B(N, \beta)$ is still the same type of an L-fuzzy set).

Proposition 3.1. *The class of L-fuzzy sets*

$$\tilde{\wp} = \{ A(M,\alpha) | M \in \Phi \text{ and } \alpha \in L \}$$

is a T_M-semiring.

Proof. For all $A(M,\alpha), B(K,\beta) \in \tilde{\wp}$ we have
$A(M,\alpha) \wedge B(K,\beta) = C(M \cap K, \alpha \wedge \beta) \in \tilde{\wp}$ and
$A(M,\alpha) \setminus B(K,\beta) = A_1(M \setminus K, \alpha) \vee A_2(M \cap K, \alpha \wedge N(\beta))$,
where $A_1(M \setminus K, \alpha), A_2(M \cap K, \alpha \wedge N(\beta)) \in \tilde{\wp}$.
Taking into account that also $\emptyset = A(\emptyset, 0_L) \in \tilde{\wp}$, it is proved that $\tilde{\wp}$ is a T_M-semiring. □

3.2 L-fuzzy valued elementary T_M-measure $\tilde{m}$

We define an L-fuzzy valued function $\tilde{m}$ on the T-semiring $\tilde{\wp}$ by the formula $\tilde{m}(A(M,\alpha)) = z_{\nu(M),\alpha}$, where

$$z_{\nu(M),\alpha}(t) = \begin{cases} 1_L, t \leq 0, \\ \alpha, 0 < t \leq \nu(M), \\ 0_L, t > \nu(M), \end{cases}$$

is an L-fuzzy real number.

Proposition 3.2. *$\tilde{m}$ is an L-fuzzy valued elementary T_M-measure.*

Proof. It is easy to see that

$$\tilde{m}^*(\emptyset) = z_{\nu(\emptyset),0_L} = z_0$$

and the equality

$$\tilde{m}(A(M,\alpha)) \oplus \tilde{m}(B(K,\beta)) = \tilde{m}(A(M,\alpha) \vee B(K,\beta))$$

is true if $A(M,\alpha) = \emptyset$ or $B(K,\beta) = \emptyset$.

Let us consider $A(M,\alpha), B(K,\beta) \in \tilde{\wp}$ such that

$$A(M,\alpha) \wedge B(K,\beta) = \emptyset,\ A(M,\alpha) \cup B(K,\beta) \in \tilde{\wp} \text{ and } \alpha \neq 0_L, \beta \neq 0_L.$$

It follows that $M \cap K = \emptyset$ or $\alpha \wedge \beta = 0_L$.

If $M \cap K = \emptyset$ then $\alpha = \beta$ and in this case

$$\tilde{m}(A(M,\alpha)) \oplus \tilde{m}(B(K,\alpha)) = z_{\nu(M),\alpha} \oplus z_{\nu(K),\alpha} =$$
$$= z_{\nu(M)+\nu(K),\alpha} = z_{\nu(M\cup K),\alpha} = \tilde{m}(A(M,\alpha) \vee B(K,\alpha)).$$

If $M \cap K \neq \emptyset$ then it is sufficient to consider the case when $\alpha \wedge \beta = 0_L$. In this case α and β are incomparable (due to the assumptions that $\alpha \neq 0_L, \beta \neq 0_L$). Because of

$$A(M,\alpha) \vee B(K,\beta) =$$
$$= C_1(M \cap K, \alpha \vee \beta) \vee C_2(M \setminus K, \alpha) \vee C_3(K \setminus M, \beta) \in \tilde{\wp}$$

we obtain that $K \setminus M = \emptyset$, $M \setminus K = \emptyset$ and hence $M = K$. Then

$$\tilde{m}(A(M,\alpha)) \oplus \tilde{m}(B(M,\beta)) = z_{\nu(M),\alpha} \oplus z_{\nu(M),\beta} =$$
$$= z_{\nu(M),\alpha\vee\beta} = C(M, \alpha \vee \beta) = \tilde{m}(A(M,\alpha) \vee B(M,\beta)).$$

By this we prove that $\tilde{m}$ is T_M-additive. □

ISBN: 978-989-95079-6-8

4 Measurable L-fuzzy sets

4.1 An extension of an L-fuzzy valued elementary T_M-measure

We define an L-fuzzy valued function $\tilde{m}^* : L^X \to \mathbb{R}_+(L)$ by the following formula

$\tilde{m}^*(E) = \bigwedge\{\bigoplus_{n=1}^{\infty} \tilde{m}(E_n) \mid (E_n)_{n\in\mathbb{N}} \subset \tilde{\wp}:$

$E \le \bigvee_{n=1}^{\infty} E_n$ and $(E_n)_{n\in\mathbb{N}}$ is T_M-disjoint$\}, E \in L^X$.

Remarks.

(i) For every $E \in L^X$ there always exists such a sequence $(E_n)_{n\in\mathbb{N}} \subset \tilde{\wp}$ that $E \le \bigvee_{n=1}^{\infty} E_n$. It is enough to take $E_1(X, 1_L)$. Thus $\tilde{m}^*$ is bounded from above in the following sense: $\tilde{m}^*(E) \le z_{\nu(X),1_L}$ for all $E \in L^X$.

(ii) For all $E \in \tilde{\wp}$ we have $\tilde{m}^*(E) = \tilde{m}(E)$.

Proposition 4.1. *For all $A, B \in L^X$ we have*

$$\tilde{m}^*(A) \oplus \tilde{m}^*(B) \ge \tilde{m}^*(A \wedge B) \oplus \tilde{m}^*(A \vee B).$$

Proof. For a given L-fuzzy sets A and B we consider two T_M-disjoint sequences:

$$(C_i(M_i, \alpha_i))_{i\in\mathbb{N}} \subset \tilde{\wp} : A \le \bigvee_{i=1}^{\infty} C_i,$$

$$(D_j(K_j, \beta_j))_{j\in\mathbb{N}} \subset \tilde{\wp} : B \le \bigvee_{j=1}^{\infty} D_j$$

and define a new T_M-disjoint sequence

$$(H_{ij}(L_{ij}, \gamma_{ij}))_{i,j\in\mathbb{N}}, \text{ where } L_{ij} = M_i \cap K_j \text{ and } \gamma_{ij} = \alpha_i \wedge \beta_i.$$

Obviously, $A \wedge B \le \bigvee_{i\in\mathbb{N}} \bigvee_{i\in\mathbb{N}} H_{ij}$.

As to L-fuzzy set $A \vee B$ we can cover it with elements of three sequences:

$$\begin{aligned} (\tilde{H}_{ij}(L_{ij}, \lambda_{ij}))_{i,j\in\mathbb{N}}, &\quad \text{with} \quad L_{ij} = M_i \cap K_j, \lambda_{ij} = \alpha_i \vee \beta_i, \\ (\tilde{C}_i(\tilde{M}_i, \alpha_i))_{i\in\mathbb{N}}, &\quad \text{with} \quad \tilde{M}_i = M_i \setminus \bigcup_{j\in\mathbb{N}} K_j, \\ (\tilde{D}_j(\tilde{K}_j, \beta_j))_{j\in\mathbb{N}}, &\quad \text{with} \quad \tilde{K}_j = K_j \setminus \bigcup_{i\in\mathbb{N}} M_i. \end{aligned}$$

Now let us transform the sum $\bigoplus_{i=1}^{\infty} \tilde{m}^*(C_i) \oplus \bigoplus_{j=1}^{\infty} \tilde{m}^*(D_j)$. We use the following notation:

$$F_{ij}^{\alpha} = (L_{ij}, \alpha_i), F_{ij}^{\beta} = (L_{ij}, \beta_j), ij \in \mathbb{N}.$$

Because of σ-additivity of measure ν we get

$$\nu(M_i) = \nu(\tilde{M}_i) + \sum_{j=1}^{\infty} \nu(L_{ij}), i \in \mathbb{N},$$

$$\nu(K_j) = \nu(\tilde{K}_j) + \sum_{i=1}^{\infty} \nu(L_{ij}), j \in \mathbb{N}.$$

It follows

$$\tilde{m}^*(C_i) = z_{\nu(M_i),\alpha_i} = z_{\nu(\tilde{M}_i) + \sum_{j=1}^{\infty} \nu(L_{ij}),\alpha_i} =$$

$$= z_{\nu(\tilde{M}_i),\alpha_i} \oplus \bigoplus_{j=1}^{\infty} z_{\nu(L_{ij}),\alpha_i} = \tilde{m}^*(\tilde{C}_i) \oplus \bigoplus_{j=1}^{\infty} \tilde{m}^*(F_{ij}^{\alpha}),$$

$$\tilde{m}^*(D_j) = z_{\nu(K_j),\beta_j} = z_{\nu(\tilde{K}_j) + \sum_{i=1}^{\infty} \nu(L_{ij}),\beta_j} =$$

$$= z_{\nu(\tilde{K}_j),\beta_j} \oplus \bigoplus_{i=1}^{\infty} z_{\nu(L_{ij}),\beta_j} = \tilde{m}^*(\tilde{D}_j) \oplus \bigoplus_{i=1}^{\infty} \tilde{m}^*(F_{ij}^{\beta}).$$

Now we obtain

$$\bigoplus_{i=1}^{\infty} \tilde{m}^*(C_i) \oplus \bigoplus_{j=1}^{\infty} \tilde{m}^*(D_j) =$$

$$= \bigoplus_{i=1}^{\infty} (\tilde{m}^*(\tilde{C}_i) \oplus \bigoplus_{j=1}^{\infty} \tilde{m}^*(F_{ij}^{\alpha})) \oplus \bigoplus_{j=1}^{\infty} (\tilde{m}^*(\tilde{D}_j) \oplus \bigoplus_{i=1}^{\infty} \tilde{m}^*(F_{ij}^{\beta})) =$$

$$= \bigoplus_{i=1}^{\infty} \tilde{m}^*(\tilde{C}_i) \oplus \bigoplus_{j=1}^{\infty} \tilde{m}^*(\tilde{D}_j) \oplus \bigoplus_{i=1}^{\infty} \bigoplus_{j=1}^{\infty} \tilde{m}^*(F_{ij}^{\alpha}) \oplus \bigoplus_{j=1}^{\infty} \bigoplus_{i=1}^{\infty} \tilde{m}^*(F_{ij}^{\beta}) =$$

$$= \bigoplus_{i=1}^{\infty} \tilde{m}^*(\tilde{C}_i) \oplus \bigoplus_{j=1}^{\infty} \tilde{m}^*(\tilde{D}_j) \oplus \bigoplus_{i=1}^{\infty} \bigoplus_{j=1}^{\infty} (\tilde{m}^*(F_{ij}^{\alpha}) \oplus \tilde{m}^*(F_{ij}^{\beta})).$$

Since

$$\tilde{m}^*(F_{ij}^{\alpha}) \oplus \tilde{m}^*(F_{ij}^{\beta}) = \tilde{m}^*(H_{ij}) \oplus \tilde{m}^*(\tilde{H}_{ij}),$$

we continue

$$\bigoplus_{i=1}^{\infty} \tilde{m}^*(\tilde{C}_i) \oplus \bigoplus_{j=1}^{\infty} \tilde{m}^*(\tilde{D}_j) \oplus \bigoplus_{i=1}^{\infty} \bigoplus_{j=1}^{\infty} (\tilde{m}^*(F_{ij}^{\alpha}) \oplus \tilde{m}^*(F_{ij}^{\beta})) =$$

$$= \bigoplus_{i=1}^{\infty} \tilde{m}^*(\tilde{C}_i) \oplus \bigoplus_{j=1}^{\infty} \tilde{m}^*(\tilde{D}_j) \oplus \bigoplus_{i=1}^{\infty} \bigoplus_{j=1}^{\infty} \tilde{m}^*(\tilde{H}_{ij}) \oplus \bigoplus_{i=1}^{\infty} \bigoplus_{j=1}^{\infty} \tilde{m}^*(H_{ij}).$$

Now taking into account the fact that

$$A \wedge B \le \bigvee_{i=1}^{\infty} \bigvee_{j=1}^{\infty} H_{ij} \text{ and } A \vee B \le (\bigvee_{i=1}^{\infty} C_i) \vee (\bigvee_{j=1}^{\infty} D_j) \vee (\bigvee_{i=1}^{\infty} \bigvee_{j=1}^{\infty} \tilde{H}_{ij}),$$

we get

$$\bigoplus_{i=1}^{\infty} \tilde{m}^*(\tilde{C}_i) \oplus \bigoplus_{j=1}^{\infty} \tilde{m}^*(\tilde{D}_j) \oplus \bigoplus_{i=1}^{\infty} \bigoplus_{j=1}^{\infty} \tilde{m}^*(\tilde{H}_{ij}) \oplus \bigoplus_{i=1}^{\infty} \bigoplus_{j=1}^{\infty} \tilde{m}^*(H_{ij}) \ge$$

$$\ge \tilde{m}^*(A \wedge B) \oplus \tilde{m}^*(A \vee B).$$

So independent of the choice of sequences $(C_i)_{i\in\mathbb{N}}$, $(D_j)_{j\in\mathbb{N}}$ it holds

$$\bigoplus_{i=1}^{\infty} \tilde{m}^*(C_i) \oplus \bigoplus_{j=1}^{\infty} \tilde{m}^*(D_j) \ge \tilde{m}^*(A \wedge B) \oplus \tilde{m}^*(A \vee B).$$

Finally, by taking infimums of sums we get

$$\tilde{m}^*(A) \oplus \tilde{m}^*(B) \ge \tilde{m}^*(A \wedge B) \oplus \tilde{m}^*(A \vee B).$$

□

ISBN: 978-989-95079-6-8

4.2 $\tilde{m}^*$-measurable L-fuzzy sets

We generalize the concept of measurability in the sense of Caratheodory (see e.g. [8]).

Definition 4.2. *A set $A \in L^X$ is called an $\tilde{m}^*$-measurable L-fuzzy set, if it satisfies the following conditions for all sets $E \in L^X$:*

(i) $\tilde{m}^*(A) \oplus \tilde{m}^*(E) = \tilde{m}^*(A \wedge E) \oplus \tilde{m}^*(A \vee E)$,

(ii) $\tilde{m}^*(N(A)) \oplus \tilde{m}^*(E) = \tilde{m}^*(N(A) \wedge E) \oplus \tilde{m}^*(N(A) \vee E)$.

We denote $\mathcal{A}_{\tilde{m}^*}$ the class of all $\tilde{m}^*$-measurable L-fuzzy sets. Some obvious properties of $\mathcal{A}_{\tilde{m}^*}$:

$\emptyset, X \in \mathcal{A}_{\tilde{m}^*}$,

$A \in \mathcal{A}_{\tilde{m}^*} \Longrightarrow N(A) \in \mathcal{A}_{\tilde{m}^*}$.

Theorem 4.3. *Class $\mathcal{A}_{\tilde{m}^*}$ is a T_M-clan.*

Proof. For given L-fuzzy sets $A_1, A_2 \in \mathcal{A}_{\tilde{m}^*}$ and $E \in L^X$ we will prove the equality

$$\begin{aligned} &\tilde{m}^*(E) \oplus \tilde{m}^*(A_1 \wedge A_2) = \\ &\tilde{m}^*(E \wedge (A_1 \wedge A_2)) \oplus \tilde{m}^*(E \vee (A_1 \wedge A_2)) \end{aligned} \tag{1}$$

Since A_1, A_2 are $\tilde{m}^*$-measurable we have

$$\tilde{m}^*(A_1) \oplus \tilde{m}^*(A_2) = \tilde{m}^*(A_1 \wedge A_2) \oplus \tilde{m}^*(A_1 \vee A_2),$$

$$\tilde{m}^*(A_1 \wedge E) \oplus \tilde{m}^*(A_1 \vee E) = \tilde{m}^*(A_1) \oplus \tilde{m}^*(E).$$

Now by summing up these two equalities and adding one more additional summand $\tilde{m}^*(A_1 \wedge (E \vee A_2))$ to both sides of the equality we obtain

$$\begin{aligned} &\tilde{m}^*(A_1) \oplus \tilde{m}^*(A_2) \oplus \tilde{m}^*(A_1 \wedge E) \oplus \\ &\oplus \tilde{m}^*(A_1 \vee E) \oplus \tilde{m}^*(A_1 \wedge (E \vee A_2)) = \\ &= \tilde{m}^*(A_1 \wedge A_2) \oplus \tilde{m}^*(A_1 \vee A_2) \oplus \tilde{m}^*(A_1) \oplus \\ &\oplus \tilde{m}^*(E) \oplus \tilde{m}^*(A_1 \wedge (E \vee A_2)). \end{aligned} \tag{2}$$

Let us transform now the left part of (2). To do this we use (3) and (4):

$$\begin{aligned} &\tilde{m}^*(A_1) \oplus [\tilde{m}^*(A_2) \oplus \tilde{m}^*(E \wedge A_1)] = \\ &= \tilde{m}^*(A_1) \oplus [\tilde{m}^*(E \wedge A_1 \wedge A_2) \oplus \tilde{m}^*((E \wedge A_1) \vee A_2)] = \\ &= \tilde{m}^*(E \wedge A_1 \wedge A_2) \oplus \tilde{m}^*((E \wedge A_1) \vee A_2 \vee A_1) \oplus \\ &\oplus \tilde{m}^*(((E \wedge A_1) \vee A_2) \wedge A_1) = \\ &= \tilde{m}^*(E \wedge A_1 \wedge A_2) \oplus \tilde{m}^*(A_1 \vee A_2) \oplus \tilde{m}^*((E \vee A_2) \wedge A_1) \end{aligned} \tag{3}$$

and

$$\begin{aligned} &\tilde{m}^*(E \vee A_1) \oplus \tilde{m}(A_1 \wedge (E \vee A_2)) = \\ &= \tilde{m}^*(E \vee (A_1 \wedge A_2) \vee A_1) \oplus \tilde{m}^*(E \vee A_1 \wedge A_2 \wedge A_1) = \\ &= \tilde{m}^*(E \vee (A_1 \wedge A_2)) \oplus \tilde{m}^*(A_1). \end{aligned} \tag{4}$$

Next we substitute (3) and (4) in (2).

$$\tilde{m}^*(E \wedge A_1 \wedge A_2) \oplus \tilde{m}^*(A_1 \vee A_2) \oplus \tilde{m}^*(A_1 \wedge (E \vee A_2)) \oplus$$

$$\oplus \tilde{m}^*(E \vee (A_1 \wedge A_2)) \oplus \tilde{m}^*(A_1) =$$

$$= \tilde{m}^*(A_1 \wedge A_2) \oplus \tilde{m}^*(A_1 \vee A_2) \oplus \tilde{m}^*(A_1) \oplus \tilde{m}^*(E) \oplus$$

$$\oplus \tilde{m}^*(A_1 \wedge (E \vee A_2)).$$

Finally, canceling like terms we obtain (1).
By analogy it can be proved that

$$\tilde{m}^*(E) \oplus \tilde{m}^*(A_1 \vee A_2) = \tilde{m}^*(E \wedge (A_1 \vee A_2)) \oplus \tilde{m}^*(E \vee (A_1 \vee A_2)).$$

Taking into account that $N(A_1 \vee A_2) = N(A_1) \wedge N(A_2)$ and $N(A_1 \wedge A_2) = N(A_1) \vee N(A_2)$, we get that $A_1 \vee A_2$ and $A_1 \wedge A_2$ are $\tilde{m}^*$-measurable. □

Proposition 4.4. $\tilde{\wp} \subset \mathcal{A}_{\tilde{m}^*}$

Proof. First we will show that for a given $A(H, \alpha) \in \tilde{\wp}$ and all $E \in L^X$ it holds

$$\tilde{m}^*(A) \oplus \tilde{m}^*(E) = \tilde{m}^*(A \vee E) \oplus \tilde{m}^*(A \wedge E).$$

Because of Proposition 4.1 it is sufficient to prove the inequality

$$\tilde{m}^*(A) \oplus \tilde{m}^*(E) \leq \tilde{m}^*(A \vee E) \oplus \tilde{m}^*(A \wedge E).$$

We consider two T_M-disjoint sequences $(C_k(M_k, \beta_k))_{k \in \mathbb{N}} \subset \tilde{\wp}$ and $(D_k(K_k, \gamma_k))_{k \in \mathbb{N}} \subset \tilde{\wp}$ such that

$$A \vee E \leq \bigvee_{k=1}^{\infty} C_k \text{ and } A \wedge E \leq \bigvee_{k=1}^{\infty} D_k$$

and define two the new sequences

$$(F_k(M_k \setminus H, \beta_k))_{k \in \mathbb{N}} \text{ and } (G_k(H \cap M_k, \beta_k))_{k \in \mathbb{N}}.$$

By T_M-additivity of $\tilde{m}$ it holds

$$\tilde{m}(G_k) \oplus \tilde{m}(F_k) = \tilde{m}(C_k), k \in \mathbb{N}.$$

Taking into account

$$A \leq \bigvee_{k=1}^{\infty} G_k \text{ and } E \leq (\bigvee_{k=1}^{\infty} F_k) \vee (\bigvee_{k=1}^{\infty} G_k),$$

we obtain

$$\tilde{m}^*(A) = \tilde{m}(A) \leq \bigoplus_{k=1}^{\infty} \tilde{m}(G_k),$$

$$\tilde{m}^*(E) \leq \bigoplus_{k=1}^{\infty} \tilde{m}(F_k) \oplus \bigoplus_{k=1}^{\infty} \tilde{m}(D_k).$$

Summing up

$$\tilde{m}^*(A) \oplus \tilde{m}^*(E) \leq \bigoplus_{k=1}^{\infty} \tilde{m}(G_k) \oplus \bigoplus_{k=1}^{\infty} \tilde{m}(F_k) \oplus \bigoplus_{k=1}^{\infty} \tilde{m}(D_k) =$$

$$= \bigoplus_{k=1}^{\infty} (\tilde{m}(G_k) \oplus \tilde{m}(F_k)) \oplus \bigoplus_{k=1}^{\infty} \tilde{m}(D_k) = \bigoplus_{k=1}^{\infty} \tilde{m}(C_k) \oplus \bigoplus_{k=1}^{\infty} \tilde{m}(D_k).$$

So independent of a choice of sequences $(C_k)_{k \in \mathbb{N}}$ and $(D_k)_{k \in \mathbb{N}}$ it holds

$$\bigoplus_{k=1}^{\infty} \tilde{m}(C_k) \oplus \bigoplus_{k=1}^{\infty} \tilde{m}(D_k) \geq \tilde{m}^*(A) \oplus \tilde{m}^*(E).$$

Finally, by taking infimums of sums we get

$$\tilde{m}^*(A) \oplus \tilde{m}^*(E) \leq \tilde{m}^*(A \vee E) \oplus \tilde{m}^*(A \wedge E).$$

Since

$$N(A) = B_1(H, N_\alpha) \vee B_2(X \setminus H, 1_L),$$

where $B_1(H, N_\alpha), B_2(X \setminus H, 1_L) \in \tilde{\wp}$, the equality

$$\tilde{m}^*(N(A)) \oplus \tilde{m}^*(E) = \tilde{m}^*(N(A) \wedge E) \oplus \tilde{m}^*(N(A) \vee E)$$

can be proved by analogy with the proof of Theorem 4.3. □

ISBN: 978-989-95079-6-8

5 A construction of an L-fuzzy valued T_M-countably additive measure

Let us denote by $\tilde{\mu}$ the restriction of $\tilde{m}^*$ to $\mathcal{A}_{\tilde{m}^*}$:

$$\tilde{\mu} : \mathcal{A}_{\tilde{m}^*} \to \mathbb{R}_+(L).$$

Theorem 5.1. *$\tilde{\mu}$ is T_M-countably additive L-fuzzy valued function.*

Proof. First we will show that for a T_M - disjoint sequence $A_1, A_2, ..., A_n \in \mathcal{A}_{\tilde{m}^*}$ it holds $\tilde{\mu}(\bigvee_{i=1}^{n} A_i) = \bigoplus_{i=1}^{n} \tilde{\mu}(A_i)$.

For $n = 2$ by usinq $\tilde{m}^*$-measurability and T_M - disjointness of A_1, A_2 we obtain

$$\tilde{\mu}(A_1) \oplus \tilde{\mu}(A_2) = \tilde{\mu}(A_1 \vee A_2) \oplus \tilde{\mu}(A_1 \wedge A_2) =$$
$$= \tilde{\mu}(A_1 \vee A_2) \oplus \tilde{\mu}(\emptyset) = \tilde{\mu}(A_1 \vee A_2).$$

Now we assume that the equality holds for a given $n \in \mathbb{N}$ and we prove it for $n+1$ T_M-disjoint sets $A_1, A_2, ..., A_{n+1} \in \mathcal{A}_{\tilde{m}^*}$

$$\tilde{\mu}(\bigvee_{k=1}^{n+1} A_k) = \tilde{\mu}((\bigvee_{k=1}^{n} A_k) \vee A_{n+1}) = \tilde{\mu}(\bigvee_{k=1}^{n} A_k) \oplus \tilde{\mu}(A_{n+1}) =$$
$$= \bigoplus_{k=1}^{n} \tilde{\mu}(A_k) + \tilde{\mu}(A_{n+1}) = \bigoplus_{k=1}^{n+1} \tilde{\mu}(A_k).$$

Now let us consider a T_M-disjoint sequence $(A_n)_{n\in\mathbb{N}} \subset \mathcal{A}_{\tilde{m}^*}$ such that $\bigvee_{n=1}^{\infty} A_n \in \mathcal{A}_{\tilde{m}^*}$. To prove the inequality

$$\tilde{\mu}(\bigvee_{n=1}^{\infty} A_n) \leq \bigoplus_{n=1}^{\infty} \tilde{\mu}(A_n)$$

we take T_M-disjoint sequences $(B_k^n)_{k\in\mathbb{N}} \in \tilde{\wp}$ such that $A_n \leq \bigvee_{k=1}^{\infty} B_k^n, n \in \mathbb{N}$. Then $\bigvee_{n=1}^{\infty} A_n \leq \bigvee_{n=1}^{\infty} \bigvee_{k=1}^{\infty} B_k^n$ and hence

$$\tilde{m}^*(\bigvee_{n=1}^{\infty} A_n)) \leq \bigoplus_{n=1}^{\infty} \bigoplus_{k=1}^{\infty} \tilde{m}(B_k^n).$$

Taking into account that this inequality holds independent of the choice of sequences $(B_k^n)_{k\in\mathbb{N}}$ we obtain

$$\tilde{m}^*(\bigvee_{n=1}^{\infty} A_n) \leq \bigoplus_{n=1}^{\infty} \tilde{m}^*(A_n).$$

Now we will prove inverse inequality. For a given $n \in \mathbb{N}$ we have

$$\tilde{\mu}(\bigvee_{k=1}^{n} A_k) \leq \tilde{\mu}(\bigvee_{k=1}^{\infty} Ak).$$

and

$$\tilde{\mu}(\bigvee_{k=1}^{n} A_k)) = \bigoplus_{k=1}^{n} \tilde{\mu}(A_k),$$

hence

$$\bigoplus_{k=1}^{n} \tilde{\mu}(A_k) \leq \tilde{\mu}(\bigvee_{k=1}^{\infty} A_k).$$

Finally by taking supremum we obtain

$$\bigoplus_{k=1}^{\infty} \tilde{\mu}(A_k) \leq \tilde{\mu}(\bigvee_{k=1}^{\infty} A_k).$$

□

As the final result, by extension of a crisp measure ν we obtain L-fuzzy valued T_M-countably additive measure $\tilde{\mu} : \mathcal{A}_{\tilde{m}^*} \to \mathbb{R}_+(L)$ such that

(i) $\tilde{\mu}/\tilde{\wp} = \tilde{m}$;

(ii) $\tilde{\mu}/\Phi = \nu$.

The last equality means that for every $M \in \Phi$ it holds $\tilde{\mu}(A(M, 1_L)) = z_{\nu(M)}$.

References

[1] Ruzha V., Asmuss S., *A Construction of a Fuzzy Valued Measure Based on Minimum t-norm.* In: New Dimensions in Fuzzy Logic and Related Technologies. Proceedings of the 5th EUSFLAT Conference, Ostrava, 2007, 175-178.

[2] Klement E. P., Butnariu D., *Triangular norm based measures.* In: Endre Pap (Hrsg.), *"Handbook of Measure Theory"*, Kapitel 23, Seite(n) 947-1010, Elsevier, Amsterdam, 2002.

[3] Hutton B., *Normality in fuzzy topological spaces*, J.Math.Anal.Appl., 50 (1975), 74-79.

[4] Gantner T.E., Steinlage R.C., Warren R.H. *Compactness in fuzzy topological spaces*, J.Math.Anal.Appl., 62 (1978), 547-562.

[5] Rodabaugh S.E., *Fuzzy addition and the L-fuzzy real line*, Fuzzy Sets and Systems, 8 (1982), 39-52.

[6] Asmuss S., Šostak A., *Extremal problems of approximation theory in fuzzy context*, Fuzzy Sets and Systems, 105 (1999), 249-257.

[7] Asmuss S., Šostak A., *On central algorithms of approximation under fuzzy information*, Fuzzy Sets and Systems, 155 (2005), 150-163.

[8] Halmosh P., *Measure theory*, Moscow, IL (1953) (in russian).

ISBN: 978-989-95079-6-8

Distributed Genetic Tuning of Fuzzy Rule-Based Systems

Ignacio Robles, Rafael Alcalá, José Manuel Benítez and Francisco Herrera

Department of Computer Science and Artificial Intelligence
University of Granada, Spain
Email: ignaciorobles@gmail.com, {alcala, J.M.Benitez, herrera}@decsai.ugr.es

***Abstract*— *The tuning of Fuzzy Rule Base-Systems is necessary to improve their performance after the extraction of rules. This optimization problem can become a hard one when the size of the considered system in terms of the number of variables, rules and data samples is big. To alleviate this growth in complexity, we propose a distributed genetic algorithm which explotes the nowadays available parallel hardware (multicore microprocessors and clusters). The empirical performance in solution quality and computing time is assessed by comparing its results with those from a highly effective sequential tuning algorithm. Both methods are applied for the modeling of four well-known regression problems.**

***Keywords*—** 2-tuples, Distributed Genetic Algorithms, Fuzzy Rule-based Systems Tuning, Lateral Tuning

1 Introduction

Fuzzy rule based-systems (FRBS) have become a wide choice when addressing modeling and system identification problems. An essential component in these systems is the fuzzy rule base, together with the data base compose the knowledge base. The construction of this knowledge base is a key step for obtaining a correct system.

It is very difficult for human beings to obtain appropiate rules when dealing with real-world complex problems with many variables and where the necessary number of rules is high. When an expert determines the rule set for a determined problem, generally it will not be the optimal set in terms of performance. Performance is an important goal design for whatever the system is intended to be used: either with approximation purposes or when searching for an interpretable system. To cope with this problem a refining process that adjusts the system is required. This process is widely known as *tuning*.

Classically, tuning processes involve changing the shape of the Membership Functions (MFs) associated to the labels in the database so that the best cooperation among rules is reached. However as the number of variables and rules increases, tuning methods show poor performance due to the growing complexity of the search space. Moreover, the comptuing time consumed by these approaches grows with the complexity of the search space which would result in a procedures that are not useful in practice.

Nowadays parallel hardware and software has become very affordable. They are broadly available which makes them perfect to deal with complex search spaces in order to improve the poor performance achieved with classical tuning approaches. Clear examples in this line are multicore processors and linux clusters.

In this paper, we address the tuning problem and present a distributed method for lateral FRBS tuning. The paper is structured as follows: in the second section of the paper the lateral tuning of FRBSs problem is stated and an efficient sequential specialized algorithm is reviewed. The third section describe our proposal for the distributed tuning for FRBS. An empirical evaluation of the distributed algorithm is presented in the fourth section. Finally some conclusions and future work is commented in the last section.

2 Lateral Tuning of FRBSs

The tuning of FRBS is a problem long studied by researchers in the community [1, 2]. A quite efficient procedure fuzzy systems tuning was presented recently by Alcalá *et al.* [3]. We choose that as a reference to compare the perfomance of our proposal with. We describe this method briefly in this section.

2.1 Lateral Tuning: The Linguistic 2-Tuples Representation

In [3], a new procedure for FRBSs tuning was proposed. It is based on the linguistic 2-tuples representation scheme introduced in [4], which allows the lateral displacement of the support of a label and maintains the interpretability at a good level. This proposal introduces a new model for rule representation based on the concept of symbolic translation [4]. The symbolic translation of a label is a number in [-0.5, 0.5), expressing this interval the domain of a label when it is moving between its two adjacent lateral labels (see Figure 1.a). Let us consider a generic linguistic fuzzy partition $S = s_0, \ldots, s_{L-1}\}$ (with L representing the number of labels). Formally, we represent the symbolic translation of a label s_i in S by means of the 2-tuple notation,

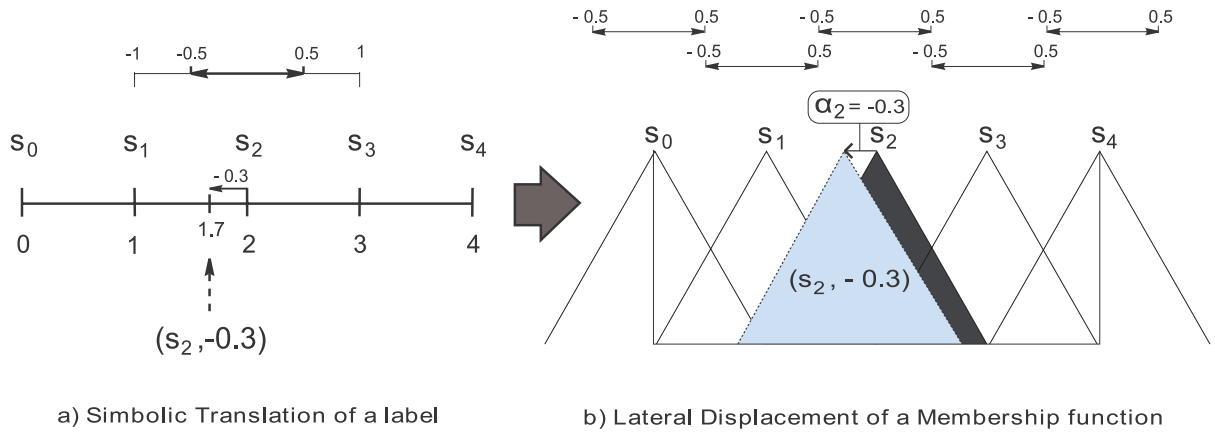
$$(s_i, \alpha_i), \;\; s_i \in S, \; \alpha_i \in [-0.5, 0.5).$$

The symbolic translation of a label involves the lateral variation of its associated MF. Figure 1 shows the symbolic translation of a label represented by the 2-tuple $(s_2, -0.3)$ together with the associated lateral variation.

In the context of FRBSs, the linguistic 2-tuples could be used to represent the MFs comprising the linguistic rules. This way to work, introduces a new model for rule representation that allows the tuning of the MFs by learning their respective lateral displacements. With respect to the classic tuning, usually considering three parameters in the case of triangular MFs, this way to work involves a reduction of the search space that eases a fast derivation of optimal models, improving the convergence speed and avoiding the necessity of a large number of evaluations.

In [3], two different rule representation approaches have been proposed, a global approach and a local approach. The global approach tries to obtain more interpretable models,

ISBN: 978-989-95079-6-8



a) Simbolic Translation of a label

b) Lateral Displacement of a Membership function

Figure 1: Symbolic Translation of a Label and Lateral Displacement of the associated MF

while the local approach tries to obtain more accurate ones. In our case, tuning is applied at the level of linguistic partitions (global approach). By considering this approach, the label s_i^v of a variable v is translated with the same α_i^v value in all the rules where it is considered, i.e., a global collection of 2-tuples is used in all the fuzzy rules. Notice that from the parameters α_i^v applied to each label we could obtain the equivalent triangular MFs. Thus, an FRBS based on linguistic 2-tuples can be represented as a classic Mamdani FRBS [5]. Refer to [3] for further details on this approach.

2.2 Sequential Algorithm for the Lateral Tuning of FRBSs

In [3], a sequential genetic algorithm was proposed to perform a lateral tuning of previously obtained FRBSs. A short description of this algorithm is given in the next (see [3] for a detailed description).

The used model was the genetic model of CHC [6]. CHC makes use of a "Population-based Selection" approach. N parents and their corresponding offsprings are combined to select the best N individuals to take part of the next population. The CHC approach makes use of an incest prevention mechanism and a restarting process to provoke diversity in the population, instead of the well known mutation operator.

This incest prevention mechanism is considered in order to apply the crossover operator, i.e., two parents are crossed if their hamming distance divided by 2 is over a predetermined threshold, T. Since a real coding scheme is considered, each gene is transformed by considering a Gray Code with a fixed number of bits per gene ($BITSGENE$) determined by the system expert. In our case, the threshold value is initialized as:

$$T = (\#GenesC_T * BITSGENE)/4.0.$$

Following the original CHC scheme, T is decreased by one when the population does not change in one generation. In order to avoid very slow convergence, T is also decreased by one when no improvement is achieved with respect to the best chromosome of the previous generation. The algorithm restarts when T is below zero.

In the following, the components used to design the evolutionary tuning process are explained. They are: DB codification, chromosome evaluation and genetic operators.

2.3 DB Codification

A real coding scheme is considered, i.e., the real parameters are the GA representation units (genes). Let us consider n system variables and a fixed number of labels per variable L. Then, a chromosome has the following form (where each gene is associated to the tuning value of the corresponding label),

$$(\alpha_1^1, \ldots, \alpha_1^L, \alpha_2^1, \ldots, \alpha_2^L, \ldots, \alpha_n^1, \ldots, \alpha_n^L)$$

To make use of the available information, the initial FRBS obtained from an automatic fuzzy rule learning method is included in the population as an initial solution. To do so, the initial pool is obtained with the first individual having all genes with value '0.0', and the remaining individuals generated at random in [-0.5, 0.5).

2.4 Chromosome Evaluation

To evaluate a determined chromosome the well-known Mean Square Error (MSE) is used:

$$\text{MSE} = \frac{1}{2 \cdot N} \sum_{l=1}^{E} (F(x^l) - y^l)^2,$$

with E being the data set size, $F(x^l)$ being the output obtained from the FRBS decoded from the said chromosome when the l-th example is considered and y^l being the known desired output.

2.5 Genetic Operators

The genetic operators considered in CHC are crossover and restarting approach (no mutation is used). A short description of these operators comes next:

- Crossover. The crossover operator is based on the the concept of environments. These kinds of operators show a good behavior in real coding. Particularly, the Parent Centric BLX (PCBLX) operator [7] (an operator based on BLX-α) is considered.
- Restarting. To get away from local optima, this algorithm uses a restart approach [6]. In this case, the best chromosome is maintained and the remaining are generated at random within the corresponding variation intervals [-0.5, 0.5). It follows the principles of CHC [6], performing the restart procedure when threshold L is below zero.

ISBN: 978-989-95079-6-8

3 A distributed genetic algorithm for FRBS tuning

The availability of extremely fast and low cost parallel hardware in the last few years benefits the investigation on new approaches to existing optimization algorithms. The key to these new approaches is achieving not only gains in time, which is somehow inherent to distributed algorithms, but gains in quality of the solutions found.

Distributed Genetic Algorithms (DGA) are excellent optimization algorithms and have proven to be one of the best options when trying to cope with large scale problems and when the classic approaches take too much time to give a proper solution.

One procedure for the parallelization of GA comes from the consideration of spatial separation of populates. Schematically:

```
1. Generate a random population, P.
2. Divide P into n subpopulations: SPi,
   i = 1, ..., n.
3. Define a topology for SP1, ..., SPn.
4. For i = 1 to n do:
4.1. Apply in parallel during FM
     generations the genetic operators.
4.2. Send in parallel NM chromosomes to
     neighbour subpopulations.
4.3. Receive in parallel chromosomes
     from neigbour subpopulations.
5. If stopping criteria is not meet then
   go back to step 4.
```

In this section the DGA [8] used for Fuzzy Rule-based System tuning is described.

3.1 Gradual Distributed Real-Coded Genetic Algorithm used

Gradual Distributed Real-Coded Genetic Algorithms (GDRCGAs) are a kind of heterogeneous DGAs based on real coding where subpopulations apply genetic operators in different levels of exploitation/exploration. This heterogeneous application of genetic operators produce a *parallel multiresolution* which allows a wide exploration of the search space and effective local precision. Due to appropiate connections between subpopulations in order to gradually exploit multiresolution,these algorithms achieve refinement or expansion of the best emerging zones of the search space.

The GDRCGA [8] used for FRBS tuning employs 8 subpopulations in a hypercube topology as seen in Figure 2.

In this topology two important groups of subpopulations can be clearly identified:

1. **Front side**: this side of the hypercube is oriented to explore the search space. In this side, four subpopulations, $E_1, ..., E_4$, apply exploration tuned genetic operators in a clockwise increasing degree.

2. **Back side**: subpopulations in the back side of the hypercube, $e_1, ..., e_4$, apply exploitation oriented genetic operators in a clockwise increasing degree.

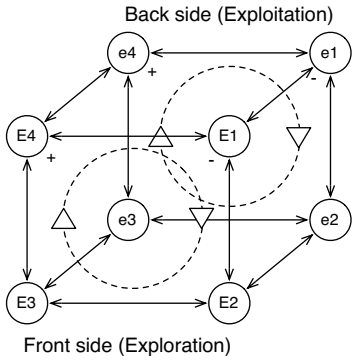


Figure 2: Hypercube topology for GDRCGA

One of the key elements of DGAs is the migration policy of individuals between subpopulations. In this particular model, a *immigration* process [9] is achieved when the best chromosome in every subpopulation abandons it and moves to an immediate neighbour. Due to this immigration policy, three different immigration movements can be identified depending on the subpopulations involved:

1. **Refinement migrations**: individuals in the back side move clockwise to the immediate neighbor, i.e. from e_2 to e_3. Chromosomes in the front side move counterclock from a more exploratory subpopulation to a less exploratory oriented one.

2. **Expansion migrations**: individuals in the back side move counterclock to the immediate neighbor and chromosomes in the front side move clockwise from a less exploratory subpopulation to a more exploratory oriented one, i.e. from E_4 to E_1.

3. **Mixed migrations**: subpopulations from one side of the hypercube exchange their best individual with the counterpart subpopulation in the other side: interchange between E_i and e_i, $i = 1 \dots 4$.

Figure 3 shows the three different migration movements described above.

The genetic operators used in the distributed model are:

- **Selection mechanism**: linear ranking selection (LRS) [10] with stochastic universal sampling [11]. Values of LRS parameter is shown if Table 1.

- **Crossover operator**: BLX-α [12] operator using values for α shown in Table 2.

- **Mutation operator**: non-uniform mutation operator applied with probability $P_{mut} = 0,125$.

Table 1: LRS parameter values for each subpopulation

Exploitation				Exploration			
$+$	$\leftarrow$		$-$	$-$	$\rightarrow$		$+$
e_4	e_3	e_2	e_1	E_1	E_2	E_3	E_4
0,9	0,7	0,5	0,1	0,9	0,7	0,5	0,1

ISBN: 978-989-95079-6-8

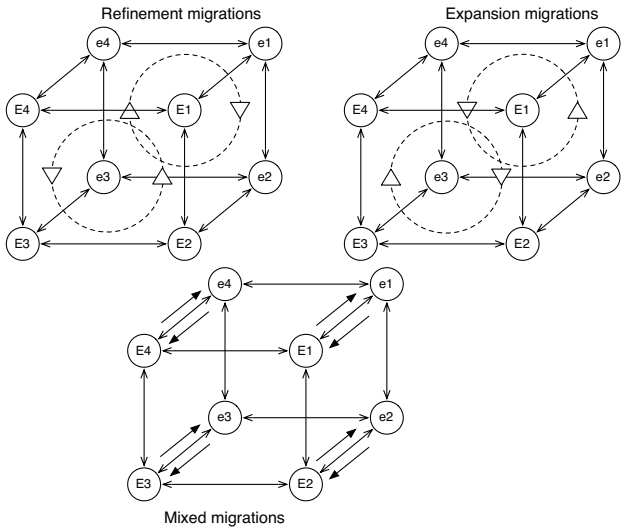


Figure 3: Three different migration movements

Table 2: Values of α for each subpopulation

Exploitation				Exploration			
$+$	$\leftarrow$		$-$	$-$	$\rightarrow$		$+$
e_4	e_3	e_2	e_1	E_1	E_2	E_3	E_4
0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8

As stated in [8], the frequency in which migration movements occur is crucial to avoid the classic withdraws of DGAs: the conquest and noneffect problems. In order to reduce the negative effect of these problems, immigrants stay in the receiving subpopulations for a brief number of generations. Besides, a restart operator is used to avoid stagnation of the search process. This restart operator randomly reinitializes all subpopulations if non-significant improvement of the best element is achieved for a number of generations. Also an elitism strategy is used in order to keep the best adapted individual of every subpopulation.

4 Empirical evaluation

In this section we describe the empirical evaluation we performed in order to assess the merits of our proposal.

We consider the tuning of FRBSs constructed for the modeling of some well-known regression problems. We have compared the quality of the tuning performed by the sequential algorithm and our proposal. We have selected four problems, some of their salient features are shown in Table 3 displayed in increasing complexity order.

Table 3: Data sets used to evaluate the algorithm

Data set	Variables	Instances
Electrical Maintenance	5	1056
Trasury	16	1049
Weather-Izmir	10	1461
Abalone	8	4177

A five-fold cross-validation approach has been used. So five runs with different independent test sets have been carried out for each problem. The performance of a fuzzy rule is measured as mean squared error (MSE) over the test set.

Our proposed distributed algorithm is compared with a specialized sequential genetic algorithm (CHC) [3] in terms of quality of the solutions achieved (MSE) as well as in time. As a starting point we have used the results achieved by applying Wang & Mendel fuzzy rule learning method [13]. These initial results are shown in Table 4.

Table 4: Initial results using Wang & Mendel fuzzy rule learning method

Dataset	Training	σ_{tra}	Test	σ_{test}
Electrical M.	57605.83	2840.78	57934.25	4732.66
Treasury	1.636	0.121	1.632	0.182
Weather-Izmir	6.944	0.720	7.368	0.909
Abalone	3.341	0.130	3.474	0.247

Generally when comparing a distributed or parallel approach with some other sequential algorithm an interesting measure is the execution time gain ratio. This ratio could be defined as follows:

$$R = \frac{T_{seq}}{T_{dist}} \quad (1)$$

where T_{seq} is the time spent by the sequential algorithm and T_{dist} is the execution time of the distributed approach. The higher the value of R, the better. Time gain ratio values obtained in the empirical experimentation are shown in Table 5.

An interesting point to which to pay attention is the *evolution* of the mean squared error as the number of evaluations increases. So, three different numbers of evaluations have been choosen: 10000, 25000 and 50000 evaluations per run. The results in terms of quality of the solutions attained are shown in Table 6. An important fact to notice is that the mean squared error achieved with the distributed method is lower than the error obtained with the specialized genetic algorithm in all data sets at 50000 evaluations. The distribuited also obtains good results with fewer iterations in some cases (e.g. Electrical Maintenance, 25000 it.), but clearly its real effectiveness will be reached when the computation load is higher. Actually, we are running further experiments with datasets of higher complexity.

Table 5: Time gain ratio with 50000 evaluations

Data set	T_{seq}	T_{dist}	R
Electrical Maintenance	187,3	391,6	0,479
Trasury	525,3	739,7	0,710
Weather-Izmir	849,8	867,1	0,980
Abalone	1980,9	942,5	2,101

As shown in Table 5, the time gain ratio, R, increases with the problem complexity. In the less complex data sets the ratio obtained is substantially low because the sequential specialized genetic algorithm is very fast and the time spent in communications of the distributed approach slows it down in comparision. As the complexity of the data set increases the time gain ratio also increases, showing that the distributed approach in the most complex data set is more than two times faster than the sequential specialized genetic algorithm.

ISBN: 978-989-95079-6-8

Table 6: Mean squared errors in training and test sets. The winner for each pair of training is *in italics*. The winner for each pair of test is **boldfaced**

		CHC		GDRCGA	
Data set	Evaluations	Training	Test	Training	Test
Electrical Maintenance	10000	*2.59363671E+04*	2.92591821E+04	2.65539710E+04	**2.89024830E+04**
	25000	2.48690100E+04	2.80510895E+04	*2.39248797E+04*	**2.67720415E+04**
	50000	2.46214328E+04	2.78282761E+04	*2.26682075E+04*	**2.54097540E+04**
Abalone	10000	*2.61355003E+00*	2.79981355E+00	2.65916770E+00	**2.79026550E+00**
	25000	2.60333453E+00	2.79298130E+00	*2.59992700E+00*	**2.76143590E+00**
	50000	2.60303744E+00	2.79117626E+00	*2.57035010E+00*	**2.75904570E+00**
Weather-Izmir	10000	*1.68875432E+00*	**1.89318352E+00**	1.89195950E+00	1.95458830E+00
	25000	*1.64117336E+00*	**1.86996710E+00**	1.66238700E+00	1.87669540E+00
	50000	1.64010963E+00	1.86891124E+00	*1.57019250E+00*	**1.86195430E+00**
Treasury	10000	*1.71238672E-01*	**1.86722425E-01**	2.12486800E-01	2.16882700E-01
	25000	*1.33618274E-01*	**1.50895419E-01**	1.42194200E-01	1.67407400E-01
	50000	1.20604483E-01	1.37784224E-01	*1.15845500E-01*	**1.31803000E-01**

5 Conclusions and final remarks

We have developed and presented a very promising distributed algorithm for FRBSs lateral tuning that achieves better results than a specialized genetic algorithm. Due to its distributed nature and consequently the spatial separation implied, it needs more evaluations to converge than a classic sequential algorithm. It always presents the same behaviour in comparision to the specialized sequential algorithm: with a small number of evaluations it offers a higher error than the sequential approach but when the evaluations are high it gives better quality solutions.

Empirical results show that when the complexity of the problem grows, our distributed method takes adventage of the large computing times and converges to a better solution in less time. This property makes the distributed tuning algorithm very convenient when dealing with large scale data sets.

The distributed algorithm takes longer than the sequential algorithm when dealing with small size data sets mainly due to two reasons: interprocess communication in the distributed approach implies additional execution time which can not be paralelized and the specialized algorithm is optimized for small size data sets where the search space is not too complex.

Since execution time and quality of the results are two properties always in conflict somehow, our approach is very convenient since it can be graduated in order to achieve faster execution times with a small cost in quality and viceversa.

We expect to achieve great execution time gains when applying the distributed FRBSs tuning algorithm with large scale data sets with a non significant loss of quality.

Acknowledgment

Supported by the Spanish Ministry of Education and Science under grant no. TIN2008-06681-C06-01.

References

[1] R. Alcalá, J. Alcalá-Fdez, M. J. Gacto, and F. Herrera. Rule base reduction and genetic tuning of fuzzy systems based on the linguistic 3-tuples representation. *Soft Computing*, 11(5):401–419, 2007.

[2] F. Herrera, M. Lozano, and J. L. Verdegay. Tuning fuzzy logic controllers by genetic algorithms. *International Journal of Approximate Reasoning*, 12:299–315, 1995.

[3] R. Alcalá, J. Alcalá-Fdez, and F. Herrera. A proposal for the genetic lateral tuning of linguistic fuzzy systems and its interaction with rule selection. *IEEE Trans. Fuzzy Syst.*, 15(4):616–635, 2007.

[4] F. Herrera and L. Martínez. A 2-tuple fuzzy linguistic representation model for computing with words. *IEEE Trans. Fuzzy Syst.*, 8(6):746–752, 2000.

[5] E. H. Mamdani and S. Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*, 7:1–13, 1975.

[6] L. J. Eshelman. The CHC adaptive search algorithm: How to have safe serach when engaging in nontraditional genetic recombination. In G.J.E. Rawlin, editor, *Foundations of genetic Algorithms*, volume 1, pages 265–283. Morgan Kaufman, 1991.

[7] F. Herrera, M. Lozano, and A. M. Sánchez. A taxonomy for the crossover operator for real-coded genetic algorithms: An experimental study. *International Journal of Intelligent Systems*, 18:309–338, 2003.

[8] F. Herrera and M. Lozano. Gradual distributed real-coded genetic algorithms. *IEEE Transactions on Evolutionary Computation*, 4(1):43–63, April 2000.

[9] B. Kröger, P. Schwenderling, and O. Vornberger. Parallel genetic packing on transputers. *Parallel Genetic Algorithms: Theory and Applications: Theory Applications*, pages 151–186, 1993.

[10] J. E. Baker. Adaptative selection methods for genetic algorithms. In MA Ed. L. Erlbraum Associates: Hillsdale, editor, *Proceedings of the First International Conference on Genetic Algorithms and their Applications*, pages 101–111, 1985.

[11] J. E. Baker. Reducing bias and inefficiency in the selection algorithm. In *Proceedings of the 2nd International Conference on Genetic Algorithms (ICGA'87)*, 1987.

[12] L.J. Eshelman and J.D. Schaffer. Real-coded genetic algorithms and interval-schemata. *Foundations of Genetic Algorithms*, 2:187–202, 1993.

[13] L. X. Wang and J. M. Mendel. Generating fuzzy rules by learning from examples. *IEEE Trans. Syst. Man and Cybernetics*, 22(6), 1992.

ISBN: 978-989-95079-6-8

Forecasting Exchange Rates: A Neuro-Fuzzy Approach

Meysam Alizadeh[1] Roy Rada[2] Akram Khaleghei Ghoshe Balagh[3] Mir Mehdi Seyyed Esfahani[1]

1 Department of Industrial Engineering, Amirkabir University of Technology (Tehran Polytechnic)
Tehran, Iran
2 Department of Information Systems, University of Maryland Baltimore County
Baltimore, USA
3 Researches & Planning Center of Bank Mellat
Tehran, Iran
Email: alizadeh.m@ieee.org, rada@umbc.edu, a.khaleghei@bankmellat.ir, msesfahani@aut.ac.ir,

Abstract—*This paper presents an adaptive neuro-fuzzy inference system (ANFIS) for USD/JPY exchange rates forecasting. Previous work often used time series techniques and neural networks (NN). ANFIS can be used to better explain solutions to users than completely black-box models, such as NN. The proposed neuro-fuzzy rule based system applies some technical and fundamental indexes as input variables. In order to generate membership functions (MFs), we make use of fuzzy clustering of the output space. The neuro-fuzzy model is tested with 28 candidate input variables for both currencies. For the purpose of comparison, Sugeno-Yasukawa model, feedforward multi-layer neural network, and multiple regression are benchmarked. The comparison demonstrates that the presented algorithm shows its superiority in terms of prediction error minimization, robustness and flexibility.*

Keywords— ANFIS Network, Exchange Rates, Forecasting, Neuro-Fuzzy Systems, Nonlinear Identification.

1 Introduction

The Efficient Market Hypothesis ([1], [2], [3]) states that one cannot predict the behavior of efficient markets, as the markets follow a random walk. However, other studies ([4], [5]) found evidences against the Efficient Market Hypothesis. Exchange rate prediction is one of the most challenging applications of modern time series forecasting. To become an efficient dealer in foreign exchange market, one must be conversant with the factors that are responsible for a currency to appreciate or depreciate.

Methods that have been proposed for currency exchange rates forecasting fall into three categories: (i) econometric methods, (ii) time series methods, and (iii) soft computing methods. The difficulty in predicting currency exchange rates, due to their high volatility and complexity, has long been a concern in international financial markets, as many econometric methods are unable to produce significantly better forecasts than the random walk model [6]. Time series methods have their limitations for multidimensional time series with mutual non-linear dependencies. Most recently, researchers have applied soft computing tools such as multilayer Artificial Neural Networks (ANN), Fuzzy Logic (FL), and their hybrids to exchange rate forecasting.

ANN support time-series analysis and forecasting and multivariate analysis. Medeiros et al [7] presented and compared alternatives to model and forecast monthly exchange rates time series. Medeiros et al [7] simulated ANN, neuro-coefficient smooth transition autoregressive, linear autoregression, and random walk models. Kamruzzaman and Sarker [8] have investigated ANN modeling of foreign exchange rates using three learning algorithms, namely, Standard Back Propagation, Scaled Conjugate Gradient, and Back Propagation with Bayesian Regularization. Majhi et al [9] proposed two low complexity ANN exchange rate prediction models for US Dollar to British Pound, Indian Rupees, and Japanese Yen exchange rates.

The literature has shown fuzzy sets as a superior modeling tool for many problems. The statistical fuzzy interval NN is proposed by [10] to perform knowledge discovery and predict currency exchange rate. To overcome the difficulty in finding matching rules for forecasting in the fuzzy time series model, [11] proposed to use the Euclidean distance between two fuzzy logic relationships as a metric for selecting matching rules and applied their model to exchange rate time series prediction.

At the computational level, a fuzzy system is a layered structure, similar to an ANN of the radial basis function type [12]. In order to optimize parameters in a fuzzy system, ANN learning algorithms can be employed. This neuro-fuzzy modeling approach can be better used to explain solutions to users than completely black-box models, such as ANN [13].

For the sake of clarity and completeness, section 2 gives a brief overview of the adaptive neuro-fuzzy inference system (ANFIS) approach. Section 3 systematically discusses neuro-fuzzy systems modeling. Section 4 carefully applies neuro-fuzzy modeling to US Dollar to Japanese Yen exchange rate forecasting. Finally, conclusions are given in Section 5.

2 Adaptive Neuro-Fuzzy Inference System

The Takagi-Sugeno-Kang (TSK) model uses fuzzy logic with crisp functions in consequent that is convenient for complex applications [14]. TSK systems are widely used in the form of a neuro-fuzzy system called Adaptive Neuro-Fuzzy Inference System (ANFIS). The structure of ANFIS and its learning algorithms are described in the following:

ISBN: 978-989-95079-6-8

2.1 ANFIS Structure

An ANFIS is a fuzzy inference system that can be trained to model some collection of input/output data. The training module allows the system to tune its parameters to learn the input/output relationships hidden in the data set. ANFIS uses two approaches: ANN and fuzzy modeling. By composing these two approaches, a suitable reasoning in quality and quantity might be achieved [15]. In ANFIS fuzzy logic is used to determine decision surfaces rather than to determine uncertainty associated with particular linguistic terms [16]. The rule-based representation of neuro-fuzzy systems offers transparency.

For pedagogical purposes, one can imagine a fuzzy inference system with two inputs x and y and one output z. The equivalent ANFIS architecture (Type-3 ANFIS) is shown in Fig. 1. The node functions in the same layer are from the same function family. The first layer implements a fuzzification, the second layer executes the T-norm of the antecedent part of the fuzzy rules, the third layer normalizes the membership functions, the fourth layer calculates the consequent parameters, and finally the last layer computes the overall output as the summation of all incoming signals. The feed forward equations of this ANFIS are as follows:

$$w_i = \mu_{A_i}(x) \times \mu_{B_i}(y), \tag{1}$$

$$\overline{w}_i = \frac{w_1}{w_1 + w_2}, i = 1,2. \tag{2}$$

$$\begin{cases} f_i = p_1 x + q_1 y + r_1 z \\ f_2 = p_2 x + q_2 y + r_2 z \end{cases} \Rightarrow \tag{3}$$

$$f = \frac{w_1 f_1 + w_2 f_2}{w_1 + w_2} = \overline{w}_1 f_1 + \overline{w}_2 f_2 \tag{4}$$

where,

- x is the input to node i,
- $\mu_{A_i}(x)$ is the node i node function,
- A_i is the linguistic label associated with node functions,
- w_i is the firing strength of the ith rule,
- $\overline{w}_i$ is the ratio of the ith rule's firing strength to the sum of all rules' firing strength,
- $\{p_i, q_i, r_i\}$ is the parameter set, and
- f_i is the consequent value.

Note that the network's output y is nonlinear in the weights w**.** The training of this ANN is thus a *nonlinear optimization problem* to which various methods can be applied [13].

2.2 Learning Algorithms

The neuro-fuzzy inference system is optimized by adapting the antecedent parameters and consequent parameters so that a specified objective function (usually a difference between the model output and the actual output) is minimized. A number of methods have been proposed for learning rules. For example, Mascioli, Varazi, and Martinelli [17] have proposed merging of Min-Max and ANFIS models to determine the optimal set of fuzzy rules. Jang and Mizutani [18] have presented an application of the Lavenberg-Marquardt method, which is essentially a nonlinear least-squares technique, for learning in an ANFIS network. Tang, Quek, and Ng [19] proposed a hybrid system combining a Fuzzy Inference System and Genetic Algorithms to tune the parameters in the TSK fuzzy ANN.

Jang [20] proposed methods to update the ANFIS parameters involving gradient descent and Least Square Error (LSE). High complexity is one of these methods' features. Several popular training algorithms for tuning parameters of ANFIS membership functions are compared in [21]. In this paper we use the hybrid learning algorithm proposed in [20] which is a combination of least square estimation and backpropagation algorithms.

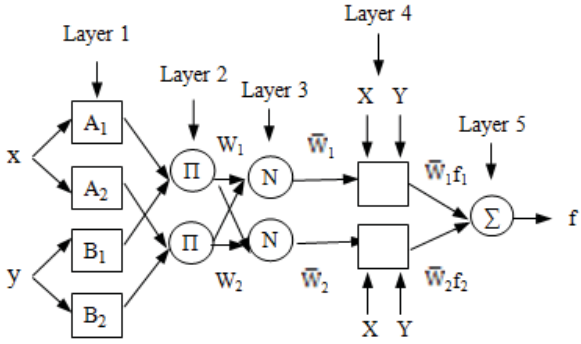


Figure 1: The equivalent ANFIS structure

3 Designing a Systematic Neuro-Fuzzy System

The two basic steps in neuro-fuzzy systems modeling are *system identification* and *fuzzy reasoning*. In the system identification stage, the significant input variables are determined, the fuzzy if–then rules are generated, and the parameters of the model, such as the number of clusters, the level of fuzziness, and the operators to be used in the reasoning, are selected. The second, *fuzzy reasoning*, is used to infer new knowledge from the identified rule base [22].

In fuzzy system modeling, the relationships between variables are represented with *if-then* rules with imprecise predicates. The basic steps are:

1) Fuzzy clustering of the output,
2) Input selection,
3) Rule base construction,
4) Tuning the parameters of membership function of input and output variables, and
5) Inference.

3.1 Fuzzy Clustering of the Output

To determine the number of rules of the initial ANFIS, one should choose the optimum number of clusters. For this purpose, a validity index which is proposed by [23] and modified by [24] is used. In other words, one minimizes:

$$V_{FNT}(U,V;X) = \frac{2}{c(c-1)} \sum_{p \neq q}^{c} S_{rel}(A_P, A_q) \tag{5}$$

The optimal number of clusters is obtained by minimizing $V_{\mathrm{FNT}}(U, V; X)$ over the range of c values: 2, …, $c_{\max}$; where

ISBN: 978-989-95079-6-8

$S_{rel}(A_P, A_q)$ is the relative similarity between two fuzzy sets A_p and A_q and is defined as:

$$S_{rel}(A_{P,}, A_q) = \sum_{i=1}^{n} S_{rel}(x_j : A_p, A_q) h(x_j) \quad (6)$$

where,

$$h(x_j) = -\sum_{p=1}^{c} u_{A_p}(x_j) \log(u_{A_p}(x_j)) \quad (7)$$

And $S_{rel}(x_j : A_p, A_q)$ is the relative similarity between two fuzzy sets A_p and A_q at x_j which is defined as:

$$S_{rel}(x_j : A_p, A_q) = \frac{f(x_j : A_p \cap A_q)}{f(x_j : A_p \cap A_q) + f(x_j : A_p - A_q) + f(x_j : A_q - A_p)} \quad (8)$$

Here, $h(x_j)$ is the entropy of datum x_j and $u_{A_p}(x_j)$ is the membership value with which x_j belongs to the cluster A_p.

After determining the optimum number of clusters, Gustafson-Kessel (GK) clustering algorithm is used for generating membership functions.

3.2 Input Selection

The performance of non-linear identification techniques is often determined by the appropriateness of the selected input variables and the corresponding time lags. High correlation coefficients between candidate input variables in addition to a non-linear relation with the output signal induces the need for an appropriate input selection methodology [25].

For variables selection, the Sugeno and Yasukawa [26] method is used. They proposed a combinatorial approach in which all possible combinations of input candidates are considered. For each combination, they built two fuzzy models based on two separated sets of data and calculated a performance index called "Regularity Criterion" (RC). After that a combination of input variables is chosen which has the minimum value of the performance index.

3.3 Rule Base Construction

There are three alternative strategies for incorporating the fuzzy clustering during rule base construction; cluster the:

1. output space and obtain the fuzzy membership functions based on the projections of the output clusters onto the input space [26].
2. input space, relating the output variables to each input cluster based on the degree of possibility [27].
3. joined input and output spaces and then project these multidimensional clusters to the separate input and output spaces [28].

Kilic, Uncu, and Turksen [22] compared these three approaches. This paper uses the first approach.

3.4 Membership Parameters Tuning

ANFIS has two kinds of parameters that needed to be trained: the antecedent parameters and the premises parameters. In this study, Gaussian membership functions are located in the antecedent part:

$$\mu_{A_i}(x) = \exp\left\{-\left[\left(\frac{x - c_i}{a_i}\right)^2\right]^{b_i}\right\} \quad (9)$$

It has three types of parameters: $\{a_i, b_i, c_i\}$, where a_i is the variance, b_i is the crossover slop and c_i is the center of MFs.

4 Exchange Rates Forecast Modeling

The authors used ANFIS in forecasting US Dollar/Japanese Yen exchange rates. While other researchers primarily used a few technical inputs ([8], [9], [10], [29], [30], [31]), (see Table 1) this work uses significant fundamental and technical inputs.

Table 1: List of some previous inputs data used in exchange rate forecasting

Authors	Model	Inputs
Tenti (1996)	Recurrent Neural Network	Compound returns of the last n periods (n=1, 2, 3, 5, 8) The running standard deviation of the k last periods (k=13, 21, 34) Average directional movement index (ADX) Trend movement index (TMI) Rate of change (ROC) Ehlers leading indicator (ELI)
Lisi and Schiavo (1999)	Chaotic models and NN	Monthly exchange rate of Franc, Deutschmark, Lira, and Pound- all against US Dollar
Kamruzzaman and Sarker (2004)	ANN	Moving average of one week, two weeks, one month, one quarter, half year closing rate Last week's closing rate
Preminger and Franck (2007)	Regression	Monthly exchange rate of Yen and Pound against the US dollar
Zhang and Wan (2007)	Fuzzy interval NN	Three, two, and one weeks ago rate of exchange
Majhi, Panda, and Sahoo (2009)	ANN	Normalized rate on the first day of a month Mean of the monthly rate Variance of the monthly rate

The candidate inputs are shown in Table 2. Similar to [32] the inputs are divided into categories. However, in each category, this study used some other indices than [32]. Additionally, some of the most popular technical input variables are used. As a result, 28 fundamental and technical candidate input variables, among which there exist 20 fundamental indices for both countries and the remaining ones have been allocated to technical indices, are fed to ANFIS. The full sample compromises daily observation for the Jan-2001 to Aug-2008 period from the Reuters 3000 Xtra Hosted Terminal Platform, where 60% of data points are used for training and the rest for testing the model.

After collecting candidate input data, the Sugeno-Yasukawa [26] input selection method is used. The method selected the following input variables:

- US Federal Reserve Bank interest rate,
- US M2 Money Supply: it technically defined as sum of M1, savings deposits, small denomination time

ISBN: 978-989-95079-6-8

deposits (where small is less than $100,000), and retirement account, where M1 is the sum of the tender that is held outside banks, travelers checks, checking accounts (but not demand deposits), minus the amount of money in the Federal Reserve float,

- West Texas Intermediate oil price,
- Rate of Change (10 days) of the USD/JPY time series,
- Momentum (5 days) of the USD/JPY time series, and
- Stochastic K% (5 days) of the USD/JPY time series.

Table 2: Candidate Input Data

No.	Factor	Indexes
1	Economic Activities	Gross domestic product, Industrial Production
2	Price	Consumer price index, Producer price index
3	Interest Rates	Short term interest rate (overnight, 1 week), Federal rate
4	Money Supply	M1, M2, M3
5	Trade Balance	Trade balance, BOP
6	Employment	Unemployment rate, Nonfarm payroll
7	Personal Consumption	Retail sale, Personal income
8	Oil	Brent, WTI oil
9	Stock	Dow Jones, Nikkei
10	Technical Indexes	Stochastic K%(5, 10 days), ROC(5, 10 days), Momentum(5, 10 days), William R%(5, 10 days)

The corresponding RC values are shown in Fig. 2. Next, based on an introduced Cluster Validity Index, we determine that the optimum number of clusters is 7 (see Fig. 3). GK fuzzy clustering algorithm is applied for initialization of antecedent membership functions. Then ANFIS with hybrid learning algorithm (combination of least square estimation and backpropagation algorithms) is used for parameters tuning. We use Mamdani-style inference, min–max operators, and centroid defuzzification method. The rules that ANFIS generated make sense relative to the economics of the foreign exchange market, as shown in Fig. 4.

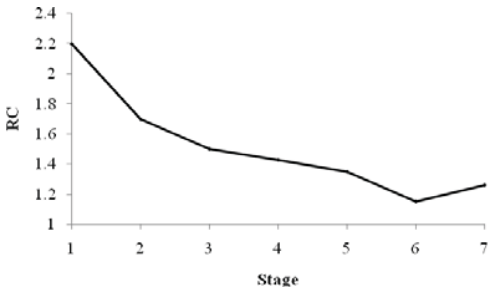


Figure 2: Behavior of RC in proposed model

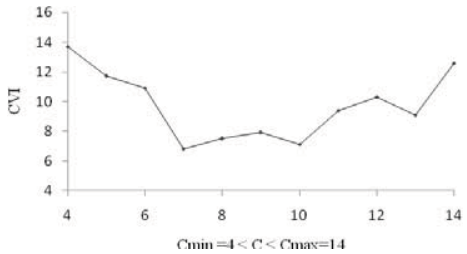


Figure 3: Identification of the optimum number of clusters

To further validate this approach for currency exchange rates forecasting, the results were compared to the results with a) the Sugeno-Yasukawa [26] approach, b) a feedforward multi-layer neural network, and c) multiple regression. The comparison uses the root mean square of the prediction error (RMSE) and mean error of the prediction (BIAS). The comparison results for training and testing dataset are shown in Table 3, 4, 5, and 6. The characteristics of the mentioned test models are described next.

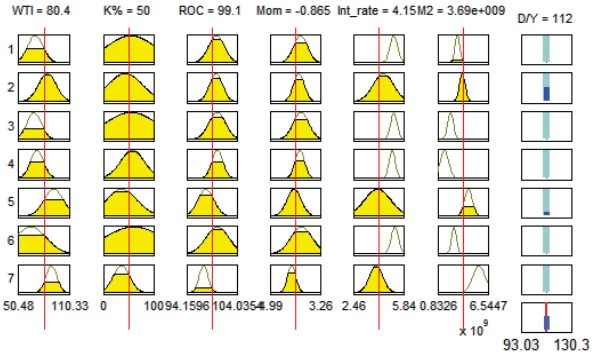


Figure 4: Fuzzy rule base of USD/JPY exchange rates forecasting

4.1 Sugeno-Yasukawa Approach

This Sugeno-Yasukawa [26] approach produces a fuzzy model with 6 rules, 6 inputs, and 1 output. The inputs are the same as with the ANFIS model. Other aspects of the approach include a Mamdani-style inference, min–max operators, and centroid defuzzification.

4.2 Feedforward Multi-layer Neural Network

A 10×5×1 feedforward multi-layer network with gradient descent learning algorithm is used to forecasting US Dollar against Japanese Yen currency exchange rates. A tangent sigmoid activation function is used in each node. The tests were performed for a maximum iteration number of 500.

4.3 Multiple Regression

The multiple regression analysis with Matlab ® is used for prediction. The regression equation is:

$$y = 1.9164\, x_1 - 0.00001\, x_2 - 0.0602\, x_3 + 1.7423\, x_4 - 0.2103\, x_5 - 0.0208\, x_6 \quad (10)$$

The output is computed by the above equation and the modeling performance measure is obtained.

Table 3: RMSE of different models for training set

Model	# Rules	# Runs	RMSE		
			Min.	Max.	Ave.
Multiple regression	-	10	6.1002	9.1770	7.2391
Neural network	-	10	1.3055	2.7595	2.1306
Sugeno-Yasukawa	6	10	3.1397	5.6691	3.8723
ANFIS	7	10	1.3805	2.4980	1.7447

ISBN: 978-989-95079-6-8

Table 4: RMSE of different models for testing set

Model	# Rules	# Runs	RMSE		
			Min.	Max.	Ave.
Multiple regression	-	10	6.3060	9.4483	7.9221
Neural network	-	10	1.3820	4.9647	2.8082
Sugeno-Yasukawa	6	10	3.9835	8.2009	4.8290
ANFIS	7	10	1.9284	5.4007	2.6301

Table 5: BIAS of different models for training set

Model	# Rules	# Runs	BIAS		
			Min.	Max.	Ave.
Multiple regression	-	10	1.7193	2.0918	1.7201
Neural network	-	10	0.5311	0.8375	0.7611
Sugeno-Yasukawa	6	10	0.8129	1.1602	0.9069
ANFIS	7	10	0.5591	0.7190	0.6207

Table 6: BIAS of different models for testing set

Model	# Rules	# Runs	BIAS		
			Min.	Max.	Ave.
Multiple regression	-	10	1.7660	2.2183	1.9219
Neural network	-	10	0.5005	1.1020	0.9011
Sugeno-Yasukawa	6	10	0.6291	1.5918	1.4820
ANFIS	7	10	0.5197	0.9518	0.8526

5 Conclusions

The average RMSE and BIAS results are better for the ANFIS approach than for the Sugeno-Yasukawa, ANN, or multiple regression approaches. However, the difference between the ANFIS and ANN average RMSE and BIAS does not appear significant. One might say that the performance of ANFIS and ANN was comparable as regards accuracy. However, with ANN one does not gain the benefit of the semantically meaningful presentation of rules that is possible with ANFIS. Given the choice between two computer systems that produce comparable test accuracies, humans are likely to prefer the system whose logic is transparent.

Future work would explore refined hypotheses in various directions. One hypothesis is that by taking advantage of further information about currency exchange rates, we can improve the initialization of antecedent membership functions and the clustering technique. Another hypothesis is that the application of other learning or optimization algorithms, such as genetic algorithms and particle swarm optimization, would demonstrate interesting variations on the representation and reasoning for the currency exchange problem. By implementing new hybrid learning programs, researchers might gain further insights about how to solve the currency exchange problem. Vast amounts of available data, clear definitions of success in testing, and practical importance are hallmarks of this problem. The problem of forecasting currency exchange rates provides a valuable test bed for soft computing methods, and researchers can thus elaborate the merit and the shortcomings of each method.

References

[1] F. Eugene, The Behavior of Stock Market Prices, *Journal of Business*, 38: 34–105, 1965.

[2] P. Samuelson, Proof That Properly Anticipated Prices Fluctuate Randomly, *Industrial Management Review*, 6: 41–49, 1965.

[3] F. Eugene, Efficient Capital Markets: A Review of Theory and Empirical Work, *Journal of Finance*, 25: 383–417, 1970.

[4] A.D. Hsieh, Chaos and non-linear dynamics: application to financial markets, *Journal of Finance,* 46:1833–1877, 1991.

[5] G. Tsibouris and M. Zeidenberg, Testing the efficient market hypothesis with gradient descent algorithm, in: A. Refenes (Ed.), *Neural networks in the capital markets*, John Wiley and Sons Ltd., England, 127–136, 1996.

[6] A.S. Chen and M.T. Leung, Regression neural network for error correction in foreign exchange forecasting and trading, *Computers and Operations Research*, 31:1049-1068, 2004.

[7] M.C. Medeiros, Á. Veiga, and C. E. Pedreira, Modeling Exchange Rates: Smooth Transitions, Neural Networks, and Linear Models, *IEEE Transactions on Neural Networks*, 12(4), July 2001.

[8] J. Kamruzzaman and R.A. Sarker, ANN-Based Forecasting of Foreign Currency Exchange Rates, *Neural Information Processing-Letters and Reviews*, 3(2), May 2004.

[9] R. Majhi, G. Panda, and G. Sahoo, Efficient prediction of exchange rates with low complexity artificial neural network models, *Expert Systems with Applications*, 36:181–189, 2009.

[10] Y.Q. Zhang and X. Wan, Statistical fuzzy interval neural networks for currency exchange rate time series prediction, *Applied Soft Computing*, 7:1149–1156, 2007.

[11] Y. Leu, C.P. Lee, and Y.Z. Jou, A distance-based fuzzy time series model for exchange rates forecasting, *Expert Systems with Applications*, doi:10.1016/j.eswa.034, 2008.

[12] J.S.R. Jang and C.T. Sun, Functional equivalence between radial basis function networks and fuzzy inference systems, *IEEE Transactions on Neural Networks*, 4(1): 156–159, 1993.

[13] R. Babuška and H. Verbruggen, Neuro-fuzzy methods for nonlinear system identification. *Annual reviews in control*, 27:73-85, 2003.

[14] R. Alcala, J. Casillas, O. Cordon, and F. Herrera, Learning TSK rule-based system from approximate ones by mean of MOGUL methodology. Granada University of Spain, 2000.

[15] M. Teshnehlab, M.A. Shoorehdeli, and A.K. Sedigh, Novel hybrid learning for tuning ANFIS parameters as an identifiers using fuzzy PSO, *IEEE International Conference on Networking, Sensing and Control, Sanya, China*, 111-116, 2008.

[16] G.S. Atsalakis and K.P. Valavanis, Surveying stock market forecasting techniques, PartII: Soft computing methods. *Expert Systems with Applications*, doi:10.1016/j.eswa.2008.07.006, 2008.

[17] F.M. Mascioli, G.M. Varazi, and G. Martinelli, Constructive Algorithm for Neuro-Fuzzy Networks, Proceedings of the *Sixth IEEE International Conference Fuzzy Systems, Spain*, 1: 459- 464, 1997.

[18] J.S.R. Jang and E. Mizutani, Levenberg-Marquardt Method for ANFIS Learning, *Biennial Conference of the North American Fuzzy Information Processing Society*, 87-91, 1996.

[19] A.M. Tang, C. Quek and G.S. Ng, GA-TSKfnn: Parameters tuning of fuzzy neural network using genetic algorithms. *Expert Systems with Applications,* 29: 769-781, 2005.

[20] J.S.R. Jang, ANFIS: Adaptive-network-based fuzzy inference systems, *IEEE Transactions on Systems, Man & Cybernetics,* 23(3): 665–685, 1993.

[21] M.S. Chen, A Comparative Study of Learning Methods in Tuning Parameters of Fuzzy Membership Functions, *IEEE SMC '99 Conference Proceedings, Tokyo, Japan*, 40-44,1999.

[22] K. Kilic, O. Uncu, and I.B. Turksen, Comparison of different strategies of utilizing fuzzy clustering in structure identification, *Information Sciences,* 177:5153-5162, 2007.

ISBN: 978-989-95079-6-8

[23] Y.I. Kim, D.W. Kim, D. Lee, and K.H. Lee, A cluster validation index for GK cluster analysis based on relative degree of sharing, *Information Sciences,* 168: 225-242, 2004.

[24] M.H. Fazel Zarandi, B. Rezaee, I.B. Turksen, and E. Neshat, A type-2 fuzzy rule-based expert system model for stock price analysis, *Expert Systems with Applications,* 36: 139–154, 2009.

[25] K. Maertens, J.D. Baerdemaeker, and R. Babuška, Genetic polynomial regression as input selection algorithm for non-linear identification, *Soft Computing,* 10:785-795, 2006.

[26] M. Sugeno and T. Yasukawa, A fuzzy logic based approach to qualitative modeling, *IEEE Transactions on Fuzzy Systems,* 1:7–31, 1993.

[27] G. Castellano, A. Fanelli, and C. Mencar, A fuzzy clustering approach for mining diagnostic rules, in: Proceedings of IEEE Conference on Systems, Man and Cybernetics, Washington, USA, 2003.

[28] O. Uncu, I.B. Turksen, and K. Kilic, Localm-fsm: A new fuzzy system modeling approach using a two-step fuzzy inference mechanism based on local fuzziness level, in: Proceedings of *International Fuzzy Systems Association World Congress*, 191–194, 2003.

[29] P. Tenti, Forecasting foreign exchange rates using recurrent neural network. *Applied Artificial Intelligence* 10: 567-581, 1996.

[30] F. Lisi and R.A. Schiavo, A comparison between neural networks and chaotic models for exchange rate prediction. *Computational Statistics & Data Analysis* 30: 87-102, 1999.

[31] A. Preminger and R. Franck. Forecasting exchange rates: A robust regression approach. *International Journal of Forecasting* 23: 71-84, 2007.

[32] K. Izumi and K. Ueda, Phase Transition in a Foreign Exchange Market—Analysis Based on an Artificial Market Approach, *IEEE Transactions On Evolutionary Computation*, 5(5), October 2001.

ISBN: 978-989-95079-6-8

Genetic Learning of Serial Hierarchical Fuzzy Systems for Large-Scale Problems

Alicia D. Benítez and Jorge Casillas[1]

[1] Department of Computer Science and Artificial Intelligence
CITIC-UGR (Research Center on Information and Communications Technology)
University of Granada, E-18071 Granada, Spain
Email: aliciades@gmail.com, casillas@decsai.ugr.es

Abstract— *When we face a problem with a high number of variables by a standard fuzzy system, the number of rules increases exponentially and then the obtained fuzzy system is scarcely interpretable. This problem can be handled by arranging the inputs in hierarchical ways. The paper presents a multi-objective Genetic Algorithm that learns Serial Hierarchical Fuzzy Systems with the aim of coping the curse of dimensionality. By means of an experimental study, we have observed that our algorithm obtains good results of interpretability and precision with problems in which the number of variables is relatively high.*

Keywords— Curse of dimensionality, hierarchical fuzzy systems, multi-objective genetic algorithms, variable selection.

1 Introduction

If a conventional fuzzy system is applied to large-scale problems (i.e, those with a high number of input variables), the number of rules grows exponentially with respect to the number of inputs received [1, 2]. Indeed, if we have n variables and k linguistic terms per variable, it requires up to n^k rules to build a complete Mamdani-based fuzzy system, and consequently, the accuracy-interpretability balance would be broken. This problem is know as the "curse of dimensionality."

In order to solve it, several approaches have been suggested such as variable selection [3, 4] and rule set reduction [5, 6]. Nevertheless, when the number of variables increases considerably, this kind of reduction is not enough to solve this problem. There exists a different approach which deals with this problem: Hierarchical Fuzzy Systems (HFS). An HFS is made up of a set of fuzzy subsystems or modules. These modules are linked in such a way that the output of a module is the input of other ones. Thanks to the decomposition of the fuzzy system made in an HFS, the complexity of each module is significantly reduced. There are several kinds of modules:

- SISO (Single Input Single Output): It has one input and one output.
- MISO (Multiple Inputs Single Output): It has several inputs and a single output [7]. We can find FLU (Fuzzy Logic Unit) in this kind of modules. One FLU special case is the one with two inputs and one output, which is equally found in the literature [2, 8–11].
- MIMO (Multiple Inputs Multiple Outputs): They have several inputs and outputs [12].

Apart from distinguishing several kinds of modules, it is also possible to find different types of hierarchical structures. There are different classifications [9, 13], though the most general one is the following [11]:

- Serial HFS (SHFS): The input of one module is the output of the previous ones, along with external variables [10].
- Parallel HFS: This system is organized in layers. The first one is made up of a set of modules receiving the input variables. Each variable is used as input only in a single module. The output of the modules in the first layer is the input of the modules which constitute the next layer, and so on. An aggregate operation might also exist in order to combine the outputs of one layer [2, 12].
- Hybrid HFS: This type of HFS is a mixture of the two previous ones [7, 14].

Other approaches study the best way to deal with the output variables in the modules (through fuzzy numbers [9] and matrixes modeling [11]) and the rules in different hierarchical levels [15]. Besides, some metaheuristics have been applied in this field with the purpose of obtaining an HFS capable of getting the best balance between precision and interpretability. For instance, Differential Evolution [10] has been employed to find the best membership functions, Genetic Algorithms [8], Ant Systems [14], Descendent Gradient Method [7] have been used to learn the hierarchical structure.

In this paper, a multi-objective genetic fuzzy system is suggested to obtain the best distribution of input variables and modules belonging to SHFS removing variables by crossover and mutation operators, so it can face the "curse of dimensionality." The paper is organized in the following sections: in Section 2, the suggested algorithm is described; in Section 3, the empirical study is shown; in Section 4, conclusions and future works are referred.

2 GSHFS algorithm

We propose a SHFS structure learning algorithm. These systems are a set of linked modules where the output of a module is one of the inputs of the next module. A module has its own fuzzy rule set which allows to infer the output variable from the input variables.

The algorithm uses a multi-objective genetic algorithm called Genetic SHFS (GSHFS) where there are two objectives to minimize: the root mean squared error (RMSE) and the number of rules. Our algorithm has three specific genetic operators: one crossover operator and two mutation operators.

ISBN: 978-989-95079-6-8

The following subsections detail the different components of the algorithm.

2.1 Coding scheme

In our algorithm, each individual represents a SHFS. The coding consists of a gene concatenation. Each gene has two fields: a variable index and a flag. The flag takes binary values (0 if it is an exogenous variable and 1 if it is an endogenous variable). Figures 1, 2, and 3 show some examples of the coding scheme.

A variable with 1 in the flag field means that is an endogenous variable (variable linking two modules) and a module exists. This variable is the input of the next module. All treated variables are variables of the problem. The algorithm decides if a variable of the problem will be endogenous or exogenous, but does not create new endogenous variables. In this way, the correlation between the variables of the problem is considered. The length of chromosome is fixed to the number of input variables of the problem. The highest hierarchy level happens when all modules are SISO. In this case, all variables will be endogenous except the first and the number of levels is equal to the number of input variables.

2.2 Initialization

The algorithm randomly generates an initial population. The variable index and the endogenous/exogenous flag is randomly chosen for each gene. The only restriction is that the variable can not be repeated and the flag of the first gene can not be 1 (i.e., the first variable have to be exogenous).

2.3 Crossover operator

The crossover operator is applied according to a probability between paired parents and provides exploitation. When crossover operator is applied to two parents, P_1 and P_2, there are a set of types of combinations which we must consider: 1) both parents, P_1 and P_2, have several modules; 2) parent P_1 has several modules and parent P_2 has one module; 3) parent P_1 has one module and parent P_2 has several modules (this case is similar to previous case); 4) both parents, P_1 and P_2, have one module. A parent-centric crossover has been used in some cases where the offspring mainly inherits the information of one of the parent and takes the secondary parent in order to add diversity.

The crossover is applied to individuals depending on the types of variables that two parents have in common. Each case has a priority. In this way, the crossover operator is applied from high to low priority as follows:

2.3.1 Both parents, P_1 and P_2, have several modules

- *Priority 1.- Endogenous-Endogenous case.* If P_1 parent has common endogenous variables with P_2 parent, a common endogenous variable is selected at random. This variable is the crossover point. The offspring O_1 is generated centered on P_1. Thus, the offspring O_1 inherits from P_2 the variables and modules from the beginning to the crossover point, but excluding the crossover point. The remaining is taken from P_1 and repeated variables are removed from the part inherited from P_2. The offspring O_2 is generated in the same way but centered on P_2.

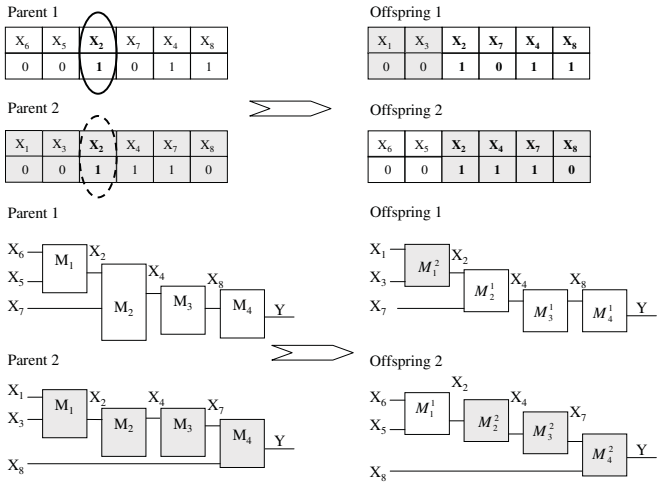


Figure 1: Example of crossover of two parents with several modules, Endogenous-Endogenous case (priority 1)

- *Priority 2.- Endogenous-Exogenous case.* If P_1 has endogenous variables which are exogenous in P_2, a common variable is randomly selected as crossover point. The offspring O_1 is generated centered on P_1. The common endogenous variable inherited from P_1 is converted into exogenous and the rest of previous modules to this variable are removed. The idea of this type of crossover is that if an endogenous variable of P_1 is equal to an exogenous variable of P_2, it indicates us that if we convert this endogenous variable into exogenous variable, RMSE will decrease, because the SHFS does not carry any error, taking the variable directly as an input.

- *Priority 3.- Exogenous-Endogenous case.* If P_1 has exogenous variables which are endogenous in P_2, one of those variables is randomly selected as crossover point. The offspring O_1 is created centered on P_1, but excluding the crossover point. The first part of O_1 is inherited from the first part of P_2 (from the first gene to the crossover point). The repeated variables are removed in O_1 from the part inherited from P_2. Figure 2 shows an example of this case. The generated SHFS offspring has four modules.

- *Priority 4.- Exogenous-Exogenous case.* If P_1 has a set of exogenous variables in common with exogenous variables in P_2, the offspring O_1 is generated as copy of P_1, but with a change: a common exogenous variable between P_1 and P_2 is randomly chosen. This variable in P_2, along with others variables, generates an endogenous variable. This endogenous variable in P_2 is selected to replace the random exogenous selected variable in O_1, which is common in P_1 and P_2. Later, the repeated variables in O_1 in part inherited from P_1 are removed. If this case comes true in P_2, O_2 offspring will be generated with the same procedure, but centered on P_2. Notice that the common exogenous variables of P_1 are looked for in P_2 excluding the exogenous variables of the module with output Y in P_2, because if this module is included, when an exogenous variable is selected there is not endogenous variable as output because it coincides with the output of

ISBN: 978-989-95079-6-8

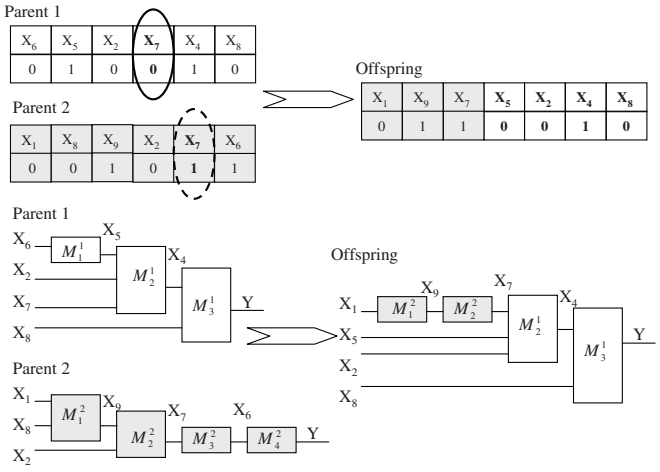


Figure 2: Example of crossover of two parents with several modules, Exogenous-Endogenous case (priority 3)

the SHFS.

- *Priority 5.- Different variables case.* If variables in P_1 and P_2 are different, an endogenous variable from P_1 and P_2 is chosen at random as crossover point. The offsprings are generated like priority 1.

2.3.2 *Parent P_1 has several modules and parent P_2 has one module*

- *Priority 1.- Endogenous-Exogenous case.* If P_1 has common endogenous variables with the exogenous variables of P_2, the offspring O_1 is generated like Endogenous-Exogenous case when P_1 and P_2 have several modules.

- *Priority 2.- Exogenous-Exogenous case.* If P_1 and P_2 have common exogenous variables, the offspring O_1 centered on P_1 is created as follows. Firstly, O_1 is generated as a copy of P_1. Then, an exogenous common variable of P_1 and P_2 is randomly chosen and moved to module with output Y. This type of crossover carries lower error and the output of SHFS is better.

- *Priority 3.- Different variables case.* In this case, O_1 is a copy of P_1 and later, an exogenous variable of P_2 is selected at random and inserted in the module of O_1 with the output Y.

2.3.3 *Parent P_1 has one module and parent P_2 has several modules*

- *Priority 1.- Exogenous-Endogenous case.* If P_1 has exogenous variables which are endogenous in P_2, the offspring O_1 is generated like Exogenous-Endogenous case when P_1 and P_2 have several modules.

- *Priority 2.- Exogenous-Exogenous case.* If P_1 has a set of exogenous variables in common with exogenous variables in P_2, the offspring O_1 is generated like Exogenous-Exogenous case when P_1 and P_2 have several modules.

- *Priority 3.- Different variables case.* If variables in P_1 and P_2 are different, an exogenous variable from P_1 and an endogenous variable from P_2 are chosen at random as crossover point. The offsprings are generated like priority 1.

2.3.4 *Both parents, P_1 and P_2, have one module*

This case, the offsprings O_1 and O_2 are generated as follows. The common exogenous variables of P_1 and P_2 are inserted in both offsprings. The rest of variables (no common variables between parents) are inserted in a set V with $|V|$ size. A number r is randomly generated: if there are not common variables between P_1 and P_2 then $r \in \{1, \ldots, n-1\}$; otherwise, $r \in \{0, \ldots, n\}$. r variables are randomly took out and inserted in O_1. Finally, the rest of variables are inserted in O_2.

2.4 *Mutation operator*

The mutation operator makes local changes in the chromosome. Two kinds of mutations have been designed:

2.4.1 *Exchange Mutation*

This operator makes an exchange of variables in a module. It chooses an exogenous variable at random in a module and exchanges between this exogenous variable and the endogenous variable of the module.

2.4.2 *Insertion Mutation*

The mutation operator distinguishes between: used and unused variables in the individual. According to it and by means of a probabilistic decision (Algorithm 1 shows a scheme), the mutation operator will choose between inserting an unused variable in a module, removing an used variable, or moving an used variable to other module.

Algorithm 1 Insertion Mutation

```
r = U[0,1];
if (r < 0.5 and there are unused variables) then
    v = Choose_at_random_a_unused_variable();
    Insert_unused_variable(v);
else
    t = U[0,1];
    if (t < 0.5 and the SHFS has only a SISO module) then
        Remove_one_used_variable();
    else
        Move_used_variable();
    end if
end if
```

- The insertion of an unused variable is as follows. Given an unused variable v, a module m is chosen randomly. Next, the operator decides if v is inserted into m with a probability of 0.5. If so, v is inserted as exogenous variable. Otherwise, if m has at least an exogenous variable as input, one of them, e, is selected at random. The mutation operator creates a new SISO module previous to m where the input is e and the output is v. In this case, v will be endogenous variable and input of module m. If m has not exogenous variables, the m's output is converted into a new exogenous variable of m and v will be the new output of m.

- To remove an used variable, first a module m of the SHFS is randomly chosen. If m has exogenous variables then one of them is removed at random. Otherwise, the module m is removed, thus linking the output of the module before m with the input of the module after m.

ISBN: 978-989-95079-6-8

- To move an used variable from a module to other, two modules of the chromosome are chosen at random: a source module m_1 where a variable is removed and a destination module m_2 where the variable is inserted. An exogenous variable of the m_1 is randomly removed and inserted as exogenous variable of the m_2. If there are not exogenous variables into the m_1 and consequently m_1 has one input and one output (a SISO module), its endogenous variable is chosen. In this situation, m_1 disappears and the chosen variable is inserted as an exogenous variable into m_2. If $m_1 = m_2$, an exogenous variable v is randomly selected, then a new module is created after m_1/m_2. The input of this new module (and therefore the output of the module m_1/m_2) is v while the output of the new module will be old output of m_1/m_2.

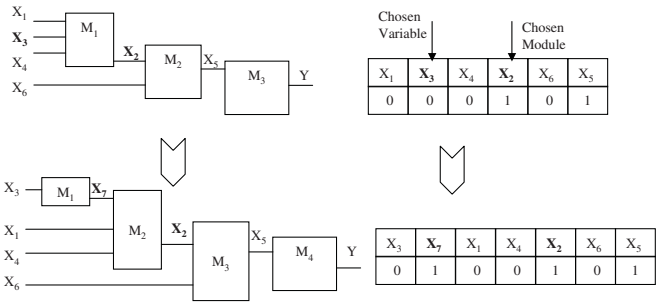


Figure 3: Example of insertion mutation of a unused variable (X_7) as endogenous variable, thus creating a new module

2.5 Fuzzy rule set learning

The learning fuzzy rule set of SHFS is as follows. Each module learns its fuzzy rule set with Mamdani rules depending on the inputs variables by WM [16]. The endogenous variables in SHFS are inferred by its respective module according to the input variables of the module. The obtained system is evaluated by training data examples, choosing the correspondent values of exogenous variables in that module. A module, according to learned fuzzy rule set, generates a value of output (an endogenous variable). The value is the next input of the module along with others exogenous variables. The membership functions are triangular shape distributed strong fuzzy partition.

2.6 Multi-objective approach

A generational approach with the multi-objective elitist replacement strategy of NSGA-II [17] is used. Crowding distance [17] in the objective function space is considered. Binary tournament selection based on the non-domination rank (or the crowding distance when both solutions belong to the same front) is applied. The crowding distance is normalized for each objective according to the extreme values of the solutions contained in the analyzed front.

2.7 Objective functions

We consider two objective functions to be minimize and so get a better precision and interpretability of the system.

2.7.1 Accuracy

It is computed as the RMSE:

$$F_1(S) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(S(x^i) - y^i)^2} \quad (1)$$

with S being the fuzzy system to be evaluated, N the data set size and (x^i, y^i) the ith input-output pair of the data set. The objective is to minimize this function to get good precision.

2.7.2 Interpretability

It is computed as follows:

$$F_2(S) = r(S) \quad (2)$$

with $r()$ being the number of rules of the fuzzy system.

Although the number of modules or the number of variables by module are notable criteria to evaluate the learned quality of the SHFS, the number of rules is more intuitive to validate the interpretability because it combines both criteria in one objective. If the number of rules is considered, the number of variables and modules is controlled automatically: as the number of rules decreases, the number of modules and variables increase.

3 Experimental Results

This section presents the experimental results. The objective of experimentation is to prove the reduction of number of rules and so, to assess the interpretability that can be compared to a referential learning method (WM in our case).

3.1 Problems

We have considered five regression problems with a moderate and high number of input real-valued variables: *Dee*[1], *Concrete*[2], *Elevators*[3], *Computer Activity (Comp-activ)*[3], and *Ailerons*[3].

Table 1 collects the main features of each problem, where *#InputVar* stands for number of input variables, *#Exam* for total number of examples, and *#LingTerms* for the number of triangular-shaped uniformly distributed linguistic terms considered for each variable in this experimental analysis.

Table 1: Data sets considered in the experimental analysis

Data set	*#InputVar*	*#Exam*	*#LingTerms*
Dee	6	365	5
Concrete	8	1030	5
Elevators	18	16559	3
Comp-activ	21	8192	3
Ailerons	40	13750	3

The experiments shown in this paper have been performed by a five fold cross validation with a total of 30 runs per problem (six runs for each data set partition). Thus, the data set is divided into five subsets of (approximately) equal size. The algorithm is then applied five times to each problem, each time

[1] KEEL: Knowledge extraction based on evolutionary learning. http://www.keel.es

[2] UCI Machine Learning Repository. Collection of regression datasets. http://archive.ics.uci.edu/ml/datasets.html

[3] L. Torgo, Collection of regression datasets. http://www.liacc.up.pt/ ltorgo/Regression/DataSets.html

ISBN: 978-989-95079-6-8

leaving out one of the subsets from training, but using only the omitted subset to compute the test error.

Our algorithm has been run with the following parameter values: 1,000 iterations, 60 as population size, 0.7 as crossover probability, and 0.2 as mutation probability per chromosome. We have not performed any previous analysis to fix these values, so better results may probably be obtained by tuning them though we have informally noticed our algorithm is not specially sensitive to any parameter.

3.2 Obtained results

For each problem we present the five validation test results (Table 2), where $RMSE_{tra}$ and $RMSE_{tst}$ are the approximation error (eq. 1) values over the training and test data set respectively, *#M*, *#R*, *#V* stand for the number of modules, the number of fuzzy rules, and the number of variables respectively.

Since our algorithm performs multi-objective optimization, several solutions are returned in each run. Therefore, we show five representative solutions from the final Pareto-optimal set: the first row of each problem is the best solution (highest precision), the second row is the first quartile, the third row is the median (it is not the average), the fourth row is the third quartile, and the fifth row is the worst solution (lowest precision).

Table 2: Obtained Results

	Method	$RMSE_{tra}$	$RMSE_{tst}$	*#M*	*#R*	*#V*
DEE	WM	0.375594	0.477303	1.0	178.4	6.0
	GSHFS	0.372482	0.474365	1.2	157.3	5.6
		0.396565	0.460321	1.8	95.6	5.2
		0.430456	0.484292	1.8	66.1	4.6
		0.495178	0.507092	2.1	28.6	3.7
		0.681968	0.695144	1.2	5.5	1.2
CONCRETE	WM	8.548268	9.822000	1.0	310.4	8.0
	GSHFS	8.479054	9.356946	1.4	257.1	7.2
		9.436752	9.812343	2.2	114.7	6.0
		10.700607	10.839647	2.3	53.3	5.2
		13.111270	13.450416	2.4	25.4	4.2
		16.669822	16.477185	1.3	5.6	1.3
ELEVATORS	WM	0.567771	0.542072	1.0	511.0	18.0
	GSHFS	0.487507	0.488641	2.3	39.7	7.7
		0.498203	0.498650	2.2	27.5	6.5
		0.517594	0.517673	2.4	18.9	5.7
		0.563252	0.563026	2.0	11.7	4.3
		1.421620	1.422627	1.6	3.8	1.8
COMP-ACTIV	WM	9.050474	9.084798	1.0	425.6	21.0
	GSHFS	5.398862	5.400107	2.5	37.5	8.1
		5.880516	5.874316	2.3	25.1	6.6
		6.708214	6.719212	2.2	16.8	5.5
		8.888890	8.935257	2.1	9.7	3.9
		31.900506	31.756379	1.7	4.1	2.0
AILERONS	WM	0.253971	0.255557	1.0	1080.6	40.0
	GSHFS	0.238855	0.239763	2.8	57.1	10.0
		0.254118	0.254958	2.7	37.7	8.2
		0.276632	0.277216	2.6	24.0	6.8
		0.318274	0.318159	2.3	12.9	5.0
		0.724993	0.731567	2.2	5.9	2.9

3.3 Analysis

The main objective is to obtain a fuzzy system with good tradeoffs between precision and interpretability. When the number of rules of a fuzzy system decreases, the system gets better interpretability, but the precision is lower and vice versa. This event can be observed clearly in the results obtained as problems with either low or high number of variables. Notice that the fuzzy rule set is learned individually for each module.

First set of Table 2 shows five solutions of the Pareto obtained by our algorithm in *Dee* problem. We can observe that the precision is high when the number of rules is high and, consequently, the fuzzy system has a lower interpretability. If we compare our algorithm with WM method, the best solution (highest precision) is better in precision, having a number of rules and a number of variables lower. In first quartile, the precision has a little increase according to solution obtained by WM in the $RMSE_{tra}$, but the $RMSE_{tst}$ is lower and the interpretability is decreased in more than a half.

However, if we observe the problems with a considerable number of variables (more than eight), our algorithm has better performance. Lets look at *Elevators* problem. It has 18 variables. The best solution (highest precision) obtained by our algorithm is better than WM solution. The $RMSE_{tra}$ and $RMSE_{tst}$ have been decreased. The interpretability is better because the SHFS has a lower number of variables and a higher number of modules. The number of rules has been decreased by 92%. The third quartile of the $RMSE_{tra}$ is better in precision than WM. The third quartile of the $RMSE_{tst}$ is a little worse but the number of rules obtained by our algorithm has decreased by 95%, 98%, and 96%, respectively.

In *Ailerons* problem (40 variables), we can see that our solution with highest precision is better than the solution obtained by WM. We want to emphasize that the number of rules is decreased by 95% and the number of variables by 75%. The median of the $RMSE_{tra}$ is a solution with a precision a little higher than WM solution, but the median of the $RMSE_{tst}$ is lower and the interpretability is better than the solution of WM, due to decrease of the number of variables and its distribution in modules.

To sum up, we have seen that in problems with a low moderate number of variables, the precision is worse when the number of rules decrease. It does not happen in problems with a higher number of variables because with a lower number of variables, the correlation between them is low, because they are important. Nevertheless, in problems with a higher number of variables exists more dependence between them: a hierarchical structure and a lower number of variables help to decrease the complexity of the learned fuzzy model.

Figure 4 shows the average Pareto front obtained in Ailerons problem by our algorithm. In it we can see that if there are many rules (lower interpretability), then the precision is higher. The SHFS gets a better interpretability with a lower precision.

Finally, we show an example of GSHFS hierarchical structure obtained by our algorithm in Figure 5. It is an example of non-dominated solution obtained by GSHFS in a data set partition of *Ailerons*. The obtained values are: $RMSE_{tra}$ = 0.214068, $RMSE_{tst}$ = 0.219682, #M = 3, #R = 63, #V = 9. The results obtained by WM in the same data set partition are: $RMSE_{tra}$ = 0.254144, $RMSE_{tst}$ = 0.261472, #M = 1, #R = 1072, #V = 40. We can observe that $RMSE_{tra}$ and $RMSE_{tst}$ obtained by our algorithm is lower than the values obtained by WM, decreasing the number of rules by 94%.

ISBN: 978-989-95079-6-8

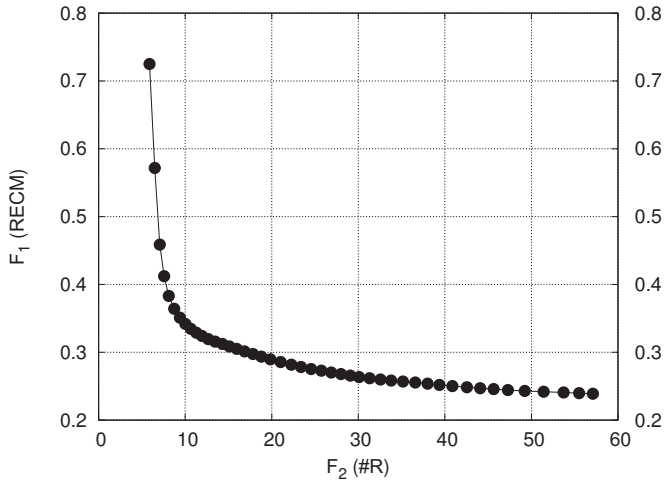


Figure 4: Average Pareto front obtained in Ailerons problem

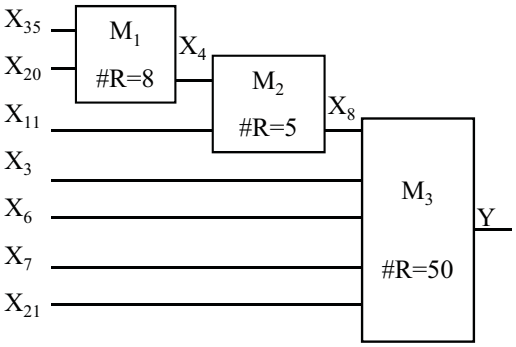


Figure 5: Example of a solution obtained by GSHFS in the *Ailerons* problem with #M = 3, #R = 63, #V = 9

4 Conclusion and Further Work

We have proposed a multi-objective algorithm applied to learning SHFS for palliate exponential increase of the number of rules when number of variables increases. The set of variables is divided into modules by the algorithm. We have proved that this division by SHFS can obtain good results in problems with a higher number of variables.

As further work, we suggest to add mechanisms for a better precision, for example, a global learning of fuzzy rule set, a detailed study of the interpretability of each rule of the fuzzy system, and to extend this algorithm for parallel and hybrid structure learning.

Acknowledgment

This work was supported in part by the Spanish Ministry of Science and Innovation (grant no. TIN2008-06681-C06-01), and by Andalusian Government (grant no. P07-TIC-3185).

References

[1] G.V.S. Raju, J. Zhou, and R.A. Kisner. Hierarchical fuzzy control. *International Journal Control*, 54(5):1201–1216, 1991.

[2] M.G. Joo and J.S. Lee. Hierarchical fuzzy control scheme using structured takagi-sugeno type fuzzy inference. In *Proceedings of IEEE International Fuzzy Systems Conference*, pages 78–83, Seoul, Korea, 1999.

[3] T.P. Hong and J.B. Chen. Finding relevant attributes and membership functions. *Fuzzy Sets and Systems*, 103(3):389–404, 1999.

[4] A. González and R. Pérez. Selection of relevant features in a fuzzy genetic learning algorithm. *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, 31(3):417–425, 2001.

[5] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka. Selecting fuzzy if-then rules for classification problems using genetic algorithms. *IEEE Transactions in Fuzzy Systems*, 3(3):260–270, 1995.

[6] T. Taniguchi, K. Tanaka, H. Ohtake, and H.O. Wang. Model construction, rule reduction, and robust compensation for generalized form of takagi-sugeno fuzzy systems. *IEEE Transactions on Fuzzy Systems*, 9(4):525–538, 2001.

[7] D. Wang, X.J. Zeng, and J.A. Keane. Learning for hierarchical fuzzy systems based on gradient-descent method. In *Proceedings of IEEE International Conference on Fuzzy Systems*, pages 92–99, 2006.

[8] K. Shimojima, T. Fukuda, and Y. Hasegawa. Self-tuning fuzzy modeling with adaptive membership function, rules, and hierarchical structure based on genetic algorithm. *Fuzzy Sets and Systems*, 71:295–309, 1995.

[9] A.E. Gaweda and R. Scherer. Fuzzy number-based hierarchical fuzzy system. *Lecture Notes in Computer Science*, 3070:302–307, 2004.

[10] F. Cheong. A hierarchical fuzzy system with high input dimensions for forecasting foreign exchange rates. *IEEE Congress on Evolutionary Computation, CEC*, pages 1642–1647, 2007.

[11] S. Aja-Fernández and C. Alberola-López. Matriz modeling of hierarchical fuzzy systems. *IEEE Transactions in Fuzzy Systems*, 16(3):585–599, 2008.

[12] P. Salgado. Rule generation for hierarchical collaborative fuzzy system. *Applied Mathematical Modelling, Science Direct*, 32:1159–1178, 2008.

[13] V. Torra. A review of the construction of hierarchical fuzzy systems. *International Journal of Intelligent Systems*, 17:531–543, 2002.

[14] Y. Chen, J. Dong, and B. Yang. Automatic design of hierarchical ts-fs model using ant programming and pso algorithm. In C. Bussler and D. Fensel, editors, *Proceedings 12th International Conference on Artificial Intelligence, Methodology, Systems and Applications, Lecture Notes on Artificial Inteligence, LNAI 3192*, pages 285–294, 2004.

[15] T.M. Jelleli and A.M. Alimi. Improved hierarchical fuzzy control scheme. *Adaptive and Natural Computing*, 1:128–131, 2005.

[16] L.X. Wang and J. Mendel. Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(6):1414–1427, 1992.

[17] K. Deb, A. Pratap, S. Agarwal, and T. Meyarevian. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002.

ISBN: 978-989-95079-6-8

An Introduction to Parameterized IFAM Models with Applications in Prediction

Peter Sussner[1] Rodolfo Miyasaki[2] Marcos Eduardo Valle[3]

1.Department of Applied Mathematics, University of Campinas
Campinas, State of São Paulo, Brazil
2.Credit Planning and Management Department, Banco Itaú
São Paulo, Brazil
3.Department of Mathematics, University of Londrina
Londrina, Brazil
Email: sussner@ime.unicamp.br, rodolfo.miyasaki@itau.com.br, valle@uel.br

Abstract— *Fuzzy associative memories (FAMs) and, in particular, the class of implicative fuzzy associative memories (IFAMs) can be used to implement fuzzy rule-based systems. In this way, a variety of applications can be dealt with. Since there are infinitely many IFAM models, we are confronted with the problem of selecting the best IFAM model for a given application. In this paper, we restrict ourselves to a subclass of the entire class of IFAMs, namely the subclass of IFAMs that are associated with the Yager family of parameterized t-norms. For simplicity, we speak of the class of Yager IFAMs. In this setting, we formulate the problem of choosing the best Yager IFAM for a given application as an optimization problem. Considering two problems in time series prediction from the literature, we solve this optimization problem and compare the performance of the resulting Yager IFAM with the performances of other fuzzy, neural, neuro-fuzzy, and statistical techniques.*

Keywords— Fuzzy associative memory, implicative fuzzy associative memory, Yager family of parameterized t-norms, time-series prediction, hydroelectric plant, monthly streamflow prediction.

1 Introduction

Implicative fuzzy associative memories (IFAMs) belong to the class of *fuzzy morphological associative memories* (FMAMs) [1, 2]. The theoretical background for FMAMs can be found in *fuzzy mathematical morphology* (FMM) [3, 4]. More precisely, FMAMs can be viewed as fuzzy neural networks whose neurons perform elementary operations of fuzzy mathematical morphology [2].

Recently, Sussner and Valle [5] showed that many well-known fuzzy associative memory (FAM) models such as the FAMs of Kosko, Junbo et al., Liu, and Bělohlavék [6, 7, 8, 9] represent particular instances of FMAMs. In this paper, we focus on the FMAM subclass of IFAMs. Upon presentation of an input pattern $\mathbf{x}$, an IFAM model performs max-T products, where T is a t-norm, at every node. A certain continuous t-norm determines a particular IFAM model. For example, the Lukasiewicz IFAM is associated with the Lukasiewicz t-norm. Learning in IFAM models occurs by forming the R-implication of the underlying t-norm and by computing a matrix product that corresponds to the Bandler-Kohout superproduct [10, 11]. In the context of fuzzy associative memories, we speak of *implicative fuzzy learning*. As we have pointed out in previous articles, implicative fuzzy learning is based on the greatest solution of a fuzzy relational inequality [1, 12, 13]. In the general setting of FMAMs, implicative fuzzy learning generalizes to *learning by adjunction*, a recording strategy that was derived from a concept in mathematical morphology known as the duality relationship of adjunction [2, 14].

We have successfully applied IFAMs to some problems in time series forecasting, in particular a problem of forecasting the demand of man power in steel manufacturing industry in the state of West Bengal, India, and a problem of stream flow prediction in a hydroelectric plant in southeastern Brazil [5, 15]. In these prediction problems, the Lukasiewicz IFAM exhibited the best performance compared to a variety of other FAM models and to a number of statistical, neural, fuzzy, and neuro-fuzzy approaches from the literature. The goal of this paper is to optimize our results by considering not only a finite number of IFAMs but an infinite number of IFAM models.

To this end, we introduce the class of Yager IFAMs, a class of parameterized IFAM models depending on a single parameter. We simply define a Yager IFAMs as an IFAM that is associated with a Yager t-norm [16]. Recall that a Yager t-norm is of the form $T^d(x, y) = 1 - \{1 \wedge [(1-x)^d + (1-y)^d]^{1/d}\}$, where the symbol $\wedge$ denotes the minimum operation and where $d > 0$.

Thus, we can formulate an optimization problem depending on the parameter d that consists in minimizing the distance between the target pattern and the pattern produced by the IFAM corresponding to the parameter d. As a result of this optimization process, we obtain the Yager IFAM (or the set of Yager IFAMs) that provides the best fit with respect to the training data.

The paper is organized as follows. After presenting a brief introduction to IFAMs, we introduce the Yager class of parameterized IFAMs in Section 3. Subsequently, we discuss the aforementioned prediction problems and determine the Yager IFAMs that minimize the mean squared distance between the outputs produced by the memory cues $\mathbf{x}^\xi$ and the desired outputs. We finish the paper with some concluding remarks and suggestions for further research.

2 A Brief Introduction to Implicative Fuzzy Associative Memories

Implicative Fuzzy Associative Memories (IFAMs) can be defined in terms of certain matrix products, namely the max-T product and the min-I product, where T is a t-norm and I is a fuzzy implication [2]. We will make use of the following

ISBN: 978-989-95079-6-8

t-norms:

$$T_M(x,y) = x \wedge y\,, \qquad \text{(Minimum)} \quad (1)$$

$$T_P(x,y) = x \cdot y\,, \qquad \text{(Product)} \quad (2)$$

$$T_L(x,y) = 0 \vee (x+y-1)\,, \qquad \text{(Lukasiewicz)} \quad (3)$$

$$T_W(x,y) = \begin{cases} x \wedge y, & x \vee y = 1\,, \\ 0, & x \vee y < 1\,. \end{cases} \qquad \text{(Nilpotent Min.)} \quad (4)$$

Reverse fuzzy implications can be easily derived from fuzzy implications. Recall the following popular implications:

$$I_M(x,y) = \begin{cases} 1, & x \le y \\ y, & x > y \end{cases}\,, \qquad \text{(Gödel)} \quad (5)$$

$$I_P(x,y) = \begin{cases} 1, & x \le y \\ y/x, & x > y \end{cases}\,, \qquad \text{(Product)} \quad (6)$$

$$I_L(x,y) = 1 \wedge (y - x + 1)\,. \qquad \text{(Lukasiewicz)} \quad (7)$$

Let $A \in [0,1]^{m\times p}$ and $B \in [0,1]^{p\times n}$. The max-T product of A and B is given by the matrix $C = A \circ B$ and the min-I product of A and B is given by the matrix $D = A \circledast B$, where C and D are defined as follows (note that the min-I product of A and B can also be viewed as a Bandler-Kohout superproduct [10, 11]).

$$c_{ij} = \bigvee_{k=1}^{p} C(a_{ik}, b_{kj}) \;\text{ and }\; d_{ij} = \bigwedge_{k=1}^{p} I(b_{kj}, a_{ik})\,. \quad (8)$$

A FAM is a fuzzy neural network that is designed to store a fundamental memory set, i.e., a set of associations $\{(\mathbf{x}^\xi, \mathbf{y}^\xi) : \xi = 1,\ldots,p\}$, where $\mathbf{x}^\xi \in [0,1]^n$ and $\mathbf{y}^\xi \in [0,1]^m$. This task can be achieved by means of a synaptic weight matrix $W \in [0,1]^{m\times n}$. Let $X = [\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^p] \in [0,1]^{n\times p}$ denote the matrix whose columns are the input patterns and let $Y = [\mathbf{y}^1, \mathbf{y}^2, \ldots, \mathbf{y}^p] \in [0,1]^{m\times p}$ denote the matrix whose columns are the output patterns. We suggested the following rule for synthesizing the weight matrix W of an IFAM [1, 12, 13]:

$$W = Y \circledast X^T\,. \quad (9)$$

Note that (9) depends on the choice of a fuzzy implication. In an IFAM model, we require that the underlying fuzzy implication I is the R-implication of a continuous t-norm T. In other words, the corresponding fuzzy implication I must be adjoint to a continuous t-norm T [2, 4]. For example, the pairs (I_M, T_M), (I_P, T_P), and (I_L, T_L) represent pairs of adjoint operators.

There are infinitely many IFAM models. In the special cases where the fuzzy implications occurring in (9) are I_M, I_P, and I_L, we speak of the Gödel IFAM, the Goguen IFAM, and the Lukasiewicz IFAM, respectively.

Once the recording phase has been completed, the IFAM weight matrix can be applied to an arbitrary input pattern $\mathbf{x} \in [0,1]^n$. If θ denotes a threshold or bias vector that is given by the entry-wise minimum over all $\mathbf{y}^\xi$, where $\xi = 1,\ldots,p$, then we obtain the following output pattern $\mathbf{y} \in [0,1]^m$:

$$\mathbf{y} = (W \circ \mathbf{x}) \vee \boldsymbol{\theta}\,, \qquad \text{where} \qquad \boldsymbol{\theta} = \bigwedge_{\xi=1}^{p} \mathbf{y}^\xi\,. \quad (10)$$

Here, the t-norm T occurring in the max-T product is such that its R-implication I is associated with the product "$\circledast$" of (9). In other words, the t-norm T of (10) and the fuzzy implication I of (9) are adjoint [2, 4].

3 The Yager Class of Parameterized IFAMs

Note that a particular IFAM model is determined by a continuous t-norm. An entire subclass of IFAM models is given by the class of Yager t-norms or intersections which are obviously continuous. Recall that the class of Yager t-norms represents a family of parameterized t-norms given by the following equation [16]:

$$T^d(x,y) = 1 - \left\{1 \wedge [(1-x)^d + (1-y)^d]^{1/d}\right\},\; d > 0\,. \quad (11)$$

For $d = 1$, the Yager t-norm coincides with the Lukasiewicz t-norm T_L. For $t \to 0$ the Yager t-norm approaches the nilpotent minimum T_W, which is the pointwise smallest t-norm. For $t \to \infty$ the Yager t-norm approaches the minimum t-norm T_M, the pointwise largest t-norm [17, 18].

For $d > 0$, the symbol I^d denotes the R-implication of T^d. The operator I^d can be computed as follows:

$$I^d(x,y) = 1 - \left\{0 \vee [(1-y)^d - (1-x)^d]\right\}^{1/d},\; d > 0\,. \quad (12)$$

The resulting IFAMs are called Yager IFAMs. These parameterized IFAM models will be considered in the applications of the next section. In fact, we will tackle the problem of choosing the best Yager IFAM for a given application in terms of an optimization problem.

To this end, let us consider the fundamental memory set $\{(\mathbf{x}^\xi, \mathbf{y}^\xi) : \xi = 1,\ldots,p\}$ which can be viewed as the set of training patterns. Let W^d denote the synaptic weight matrix of the Yager IFAM with parameter $d > 0$. A possible approach towards finding the best Yager IFAM consists in minimizing the distances between $(W^d \circ^d \mathbf{x}^\xi) \vee \theta$ and $\mathbf{y}^\xi$ for $\xi = 1,\ldots,p$, where the symbol "$\circ^d$" denotes the max-T^d product. This optimization problem can be formulated as follows:

$$\begin{cases} \text{minimize } \sum_{\xi=1}^{p} \left\|\mathbf{y}^\xi - [(W^d \circ^d \mathbf{x}^\xi) \vee \theta]\right\|_2 \\ \text{subject to } \quad d > 0 \end{cases}\,. \quad (13)$$

If $\mathbf{y}^{\xi,d} = (W \circ^d \mathbf{x}^\xi) \vee \theta$ and Y^d is the matrix $Y^d = [\mathbf{y}^{1,d}, \ldots \mathbf{y}^{p,d}]$ then the optimization problem of (13) can be expressed in terms of the Frobenius norm as follows:

$$\begin{cases} \text{minimize } \left\|Y - Y^d\right\|_F \\ \text{subject to } \quad d > 0 \end{cases}\,. \quad (14)$$

In the applications as fuzzy rule-based systems that are described below, an input fuzzy set $\mathbf{x} \in [0,1]^n$ is derived from a real-valued input vector $\mathbf{v}$ and the respective output fuzzy set $\mathbf{y} = (W^d \circ \mathbf{x}) \vee \theta$ is defuzzified yielding a real-valued output $s = defuzz(\mathbf{y})$. The training patterns are given in the form $(\mathbf{v}^1, s^1), \ldots, (\mathbf{v}^p, s^p)$. Let $s^{\xi,d}$ denote the defuzzification of $\mathbf{y}^{\xi,d}$, i.e. $s^{\xi,d} = defuzz(\mathbf{y}^{\xi,d})$ for all $\xi = 1,\ldots,p$. If we have $\mathbf{s} = (s^1,\ldots,s^p)^T$ and $\mathbf{s}^d = (s^{1,d},\ldots,s^{p,d})^T$ then we choose to minimize the following expression instead of (14):

ISBN: 978-989-95079-6-8

$$\begin{cases} \text{minimize } \left\|\mathbf{s}-\mathbf{s}^d\right\|_2 \\ \text{subject to } \quad d>0 \end{cases} . \tag{15}$$

In other words, we minimize the Euclidean distance between the desired results $\mathbf{s}$ and the results produced by the Yager IFAMs for $d > 0$. To this end, we applied the routine FMINBND of MATLAB's Optimization Toolbox in order to determine a local minimum of the real-valued objective function $f(d) = \left\|\mathbf{s}-\mathbf{s}^d\right\|_2$ in an interval. Note that plotting the objective function $f : \mathbb{R} \to \mathbb{R}$ allows us to select an interval $[x_1, x_2]$ such that an application of FMINBND yields a candidate for a global minimum. The resulting parameter d yields the Yager IFAM that exhibits the best performance on the training data. The next sections describes applications of this strategy to two forecasting problems that can be found in the literature.

4 Applications of Yager IFAMs in Prediction

Fuzzy associative memories such as IFAMs can be used to implement mappings of fuzzy rules. In this case, a set of rules in the form of human-like IF-THEN conditional statements are stored. In this subsection, we consider two problems in time-series prediction. We use (15) to select the Yager IFAM that produces the least MSE on the training data.

4.1 Prediction of Manpower Requirement in Steel Manufacturing

Let us consider the problem of predicting the manpower requirement in steel manufacturing industry in the state of West Bengal, India [19]. Initially, we have five linguistic values A_i, $i = 1, \ldots, 5$ and a set of fuzzy conditional statements such as "If the manpower of year n is A_i, then that of year $n + 1$ is A_j". The linguistic values A_i correspond to fuzzy sets. Table 1 shows the set of input-output pairs that we stored in a number of different FAM models including the Gödel, the Goguen, the Lukasiewicz, and the Yager IFAMs with $d > 0$.

If W is the synaptic weight matrix of an IFAM model and $\boldsymbol{\theta}$ is the threshold obtained after the learning process, then the predicted manpower of year $n + 1$ is given by the following equation:

$$A_{n+1} = (W \circ A_n) \vee \boldsymbol{\theta} , \tag{16}$$

where A_n is the manpower of year n and $\circ$ is the max-t composition. Note that here we have $\boldsymbol{\theta} = (0, 0, 0, 0, 0)^T$ and therefore

$$A_{n+1} = W \circ A_n . \tag{17}$$

The fuzzy set A_n is computed by fuzzifying the numerical input value v corresponding to the manpower requirement of year n according to the method described by Choudhury et. al. [19]. We calculated the predicted value s of the required manpower for year $n + 1$ by applying (17) to A_n and by defuzzifying the result A_{n+1} using the "mean of maxima" (MOM) scheme.

Choudhury et. al. used this problem to compare the average or mean percentage error (MPE) produced by Kosko's FAM [6, 20] and by the statistical methods ARIMA1 and ARIMA2 [21, 22]. In a recent paper [1], we included some IFAM models, in particular the Lukasiewicz IFAM, as well as the Lukasiewicz Generalized FAM (GFAM) of Chung and Lee [23] and the max-min FAM with threshold of Liu [8] in this comparison. The Lukasiewcz IFAM, which is closely related to the original morphological associative memory (MAM) [1, 24, 25], exhibited the best performance among all the models mentioned above. Recall that the Lukasiewicz IFAM coincides with the Yager IFAM for $d = 1$.

Table 1: Set of input and output pairs used in the forecasting application

ξ	$\mathbf{x}^\xi$	$\mathbf{y}^\xi$
1	$[1.0, 0.5, 0, 0, 0]^T$	$[0.5, 1.0, 0.5, 0, 0]^T$
2	$[0.5, 1.0, 0.5, 0, 0]^T$	$[0.5, 1.0, 0.5, 0, 0]^T$
3	$[0.5, 1.0, 0.5, 0, 0]^T$	$[0, 0.5, 1.0, 0.5, 0]^T$
4	$[0, 0.5, 1.0, 0.5, 0]^T$	$[0.5, 1.0, 0.5, 0, 0]^T$
5	$[0, 0.5, 1.0, 0.5, 0]^T$	$[0, 0.5, 1.0, 0.5, 0]^T$
6	$[0, 0.5, 1.0, 0.5, 0]^T$	$[0, 0, 0.5, 1.0, 0.5]^T$
7	$[0, 0, 0.5, 1.0, 0.5]^T$	$[0, 0, 0.5, 1.0, 0.5]^T$
8	$[0, 0, 0.5, 1.0, 0.5]^T$	$[0, 0, 0, 0.5, 1.0]^T$
9	$[0, 0, 0, 0.5, 1.0]^T$	$[0, 0, 0, 0.5, 1.0]^T$

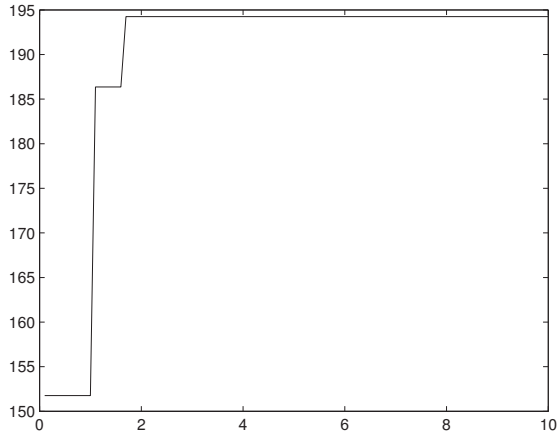


Figure 1: The distance $\left\|\mathbf{s}-\mathbf{s}^d\right\|_2$ between the actual demand of manpower between 1984 and 1995 and the demand of manpower predicted by the Yager IFAM as a function of d.

In this paper, we determine the best Yager IFAM for this prediction problem by solving the optimization problem of (15). Since the Lukasiewicz IFAM is taken into account in the optimization process, the IFAMs representing solutions of (15) are guaranteed to perform at least as well as the Lukasiewicz IFAM with respect to the MSE. In fact, the minimum mean squared error (MSE) is adopted for any value of d in the interval $(0, 1]$ as shown in Figure 1.

Table 2 presents a comparison of the FAM models and the statistical autoregressive integrated moving average (ARIMA) methods in terms of the MSE, the MPE, and the mean absolute error (MAE). Figure 2 provides for a graphical interpretation of the predictions.

The results of this experiment indicate the utility of IFAMs for prediction problems and the validity of our approach for determining the best Yager IFAM. However - as the reader may have noticed - this prediction problem does not include

ISBN: 978-989-95079-6-8

Table 2: MSE, MAE and MPE produced by the prediction models

Method	MSE ($\times 10^5$)	MAE (m^3/s)	MPE (%)
Yager IFAM ($d \in (0,1]$):	1.92	32.75	2.29
Lukasiewicz IFAM:	1.92	32.75	2.29
Kosko's FAM:	2.57	38.75	2.67
Lukasiewicz GFAM:	2.57	38.75	2.67
Gödel IFAM:	2.89	38.58	2.73
Max-min FAM/threshold:	2.89	38.58	2.73
Goguen IFAM:	3.14	42.75	2.99
ARIMA2	9.36	83.55	5.48
ARIMA1	23.26	145.25	9.79

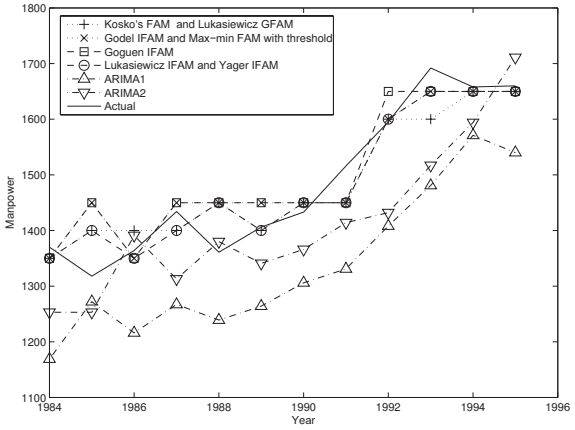


Figure 2: Comparison in forecasting manpower. The continuous line represents the actual manpower. The dashed line marked by '$\circ$' corresponds to the Lukasiewicz IFAM model, i.e., the Yager IFAM for $d = 1$ and the dashed line marked by '$\square$' corresponds to the Goguen IFAM. The dotted line marked by '$\times$' refers to Kosko's FAM model as well as the Lukasiewicz Generalized FAM, and the dotted line marked by '$+$' refers to the max-min FAM with threshold and the Gódel IFAM. The lines marked by '$\triangle$' and '$\triangledown$' represent ARIMA1 and ARIMA2.

any test data that would allow us to assess the performance of the models for years beyond the year of 1995. More realistic problems such as the one described below consist of training data and test data. In this case, the optimization should be performed on the training data and the test data should serve to validate this approach (possibly, separate validation data can be extracted from the training data and the validation data can be used to choose the best model).

4.2 Prediction of Average Monthly Streamflow of a Hydroelectric Plant

Furnas is a large hydroelectric plant that is located in southeastern Brazil. Magalhaẽs et al. as well as Sussner and Valle have previously discussed the problem of forecasting the average monthly streamflow [26, 15]. The seasonality of the monthly streamflow suggests the use of 12 different models, one for each month of the year. The training data correspond to the years $1931 - 1990$ and the test data correspond to the years $1991 - 1998$.

Let s^ξ, where $\xi = 1, \ldots, p$, be samples of a seasonal streamflow time series. The goal is to estimate the value of s^γ from a subsequence of $(s^1, s^2, \ldots, s^{\gamma-1})$. Here, we employ subsequences that correspond to a vector of the form

$$\mathbf{v}^\gamma = (s_{\gamma-h}, \ldots, s_{\gamma-1})^T, \tag{18}$$

where $h \in \{1, 2, \ldots, \gamma - 1\}$. In this experiment, we employed a fixed number of three antecedents in our IFAM models. For example, only the values of January, February, and March were taken into account to predict the streamflow of April.

The uncertainty that is inherent in hydrological data suggests the use of fuzzy sets to model the streamflow samples. For $\xi < \gamma$, a fuzzification of $\mathbf{v}^\xi$ and s^ξ using Gaussian membership functions yields fuzzy sets $\mathbf{x}^\xi : \mathcal{U} \to [0, 1]$ and $\mathbf{y}^\xi : \mathcal{V} \to [0, 1]$, respectively, where $\mathcal{U}$ and $\mathcal{V}$ represent finite universes of discourse. A subset S of the resulting input-output pairs $(\mathbf{x}^\xi, \mathbf{y}^\xi)$, where $\xi = 1, \ldots, p$, is implicitly stored in an IFAM model (we only construct the parts of the weight matrix that are actually used in the recall phase). We employed the *subtractive clustering method* to determine the set S [27]. The centers and the widths of the Gaussian-type membership functions of the input patterns $\mathbf{x}^\xi$ and the output patterns $\mathbf{y}^\xi$ in S were estimated by means of the MATLAB function `subclust`. Here, we used a constant width of $r = 0.5$.

Upon presentation of the input pattern $\mathbf{x}^\gamma$, the IFAM with parameter d yields the corresponding output pattern $\mathbf{y}^{\gamma,d}$. For computational reasons, $\mathbf{x}^\gamma$ is modeled as a discrete Dirac-δ (impulse) function. A defuzzification of $\mathbf{y}^{\gamma,d}$ using the centroid method yields $s^{\gamma,d}$.

As before, we generated the vectors $\mathbf{s} = (s^1, \ldots, s^p)^T$ and $\mathbf{s}^d = (s^{1,d}, \ldots, s^{p,d})^T$. We employed the MATLAB function FMINBND to solve the optimization problem given by (15). This optimization process resulted in the parameter $d = 4.3568$. Figure 3 depicts $\left\|\mathbf{s} - \mathbf{s}^d\right\|_2$ for $d \in (0, 10]$. For the training data concerning the years $1931 - 1990$, the MSE of the Yager IFAM with $d = 4.3568$ is 12087.3 whereas the MSE for the Lukasiewicz IFAM, i.e., the Yager IFAM with $d = 1$, is 12387.4.

Table 3 provides the MSEs, the MPEs, and the MAEs produced by some IFAMs and other models during the testing phase. The values of the periodic auto-regressive moving average model (PARMA) [21], the fuzzy neural network NFN [28], and the predictive fuzzy clustering method FPM-PRP were drawn from the literature [26].

Figure 4 shows the forecasted streamflows estimated by the prediction model based on the FMAM for the Furnas reservoir from 1991 to 1998. Table 3 compares the errors that were generated by the IFAMs and several other models [26]. In contrast to the IFAM models, the MLP, NFN, and FPM-PRP models were initialized by optimizing the number of the parameters for each monthly prediction. For example, the MLP considers 4 antecedents to predict the streamflow of January and 3 antecedents to predict the streamflow for February. Moreover, the FPM-PRP model also takes advantage of slope information which requires some additional "fine tuning". Nevertheless, the Yager IFAM resulting from the min-

ISBN: 978-989-95079-6-8

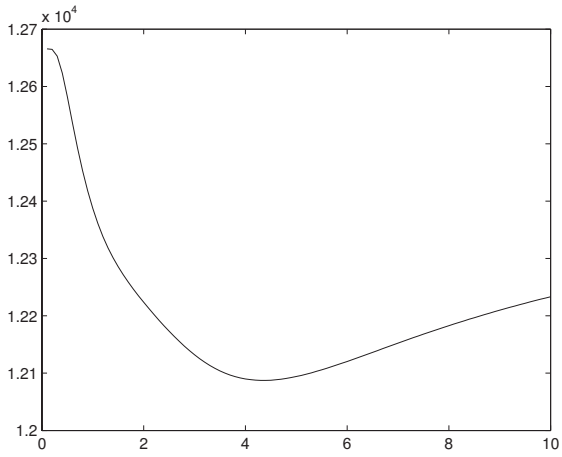


Figure 3: The distance $\left\|\mathbf{s}-\mathbf{s}^{d}\right\|_{2}$ between the actual streamflows given by the training data and the streamflows predicted by the Yager IFAM as a function of d.

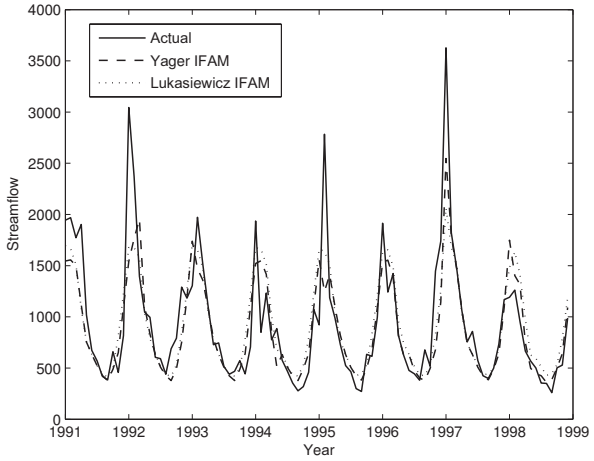


Figure 4: The streamflow prediction for the Furnas reservoir from 1991 to 1998. The continuous line corresponds to the actual values and the dashed line corresponds to the prediction of the Yager IFAM with $d = 4.3568$. The dotted line refers to the values predicted by the Lukasiewcz IFAM.

Table 3: Mean square, mean absolute, and mean relative percentage errors produced by the prediction models.

Method	MSE ($\times 10^5$)	MAE (m^3/s)	MPE (%)
FPM-PRP	1.20	200	18
Yager IFAM ($d = 4.3568$)	1.28	216	21
Lukasiewicz IFAM	1.27	229	24
PARMA	1.85	280	28
MLP	1.82	271	30
NFN	1.73	234	20

imization performed in (15) and the Lukasiewicz IFAM produced very satisfactory predictions that are visualized in Figure 4. Note that in the testing phase the MSE produced by the Yager IFAM with $d = 4.3568$ is slightly higher than the one produced by the Lukasiewicz IFAM although the Yager IFAM with $d = 4.3568$ outperforms the Lukasiewicz IFAM in terms of MAE and MPE. This fact may be due to overfitting to the training data. With some fine tuning of the parameters, it is possible to experimentally determine a Lukasiewicz IFAM (and thus a Yager IFAM) that outperforms the FPM-PRP with respect to the three error measures MSE, MAE, and MPE [15]. In our opinion, however, the fine tuning of the parameters should preferably not be performed experimentally but should be part of the optimization process.

5 Concluding Remarks

This paper represents the first attempt of tackling the problem of selecting the best IFAM for a given application. To this end, we introduced parameterized IFAM models, specifically the Yager class of IFAMs, and we formulated the problem of determining the Yager IFAM that yields the best fit to the training data as an optimization problem.

Recall that an IFAM is uniquely determined by a continuous t-norm, such as a Yager t-norm T^d with parameter d for $0 < d < \infty$, which gives rise to an R-implication. Although other continuous parameterized t-norms could have been chosen [29], the Yager t-norms T^d have the advantage that T^d approaches T_W, the pointwise smallest t-norm, for $d \to 0$ and T^d approaches T_M, the pointwise largest t-norm, for $d \to \infty$. Moreover, T^1 equals the Lukasiewicz t-norm which implies that the Lukasiewicz IFAM, i.e., the IFAM that exhibited the best performance in our previous experiments, is taken into account in the optimization process.

We applied our approach to two prediction problems from the literature. In the first problem concerning the prediction of the manpower requirement in steel manufacturing industry in West Bengal, India, only training data are available. The minimization of the MSE for the Yager IFAMs resulted in an infinite number of Yager IFAMs corresponding to $d \in (0, 1]$ that exhibit the same errors on the training data as the Lukasiewicz IFAM ($d = 1$). For the streamflow prediction problem in a major Brazilian hydroelectric plant, we dispose of training data and test data. The optimization process generated a Yager IFAM that outperforms the Lukasiewicz IFAM with respect to the MSE on the training data but exhibits a slightly higher MSE on the test data. This result may be due to overfitting and in fact we know that better Yager IFAMs exist for this problem. Nevertheless, the particular Yager IFAM derived from the optimization process exhibits very satisfactory results (which are better than those of the Lukasiewicz IFAM with respect to the MAE and MPE error measures), especially in view of the fact that we used a fixed number of antecedents and other parameters in conjunction with the Yager IFAM.

In the near future, we intend to incorporate the selection of additional parameters into the optimization process and we intend to use validation and/or regularization techniques to avoid overfitting. Furthermore, we plan to investigate other classes of parameterized IFAMs that are based on other types of parameterized t-norms.

ISBN: 978-989-95079-6-8

References

[1] P. Sussner and M. Valle. Implicative fuzzy associative memories. *IEEE Transactions on Fuzzy Systems*, 14(6):793–807, 2006.

[2] M. Valle and P. Sussner. A general framework for fuzzy morphological associative memories. *Fuzzy Sets and Systems*, 159(7):747–768, 2008.

[3] M. Nachtegael and E. Kerre. Connections between binary, gray-scale and fuzzy mathematical morphologies. *Fuzzy Sets and Systems*, 124(1):73–85, Nov. 2001.

[4] P. Sussner and M. Valle. Classification of fuzzy mathematical morphologies based on concepts of inclusion measure and duality. *Journal of Mathematical Imaging and Vision*, 32(2):139–159, 2008.

[5] P. Sussner and M. Valle. Fuzzy associative memories and their relationship to mathematical morphology. In W. Pedrycz, A. Skowron, and V. Kreinovich, editors, *Handbook of Granular Computing*, chapter 33. John Wiley and Sons, Inc., New York, 2008.

[6] B. Kosko. *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*. Prentice Hall, Englewood Cliffs, N.J., 1992.

[7] F. Junbo, J. Fan, and S. Yan. A learning rule for fuzzy associative memories. In *Proceedings of the IEEE International Joint Conference on Neural Networks*, volume 7, pages 4273 – 4277, June 1994.

[8] P. Liu. The fuzzy associative memory of max-min fuzzy neural networks with threshold. *Fuzzy Sets and Systems*, 107:147–157, 1999.

[9] R. Belohlávek. Fuzzy logical bidirectional associative memory. *Information Sciences*, 128(1-2):91–103, 2000.

[10] W. Bandler and L. Kohout. Fuzzy relational products and fuzzy implication operators. In *International Workshop on Fuzzy Reasoning Theory and Applications*, Queen Mary College, University of London, London, 1978.

[11] L. Běhounek and M. Daňková. Relational compositions in fuzzy class theory. *Fuzzy Sets and Systems*, 160(8):1005–1036, April 2009.

[12] J. Drewniak. Fuzzy relation equations and inequalities. *Fuzzy Sets and Systems*, 14(3):237–247, December 1984.

[13] M. Valle, P. Sussner, and F. Gomide. Introduction to implicative fuzzy associative memories. In *Proceedings of the IEEE International Joint Conference on Neural Networks*, pages 925 – 931, Budapest, Hungary, July 2004.

[14] H. Heijmans. *Morphological Image Operators*. Academic Press, New York, NY, 1994.

[15] P. Sussner and M. Valle. Morphological and certain fuzzy morphological associative memories for classification and prediction. In V. Kaburlassos and G. Ritter, editors, *Computational Intelligence Based on Lattice Theory*, volume 67, pages 149 – 173. Springer Verlag, Heidelberg, Germany, 2007.

[16] R. Yager. On a general class of fuzzy connectives. *Fuzzy Sets and Systems*, 4(3):235–242, 1980.

[17] G. Klir and T. Folger. *Fuzzy Sets, Uncertainty and Information*. Prentice Hall, Englewood Cliffs, NJ, 1988.

[18] C. Lin and C. Lee. *Neural Fuzzy Systems*. Prentice Hall, Upper Saddle River, NJ, 1996.

[19] J. Choudhury, B. Sarkar, and S. Mukherjee. Forecasting of engineering manpower through fuzzy associative memory neural network with arima: a comparative study. *Neurocomputing*, 47:241–257, 2002.

[20] B. Kosko. *Fuzzy Engineering*. Prentice Hall, Upper Saddle River, N.J., 1992.

[21] J. Box, G. Jenkins, and C. Reinsel. *Time Series Analysis, Forecasting, and Control*. Holden-Day, Oakland, CA, third edition, 1994.

[22] T. Mills. *Time Series Techniques for Economists*. Cambridge University Press, Cambridge, United Kingdom, 1990.

[23] F. Chung and T. Lee. On fuzzy associative memory with multiple-rule storage capacity. *IEEE Transactions on Fuzzy Systems*, 4(3):375–384, August 1996.

[24] G. Ritter, P. Sussner, and J. Diaz de Leon. Morphological associative memories. *IEEE Transactions on Neural Networks*, 9(2):281–293, 1998.

[25] P. Sussner and M. Valle. Grayscale morphological associative memories. *IEEE Transactions on Neural Networks*, 17(3):559–570, May 2006.

[26] M. Magalhães, R. Ballini, R. Gonçalves, and F. Gomide. Predictive fuzzy clustering model for natural streamflow forecasting. In *Proceedings of the IEEE International Conference on Fuzzy Systems*, pages 390–394, Budapest, Hungary, 2004.

[27] S. Chiu. Fuzzy model identification based on cluster estimation. *Journal of Intelligent and Fuzzy Systems*, 1994.

[28] M. Figueiredo and F. Gomide. Adaptive neuro-fuzzy modeling. *Proceedings of the Sixth IEEE International Conference on Fuzzy Systems*, 3:1567–1572, 1997.

[29] E. Klement, R. Mesiar, and E. Pap. *Triangular Norms*. Kluwer Academic Publishers, 2001.

ISBN: 978-989-95079-6-8

Defining A Fuzzy Partition for Coarseness Modelling in Texture Images

J. Chamorro-Martínez, P. Martínez-Jiménez[1] J.M. Soto-Hidalgo [2]

1.Department of Computer Science and Artificial Intelligence, University of Granada
Granada, Spain
2.Department of Computer Architecture, Electronics and Electronic Technology, University of Córdoba
Córdoba, Spain
Email: {jesus,pedromartinez}@decsai.ugr.es, jmsoto@uco.es

Abstract— *In this paper, the texture feature "coarseness" is modelled by means of a fuzzy partition on the domain of coarseness measures. The number of linguistic labels to be used, and the parameters of the membership functions associated to each fuzzy set are calculated relating representative coarseness measures (our reference set) with the human perception of this texture property. A wide variety of measures is studied, analyzing its capability to discriminate different coarseness categories. Data about the human perception of fineness is collected by means of a pools, performing an aggregation of their assessments by means of OWA operators. This information is used to obtain a fuzzy partition adapted to the human perception of coarseness-fineness*

Keywords— Coarseness, fineness, fuzzy partition, fuzzy texture, image features, texture features.

1 Introduction

For analyzing an image several kind of features can be used. From all of them, texture is one of the most popular and, in addition, one of the most difficult to characterize due to its imprecision. For describing texture, humans use vague textural properties like *coarseness/fineness*, *orientation* or *regularity* [1, 2]. From all of them, the *coarseness/fineness* is the most common one, being usual to associate the presence of fineness with the presence of texture. In this framework, a *fine* texture corresponds to small texture primitives (e.g. the image in figure 1(A)), whereas a *coarse* texture corresponds to bigger primitives (e.g. the image in figure 1(I)).

There are many measures in the literature that, given an image, capture the fineness (or coarseness) presence in the sense that the greater the value given by the measure, the greater the perception of texture [3]. However, given a certain measure value, there is not an immediate way to decide whether there is a fine texture, a coarse texture or something intermediate; in other words, there is not a textural interpretation.

To face this problem, fuzzy logic has been recently employed for representing the imprecision related to texture. In many of these approaches, fuzzy logic is usually applied just during the process, being the output a crisp result [4, 5]. Other approaches try to model the texture and its semantic by means of fuzzy sets defined on the domain of a given texture measure. In this last framework, some proposals model the texture property by means of an unique fuzzy set [6], and other approaches define fuzzy partitions providing a set of linguistic terms [7, 8].

Focusing our study in the last type of approaches, two questions need to be faced for defining properly a fuzzy partition: (i) the number of linguistic labels to be used, and (i) the parameters of the membership functions associated to each fuzzy set (and, consequently, the kernel localization). However, these question are not treat properly in the literature. Firstly, the number of fuzzy sets are often chosen arbitrarily, without take into account the capability of each measure to discriminate between different categories. Secondly, in many of the approaches, just an uniform distribution of the fuzzy sets is performed on the domain of the measures, although is wellknown that measure values corresponding to representative labels are not distributed uniformly. In addition, from our knowledge, none of the fuzzy approaches in the literature considers the relationship between the computational feature and the human perception of texture, so the labels and the membership degrees do not necessarily will match with the human assessments.

In this paper, we propose a fuzzy partition taking into account the previous questions. Firstly, in order to select the number of linguistic labels, we analyze the ability of each measure to discriminate different coarseness categories. For this purpose, data about the human perception of fineness is collected by means of a pools. This information is also used to localize the position and size of the kernel of each fuzzy set, obtaining a fuzzy partition adapted to the human perception of coarseness-fineness

The rest of the paper is organized as follows. In section 2 we present our methodology to obtain the fuzzy partition. Results are shown in section 3, and the main conclusions and future work are sumarized in section 4.

2 Fuzzy Partitions for Coarseness

As it was pointed, there is not a clear perceptual interpretation of the value given by a fineness measure. To face this problem, we propose to define a fuzzy partition on the domain of a given fineness measure. For this purpose, several questions will be faced: (i) what reference set should be used for the fuzzy partition, (ii) how many fuzzy sets will compound the partition, and (ii) how to obtain the membership functions for each fuzzy set.

Concern to the reference set, we will define the partition on the domain of a given coarseness-fineness measure. From now on, we will note $\mathcal{P} = \{P_1, \ldots, P_K\}$ the set of K measures analyzed in this paper, Π_k the partition defined on the domain of $\mathcal{P}_k$, N_k the number of fuzzy sets which compounds the partition Π_k, and $\mathcal{T}_k^i$ the i-th fuzzy set in Π_k. In this paper, the set $\mathcal{P} = \{P_1, \ldots, P_K\}$ is formed by the $K = 17$ mea-

ISBN: 978-989-95079-6-8



Figure 1: Some examples of images with different degrees of fineness

sures shown in the first column of table 1. It includes classical statistical measures, frequency domain approaches, fractal dimension analysis, etc. All of them are automatically computed from the texture image.

With regard to the number of fuzzy sets which compounds the partition, we will analyze the ability of each measure to distinguish between different degrees of fineness. This analysis will based on how the human perceives the fineness-coarseness. To get information about human perception of fineness, a set of images covering different degrees of fineness will be gathered. This images will be used to collect, by means of a pool, human assessments about the perceived fineness. From now on, let $\mathcal{I} = \{I_1, \ldots, I_N\}$ be the set of N images representing fineness-coarseness examples, and let $\Gamma = \{v^1, \ldots, v^N\}$ be the set of perceived fineness values associated to $\mathcal{I}$, with v^i being the value representing the degree of fineness perceived by humans in the image $I_i \in \mathcal{I}$. The way to obtain Γ will be describe in section 2.1

Using the data about human perception, and the measures values obtained for each image $I_i \in \mathcal{I}$, we will apply a set of multiple comparison tests in order to obtain the number of fineness degrees that each measure can discriminate (section 2.2). In addition, with the information given by the tests, we will define the fuzzy sets which will compund the partition (2.3).

2.1 *Assessment collection*

In this section, the way to obtain the set $\Gamma = \{v1, \ldots, v^N\}$ of perceived fineness values associated to $\mathcal{I}$ will be described. For this purpose, firstly the image set $\mathcal{I}$ will be selected. After that, a poll for getting assessments about the perception of fineness will be designed. Finally, for a given image, the assessments of the different subjects will be aggregated.

2.1.1 *The texture image set*

A set $\mathcal{I} = \{I_1, \ldots, I_N\}$ of $N = 80$ images representative of the concept of *fineness* has been selected. Figure 1 shows some images extracted from the set $\mathcal{I}$. Such set has been selected satisfying the following properties:

1. It covers the different presence degrees of fineness.
2. The number of images for each presence degree is representative enough.
3. Each image shows, as far as possible, just one presence degree of fineness.

Due to the third property, each image can viewed as "homogeneous" respect to the fineness degree represented, i.e., if we select two random windows (with a dimension which does not "break" the original texture primitives and structure), the perceived fineness will be the same for each window (and also respect to the original image). In other words, we can see each image $I_i \in \mathcal{I}$ as a set of lower dimension images (windows) with the same fineness degree of the original one.

As we explained, given an image $I_i \in \mathcal{I}$, a set of measures $\mathcal{P}$ will be applied on it. In fact, and thanks to the third property, we really can apply these measures for each subimage, assuming that the human assessment associated to that subimage will be the human assessment associated to the whole image. From now on, we will note as $\mathbf{M}_w^i = [m_1^{i,w}, \ldots, m_{K'}^{i,w}]$ the vector of measures for the w-th window of the image I_i, with $m_k^{i,w}$ being the result of applying the measure $P_k \in \mathcal{P}$ to the w-th window of the image I_i.

2.1.2 *The poll*

Given the image set $\mathcal{I}$, the next step is to obtain assessments about the perception of fineness from a set of subjects. From now on we shall note as $\Theta^i = [o_1^i, \ldots, o_L^i]$ the vector of assessments obtained from L subjects for the image I_i. To get Θ^i, subjects will be asked to assign images to classes, so that each class has associated a perception degree of fineness. In particular, 20 subjects have participated in the poll and 9 classes have been considered. The first nine images in figure 1 show the nine representative images for each class used in this poll. It should be noticed that the images are decreasingly ordered according to the degree of fineness.

As result, a vector of 20 assessments $\Theta^i = [o_1^i, \ldots, o_{20}^i]$ is obtained for each image $I_i \in \mathcal{I}$. The degree o_j^i associated to the assessment given by the subject S_j to the image I_i is computed as $o_j^i = (9 - k) * 0.125$, where $k \in \{1, \ldots, 9\}$ is the index of the class C_k to which the image is assigned by the subject.

ISBN: 978-989-95079-6-8

Algorithm 1 Distinguishable clusters selection

Input:

$Part^0 = C_1, C_2, \ldots, C_n$: Initial Partition

δ: distance function between clusters

ϕ: Set of multiple comparison tests

NT: Number of positive tests to accept distinguishability

1.-Initialization

$k = 0$

$distinguishable = false$

2.- While (distinguishable = false) and (k < n)

Apply the multiple comparison tests ϕ to $Part^k$

If for each pair $C_i, C_j \in Part^k$ more than NT of the multiple comparison tests ϕ show distinguishability

$distinguishable = true$

Else

Search for the pair of clusters C_r, C_{r+1}, verifying $\delta(C_r, C_{r+1}) = \min\{\delta(C_i, C_{i+1}), C_i, C_{i+1} \in Part^k\}$

Join C_r and C_{r+1} on a cluster $C_u = C_r \cup C_{r+1}$

$Part^{k+1} = Part^k - C_r - C_{r+1} + C_u$

$k = k + 1$

3.- Output: $\widetilde{Part}_k = C_1, C_2, \ldots, C_{n-k}$

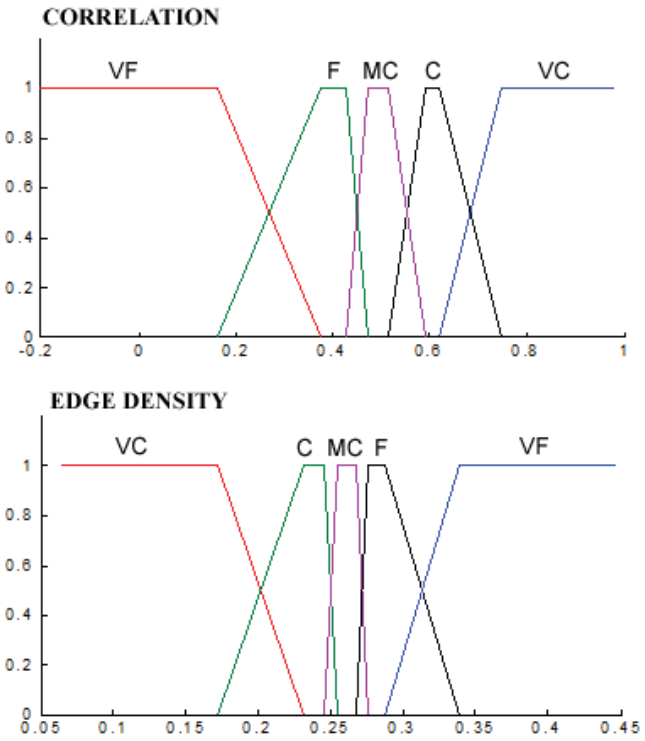


Figure 2: Fuzzy partitions for the measures Correlation and Edge Density. The linguistic labels are VC = very coarse, C = coarse, MC = medium coarse, F = fine, VF = very fine

2.1.3 *Assessment aggregation*

Our aim at this point is to obtain, for each image in the set $\mathcal{I}$, one assessment v^i that summarizes the assessments Θ^i given by the different subjects about the presence degree of fineness.

To aggregate opinions we have used an OWA operator guided by a quantifier [9]. Concretely, the quantifier "the most" has been employed, which allows to represent the opinion of the majority of the polled subjects. This quantifier is defined as

$$Q(r) = \begin{cases} 0 & \text{if } r < a, \\ \frac{r-a}{b-a} & \text{if } a \leq r \leq b, \\ 1 & \text{if } r > b \end{cases} \quad (1)$$

with $r \in [0, 1]$, $a = 0.3$ and $b = 0.8$. Once the quantifier Q has been chosen, the weighting vector of the OWA operator can be obtained following Yager [9] as $w_j = Q(j/L) - Q((j-1)/L), j = 1, 2, ..., L$. According to this, for each image $I_i \in \mathcal{I}$, the vector Θ^i obtained from L subjects will be aggregated into one assessment v^i as follows:

$$v^i = w_1\hat{o}^i_1 + w_2\hat{o}^i_2 + ... + w_L\hat{o}^i_L \quad (2)$$

where $[\hat{o}^i_1, \ldots, \hat{o}^i_L]$ is a vector obtained by ranking in nonincreasing order the values of the vector Θ^i.

2.2 *Distinguishability Analysis of the Fineness Measures*

As it was expected, some measures have better ability to represent fineness-coarseness than the others. To study the ability of each measure to discriminate different degrees of fineness-coarseness (i.e. how many classes can P_k actually discriminate), we propose to analyze each $P_k \in \mathcal{P}$ by applying a set of multiple comparison tests following the algorithm 1. This algorithm starts with an initial partition[1] and iteratively joins clusters until a partition in which all classes are distinguishable is achieved. In our proposal, the initial partition will be formed by the 9 classes used in our poll (where each class will contain the images assigned to it by the majority of the subjects), as δ the Euclidean distance between the centroids of the involved classes will be used, as ϕ a set of 5 multiple comparison tests will be considered (concretely, the tests of Scheffé, Bonferroni, Duncan, Tukey's least significant difference, and Tukey's honestly significant difference [10]), and finally the number of positive tests to accept distinguishability will be fixed to $NT = 3$.

From now on, we shall note as $\Upsilon_k = C^k_1, C^k_2, \ldots, C^k_{N_k}$ the N_k classes that can be discriminated by P_k. For each C^k_r, we will note as $\bar{c}^k_r$ the class representative value. In this paper, we propose to compute $\bar{c}^k_r$ as the mean of the measure values in the class C^k_r.

Table 1 shows the parameters obtained by applying the proposed algorithm 1 with the different measures considered in this paper. The second column of this table shows the N_k classes that can discriminated each measure and the third column shows how the initial classes have been grouped. The columns from fourth to eighth show the representative values $\bar{c}^k_r$ associated to each cluster.

2.3 *The Fuzzy Partitions*

In this section we will deal with the problem of defining the membership function for each fuzzy sets compounding the partition. As it was explained, the number of fuzzy sets will be given by the number of categories that each measure can discriminate.

In this paper, trapezoidal functions are used for defining the

[1] Let us remark that this partition is not the "fuzzy partition". In this case, the elements are measure values and the initial clusters the ones given by the pool

ISBN: 978-989-95079-6-8

Table 1: Result obtained by applying the algorithm 1

Measure	N_k	Classes	$\bar{c}_5 \pm KW_5/2$	$\bar{c}_4 \pm KW_4/2$	$\bar{c}_3 \pm KW_3/2$	$\bar{c}_2 \pm KW_2/2$	$\bar{c}_1 \pm KW_1/2$
Correlation [3]	5	{1,2-4,5-6,7-8,9}	0.122±0.038	0.403±0.0272	0.495±0.0225	0.607±0.0133	0.769±0.0210
ED [11]	5	{1,2,3-5,6-8,9}	0.348±0.0086	0.282±0.0064	0.261±0.0063	0.238±0.0066	0.165±0.0061
Abbadeni [12]	4	{1,2-6,7-8,9}	-	5.672±0.2738	9.208±0.4247	11.12±0.2916	25.23±1.961
Amadasun [1]	4	{1,2-6,7-8,9}	-	4.864±0.271	7.645±0.413	9.815±0.230	19.62±1.446
Contrast [3]	4	{1,2-5,6-8,9}	-	3312±265.5	2529±295.5	1863±94.84	790.8±129.4
FD [13]	4	{1,2,3-8,9}	-	3.383±0.0355	3.174±0.0282	2.991±0.0529	2.559±0.0408
Tamura [2]	4	{1,2-6,7-8,9}	-	1.540±0.0634	1.864±0.0722	2.125±0.0420	3.045±0766
Weszka [14]	4	{1,2-6,7-8,9}	-	0.153±0.0064	0.113±0.0093	0.099±0.0036	0.051±0.0041
DGD [15]	3	{1,2-8,9}	-	-	0.020±0.0010	0.038±0.0017	0.091±0.0070
FMPS [16]	3	{1,2-8,9}	-	-	0.256±0.0477	0.138±0.0122	0.0734±0.0217
LH [3]	3	{1,2-8,9}	-	-	0.023±0.0010	0.052±0.0025	0.127±0.0096
Newsam [17]	3	{1,2-6,7-9}	-	-	0.1517±0.0425	0.2654±0.0466	0.4173±0.0497
SNE [18]	3	{1,2-8,9}	-	-	0.879±0.0182	0.775±0.0087	0.570±0.0232
SRE [19]	3	{1,2-8,9}	-	-	0.995±0.00026	0.987±0.00066	0.966±0.0030
Entropy [3]	2	{1,2-9}	-	-	-	9.360±0.124	8.656±0.301
Uniformity[3]	2	{1,2-9}	-	-	-	$1.3_{E^{-4}}$ ±$2.6_{E^{-5}}$	$3.9_{E^{-4}}$ ±$1.9_{E^{-4}}$
Variance[3]	1	-	-	-	-	-	-

membership functions. In addition, a fuzzy partition in the sense of Ruspini is proposed. Figure 2 shows some examples of the type of fuzzy partition used. To establish the localization of each kernel, the representative value $\bar{c}_r^k$ will be used (in our case, the mean). Concretely, this value will be localized at the center position of the kernel.

To establish the size of the kernel, we propose a solution based on the multiple comparison tests used in section 2.2. As it is known, in these tests confidence intervals around the representative value of each class are calculated (being accomplish that these intervals do not overlap for distinguishable classes). All values in the interval are considered plausible values for the estimated mean. Based on this idea, we propose to set the kernel size as the size of the confidence interval.

The confidence interval CI_r^k for the class C_r^k is defined as

$$CI_r^k = \bar{c}_r^k \pm 1.96 \frac{\bar{\sigma}_r^k}{\sqrt{||C_r^k||}} \quad (3)$$

where $\bar{c}_r^k$ is the class representative value, and $\bar{\sigma}_r^k$ is the estimated standard deviation for the class. Thus, the kernel size KW_r^k is

$$KW_r^k = 3.92 \frac{\bar{\sigma}_r^k}{\sqrt{||C_r^k||}} \quad (4)$$

and the endpoints of the kernel will be given by $\bar{c}_r^k \pm KW_r^k/2$ Table 1 shows these values for each measure and each class.

Figure 2 shows the fuzzy partitions for the measures of correlation and ED (the ones with higher capacity to discriminate fineness clases).

3 Results

In this section, the fuzzy partition defined for the measure "Correlation" (showed in Figure 2) will be applied in order to analyze the performance of the proposed model.

Let's consider Figure 3(A) corresponding to a mosaic made by several images, each one with a different increasing degree of fineness. Figure 3(B-F) shows the membership degree to the fuzzy sets "very coarse", "coarse", "medium coarse", "fine" and "very fine", respectively, using the proposed model. For each pixel in the original image, a centered window of size 32×32 has been analyzed and its membership degree to each fuzzy set has been calculated. Thus, Figure 3(B) represents the degree in which the texture is perceived as "very coarse", with a white level meaning maximum degree, and a dark one meaning zero degree. It can be noticed that our model captures the evolution of the perception degrees of fineness.

Figure 4 presents an example where the proposed fuzzy partition has been employed for pattern recognition. In this case, Figure shows a microscopy image (Figure 4(A)) corresponding to the microstructure of a metal sample. The lamellae indicates islands of eutectic, which are to be separated from the uniform light regions. The brightness values in regions of the original image are not distinct, so texture information is needed for extracting the uniform areas. This fact is showed in Figure 4(B1,B2), where a thersholding on the original image is displayed (homogeneous regions cannot be separated from the textured ones as they "share" brightness values). Figure 4(C1) shows a mapping from the original image to its membership degree to the fuzzy set associated "very coarse". Thus, Figure 4(C1) represents the degree in which the texture is perceived as "very coarse" and it can be noticed that uniform regions correspond to areas with the maximum degree (bright grey levels), so if only the pixels with degree upper than 0.9 are selected, the uniform light regions emerge with ease (Figure 4(C2,C3)).

4 Conclusions and future works

In this paper, a fuzzy partition for representing the fineness/coarseness concept have been proposed. The number of fuzzy sets and the parameters of the membership functions have been defined relating fineness measures with the human perception of this texture property. Pools have been used for collecting data about the human perception of fineness, and the capability of each measure to discriminate different coarseness degrees has been analyzed. The results given by our approach show a high level of connection with the human perception of fineness/coarseness. As future work, the perfor-

ISBN: 978-989-95079-6-8

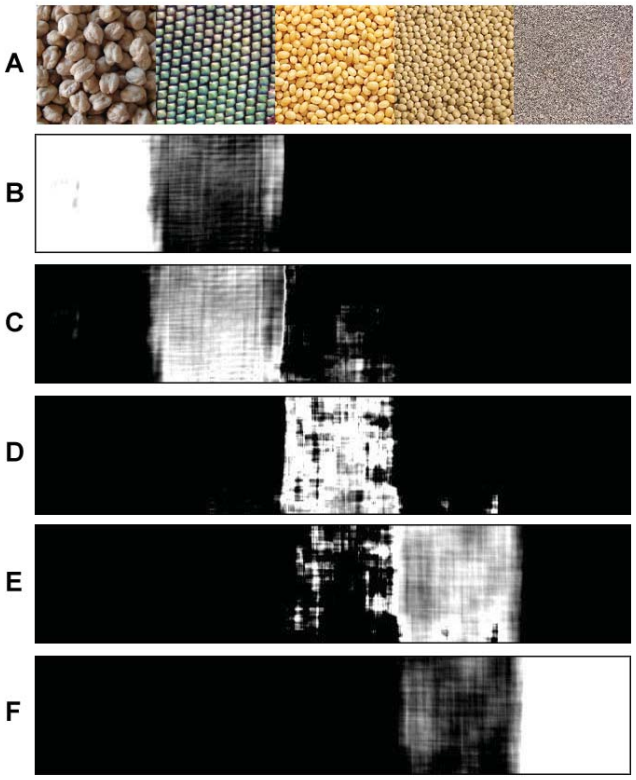


Figure 3: Results for a mosaic image. (A) Original image (B)(C)(D)(E)(F) Membership degree of each pixel to the sets ”very coarse”, ”coarse”, ”medium coarse”, ”fine” and ”very fine”, respectively (the darker the pixel, the lower the membership degree)

mance of the fuzzy partition will be analyzed in applications like textural classification or segmentation.

References

[1] M. Amadasun and R. King. Textural features corresponding to textural properties. *IEEE Transactions on Systems, Man and Cybernetics*, 19(5):1264–1274, 1989.

[2] H. Tamura, S. Mori, and T. Yamawaki. Textural features corresponding to visual perception. *IEEE Transactions on Systems, Man and Cybernetics*, 8:460–473, 1978.

[3] R.M. Haralick. Statistical and structural approaches to texture. *Proceedings IEEE*, 67(5):786–804, 1979.

[4] M. Hanmandlu, V. K. Madasu, and S. Vasikarla. A fuzzy approach to texture segmentation. In *Proc. International Conference on Information Technology: Coding and Computing*, volume 1, pages 636–642, 2004.

[5] A. Barcelo, E. Montseny, and P. Sobrevilla. Fuzzy texture unit and fuzzy texture spectrum for texture characterization. *Fuzzy Sets and Systems*, 158:239–252, 2007.

[6] J. Chamorro-Martnez, E. Galn-Perales, J.M. Soto-Hidalgo, and B. Prados-Suarez. Using fuzzy sets for coarseness representation in texture images. In *Proceedings IFSA'07*, pages 783–792, 2007.

[7] S. Kulkarni and B. Verma. Fuzzy logic based texture queries for cbir. In *Proc. 5th Int. Conference on Computational Intelligence and Multimedia Applications*, pages 223–228, 2003.

[8] H.C. Lin, C.Y. Chiu, and S.N. Yang. Finding textures by textual descriptions, visual examples, and relevance feedbacks. *Pattern Recognition Letters*, 24(14):2255–2267, 2003.

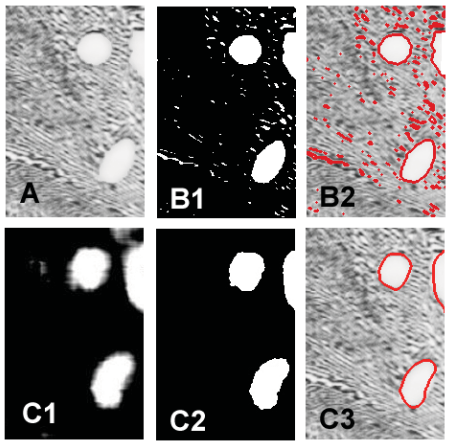


Figure 4: Example of pattern recognition (A) Original image (B1) Binary image obtained by thresholding the original one (B2) Region outlines of B1 superimposed on original image (C1) Membership degrees to the set ”very coarse” obtained with our model from the original image (C2) Binary image obtained by thresholding C1 (C3) Region outlines of C2 superimposed on original image

[9] R.R. Yager. On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Transactions on Systems, Man and Cybernetics*, 18(1):183–190, 1988.

[10] Y. Hochberg and A.C. Tamhane. *Multiple Comparison Procedures*. Wiley, 1987.

[11] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[12] N. Abbadeni, N. Ziou, and D.S. Wang. Autocovariance-based perceptual textural features corresponding to human visual perception. In *Proc. of 15th International Conference on Pattern Recognition*, volume 3, pages 901–904, 2000.

[13] S. Peleg, J. Naor, R. Hartley, and D. Avnir. Multiple resolution texture analysis and classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (4):518–523, 1984.

[14] J.S. Weszka, C.R. Dyer, and A. Rosenfeld. A comparative study of texture measures for terrain classification. *IEEE Transactions on Systems, Man and Cybernetics*, 6:269–285, 1976.

[15] S.I. Kim, K.C. Choi, and D.S. Lee. Texture classification using run difference matrix. In *Proc. of IEEE 1991 Ultrasonics Symposium*, volume 2, pages 1097–1100, December 1991.

[16] H. Yoshida, D.D. Casalino, B. Keserci, A. Coskun, O. Ozturk, and A. Savranlar. Wavelet-packet-based texture analysis for differentiation between benign and malignant liver tumours in ultrasound images. *Physics in Medicine and Biology*, 48:3735–3753, 2003.

[17] S.D. Newsam and C. Kammath. Retrieval using texture features in high resolution multi-spectral satellite imagery. In *Data Mining and Knowledge Discovery: Theory, Tools, and Technology VI, SPIE Defense and Security*, April 2004.

[18] C. Sun and W.G. Wee. Neighboring gray level dependence matrix for texture classification. *Computer Vision, Graphics and Image Processing*, 23:341–352, 1983.

[19] M.M. Galloway. Texture analysis using gray level run lengths. *Computer Graphics and Image Processing*, 4:172–179, 1975.

ISBN: 978-989-95079-6-8

A Fuzzy Particle Swarm Optimization Algorithm for a Cell Formation Problem

Esmaeil Mehdizadeh[1], Reza Tavakkoli-Moghaddam [2]

1Faculty of Industrial and Mechanical Engineering, Islamic Azad University, Qazvin Branch, Iran
2 Department of Industrial Engineering, College of Engineering, University of Tehran, Tehran, Iran
Email: mehdizadeh@qazviniau.ac.ir

Abstract— *Group technology (GT) is a useful way to increase productivity with high quality in cellular manufacturing systems (CMSs), in which cell formation (CF) is a key step in the GT philosophy. When boundaries between groups are fuzzy, fuzzy clustering has been successfully adapted to solve the CF problem; however, it may result uneven distribution of parts/machines where the problem becomes larger. In this case, particle swarm optimization (PSO) can be used to tackle such a hard problem. This paper proposes a hybrid algorithm based on the fuzzy clustering and particle swarm optimization (FPSO) to solve the given CF problem. We experiment a number of examples to show the efficiency of the proposed algorithm and find that our proposed FPSO algorithm is able to obtain good results at reasonable time.*

Keywords— Cellular manufacturing systems, cell formation, fuzzy clustering, particle swarm optimization (PSO)

1 Introduction

Cell formation (CF) is a key step in Group technology (GT) that is used to design a good cellular manufacturing system by using the similarities of parts related to machines so that it can identify part families and machine groups. The parts in the same machine group have similar requirement so that GT can reduce travel and setup time. In CF the part/machine matrix, which has $m \times p$ dimensions with binary components, is usually described and given. The m rows indicate machines and p columns represent p parts that need to be operated upon. In matrix m, ‘‘1’’ (‘‘0’’) represents that this part should be (not) worked on the machine. The matrix exhibits parts requirement relative to machines. Our objective is to group parts and machines just like a cell.

Cell formation problem (CFP) is shown to be NP-hard [1] in the strong sense, and obtaining an optimal solution for the large-sized problems in reasonable computational time is extremely difficult.

Clustering involves the task of dividing data points into homogeneous classes or clusters so that items in the same class are as similar as possible and items in different classes are as dissimilar as possible. In real applications there is very often no sharp boundary between clusters so that fuzzy clustering is often better studied for the data. Membership degrees between zero and one are used in fuzzy clustering instead of crisp assignments of the data to clusters. In fuzzy clustering, the data points can belong to more than one cluster, and associated with each of the points are membership grades which indicate the degree to which the data points belong to the different clusters. In deterministic CF methods assumed well- defined boundaries between part–machine cells. These crisp boundary assumptions may fail to fully describe the case where the part–machine cell boundaries are fuzzy. This is why fuzzy clustering algorithms were applied for CF.

There are many CF methods in the literature [2]. The CF models can also be categorized into those of crisp or fuzzy. Crisp models assume that there are well-defined boundaries between groups and therefore assign each part or machine to only one family. In reality, some parts may belong to one part family, but there may have parts whose linkages are much less evident.

Various Clustering methods have been proposed to solve the CF problem. The fuzzy *c*-means (FCM) algorithm was first used in part-family formation by Xu and Wang [3]. FCM algorithm performs well with small and well-structured data sets. However, when the data set becomes larger, the algorithm may result in clustering errors, infeasible solutions, and uneven distribution of parts/machines. Then, several researchers have proposed alternatives to remedy these weaknesses with mixed success. For example, Chu and Hayya [4] improved the study carried out by Xu and Wang [3].

Al- Ahmari [5], Yang et al. [6] and Feng et al. [7] applied concepts of fuzzy clustering on the cell formation problem. Li et al. [8] improved fuzzy clustering algorithm to overcome the deficiencies of FCM. Since large instances are so difficult to optimally solve, approximate methods are needed. Perhaps, meta-heuristics are the most successful approximate methods that have been used so far. Thus, for example, Boctor [9] and Chen and Srivastava [10] used simulated annealing. Genetic algorithms have been used by Kazerooni et al. [11], Brown and Sumichrast [12], Ravichandran et al. [13]. Aljaber et al. [14] and Lozano et al. [15] applied tabu search. Attila [16] proposed an ant system algorithm. Zhao et al. [17] used swarm intelligence, and finally Andres and Lozano [18] presented particle swarm optimization (PSO) algorithm to solve the cell formation problem addressed in GT.

In the previous work, fuzzy clustering and PSO have been applied in the CF in separate. Our aim is to design a fuzzy particle swarm optimization (FPSO) clustering algorithm to solve the part–machine grouping problem,

ISBN: 978-989-95079-6-8

which is known as a hard combinatorial problem.

The rest of this paper is given bellow. The fuzzy cell formation model is presented in Section 2. In Section 3, the PSO algorithm is presented. The proposed FPSO algorithm is presented in Section 4. Computational results with a number of test problems taken from the literature are shown in Section 5. Finally, Section 6 draws conclusions, suggests directions for future research and discusses the limitations of the research.

2 Fuzzy clustering problem

The fuzzy cell formation (FCF) problem described by Li et al. [8] as follows: Given the routing information of n parts and m machines, the goal of CF is to cluster the parts into c part families and the corresponding machines into machine cells. The classification result can be expressed as a matrix $U=[\mu_{ik}]_{c\times n}$, (k= 1, 2,…, n and i= 1, 2,…, c), and μ_{ik} is the membership degree of part k to group i, which satisfies:

$$0 \le \mu_{ik} \le 1 \tag{1}$$

$$\sum_{i=1}^{c} \mu_{ik} = 1 \tag{2}$$

$$0 < \sum_{k=1}^{n} \mu_{ik} \le n \tag{3}$$

The FCM clustering algorithm is based on the minimization of the following equation:

$$J(p) = \sum_{k=1}^{n} \sum_{i=1}^{c} [\mu_{ik}]^m \|x_k - V_i\|^2 \tag{4}$$

where, m>1 is a real number governing the influence of membership grades, V_i is the cluster center of the part family i, and x_k is the vector of part k. The necessary conditions for minimizing $J(p)$ are the following update equations:

$$V_i = \frac{\sum_{k=1}^{n} [\mu_{ij}]^m x_k}{\sum_{k=1}^{n} [\mu_{ij}]^m} \tag{5}$$

which, c= 1, 2, … , n.

$$\mu_{ik}^{(t+1)} = \left[\sum_{j=1}^{c} \left(\frac{\|x_k - V_i^{(t)}\|^2}{\|x_k - V_j^{(t)}\|^2} \right)^{\frac{1}{m-1}} \right]^{-1} \tag{6}$$

where, $\|x_k - V_i\|^2$ represents the Euclidean distance between x_k and V_i , and $\mu_{ik}^{(t+1)}$ is the membership degree of part k in group i.

Many variations of FCM algorithms can be found in Bezdek [19]. The algorithm is based on the assumption that the desired number of clusters c, real number m, stopping criterion ε, and a particular distance are given.

Step 1: Let t=0 and select an initial fuzzy pseudo-partition $p^{(0)}$.

Step 2: Calculate c cluster centers, $V_1^{(t)}$, …, $V_c^{(t)}$, by (5) for $p^{(t)}$ and the chosen value of m.

Step 3: Define $\mu_i^{(t+1)}$ by (6) and update $p^{(t+1)}$

Step 4: Compare $p^{(t)}$ and $p^{(t=1)}$. If $\left| p^{(t+1)} - p^{(t)} \right| \le \varepsilon$, then stop the algorithm; otherwise, increase t by one and then return to Step 2.

Since the fuzzy clustering problem is a combinatorial optimization problem that is hard to solve [20]. Large instances are so difficult to optimally solve, approximate methods are then needed.

3 Particle swarm optimization

In particle swarm optimization (PSO) a number of simple entities—the particles—are placed in the search space of some problem or function, and each evaluates the objective function at its current location. Each particle then determines its movement through the search space by combining some aspect of the history of its own current and best (best-fitness) locations with those of one or more members of the swarm, with some random perturbations. The next iteration takes place after all particles have been moved. Eventually the swarm as a whole likes a flock of birds collectively foraging for food is likely to move close to an optimum of the fitness function.

There are many variants of the PSO proposed in the literature so far, when Eberhart and Kennedy [21] first introduced this technique. A version of this algorithm is used for part-machine grouping by Andres and Lozano [18].

For description of The PSO algorithm, first, let me define the notation adopted in this paper: the i-th particle of the swarm is represented by the D-dimensional vector $x_i=(x_{i1}, x_{i2}, \ldots, x_{iD})$ and the best particle of the swarm (i.e., the particle with the smallest function value) is denoted by index p_g. The best previous position (i.e., the position giving the best function value) of the i-th particle is recorded and represented $p_i=(\rho_{i1}, \rho_{i2}, \ldots, \rho_{iD})$, and the position change (velocity) of the i-th particle is $Vel_i=(Vel_{i1}, Vel_{i2}, \ldots, Vel_{iD})$. The particles are then manipulated according to the following equations:

$$Vel_{id}(t+1) = \chi\{wVel_{id}(t) + c_1\phi_1[\rho_{id}(t) - x_{id}(t)] + c_2\phi_2[(\rho_{gd}(t) - x_{id}(t)]\} \tag{7}$$

$$x_{id}(t+1) = x_{id}(t) + Vel_{id}(t+1) \tag{8}$$

where, d=1, 2, …, D ; i=1, 2, …, n; and n is the size of the swarm; w is the inertia weight; c_1 and c_2 are two positive acceleration constants; ϕ_1 and ϕ_2 are two random values into the range [0, 1]; χ is a constriction factor that is used in constrained optimization problems in order to control the magnitude of the velocity (in unconstrained optimization problems it is usually set to 1.0).

4 Proposed FPSO algorithm

In the fuzzy clustering, a single particle represents a cluster center vector and a swarm represents a number of candidates clustering for the current data vector. Here; each point or data vector belongs to every various cluster by different membership function, thus; assign a fuzzy membership to each point or data vector. Each cluster has a cluster center, and per iteration presents a solution that gives

ISBN: 978-989-95079-6-8

a vector of cluster centers. We determine the position of each vector for every particle and update it, then change the position of cluster centers based of particles. For the purpose of our algorithm, we define the following notations:

n	Number of part
c	Number of cluster center
$V_l^{(t)}$	Position of l-th particle' cluster center in stage t
$Vel_l^{(t)}$	Velocity of l-th particle in stage t
x_k	Vector of parts ($k = 1,2,.....,n$)
$\rho_l^{(t)}$	Best position funded by l-th particle in stage t
$\rho_g^{(t)}$	Best position funded by all particles in stage t
$P^{(t)}$	Fuzzy pseudo partition in stage t
$\mu_{ik}^{(t)}$	Membership function of k-th part in stage t into i-th cluster

The fitness of particles is easily measured by (4). The c-means algorithm tends to converge faster than the proposed FPSO algorithm with a less accurate clustering. In this section, the performance of the PSO clustering algorithm is improved by seeding the initial swarm with the result of the c-means algorithm. The FPSO algorithm first executes the c-means algorithm once. In this case, the c-means clustering algorithm is terminated by one of two stopping criteria: I) the maximum number of iterations; or II) $\left|p^{(t+1)} - p^{(t)}\right| \le \varepsilon$. The result of c-means algorithm is then used as one of the particles, while the rest of the swarms are initialized randomly. The following algorithm can use to finding cluster for each data vector or part:

Step 1: Let t=0, select initial parameters such as number of cluster center c, initial position of particle by the FCM, initial velocity of particles, c_1, c_2, w, χ, and a real number $m \in (1, \infty)$, and a small positive number ε for stopping criterion.
Step 2: Calculate $\mu_{ik}^{(t)}$ for all particles and all i=1, 2, …, c and k=1, 2, …, n by (6) and update $p^{(t+1)}$.
Step 3: For each particle, calculate the fitness by using (4).
Step 4: Update the global best and local best position.
Step 5: Update $Vel_l^{(t)}$ and $V_l^{(t)}$ for all l=1, 2, …, *n_particle* by using (7) and (8).
Step 6: Update $p^{(t+1)}$ by the Step 2.
Step 7: Compare $p^{(t)}$ and $p^{(t+1)}$. If $\left|p^{(t+1)} - p^{(t)}\right| \le \varepsilon$, then stop; otherwise, increase t by one and continue form Step 3.

5 Numerical example

In this section, examples from the literature are considered to illustrate the application of the proposed fuzzy algorithm in the cell formation problem. We compare the results of the FCM and FPSO algorithms on various problems taken from the literature. Their performances are measured by the objective function value given in (4) and CPU time. A general rule of thumb is that a clustering result with lower *J(p)* and lower CPU time is preferable. For a comparable assessment, we code these methods by using the fuzzy tools available in MATLAB 7 and the FPSO, respectively, with 10 particles, w=0.72, and $c_1 = c_2 = 1.49$. For our experimental tests, we use a PC Pentium III, CPU 1133 MHz and 256 MB of RAM for the same parameters for all algorithms implementations: m=2, the maximum number of iterations is 100 and $\varepsilon = 0.00001$.

Table 1: Data from Chu and Hayya [4]

	Parts								
Machines	1	2	3	4	5	6	7	8	9
1	1	1	0	0	1	0	0	0	0
2	1	1	0	0	0	1	0	0	1
3	0	0	1	0	0	0	1	1	0
4	0	1	1	1	0	0	0	1	0
5	1	0	0	1	1	0	0	1	0
6	0	1	0	0	0	1	0	0	1
7	0	0	1	0	0	0	1	1	0
8	0	0	1	1	1	0	1	1	0
9	0	1	0	0	0	1	0	0	1

Table 2: The membership matrix for cells for first example

Machines	Memberships for cells		
	1	2	3
1	.2239	.2755	.5006
2	.0912	.7838	.1250
3	.9335	.0249	.0417
4	.4413	.1871	.3716
5	.0712	.0392	.8895
6	.0160	.9666	.0174
7	.9335	.0249	.0417
8	.5359	.0947	.3694
9	.0160	.9666	.0174

Table 3: The membership matrix for part families for first example

Parts	Memberships for part families		
	1	2	3
1	.7110	.1656	.1234
2	.2502	.6234	.1264
3	.0698	.0441	.8861
4	.5026	.1370	.3603
5	. 7659	.0922	.1419
6	.0437	.9346	.0216
7	.2061	.1233	.6706
8	.0751	.0330	.8919
9	.0437	.9346	.0216

Table 4: Comparison of Chu and Hayya's approach and FPSO approach

	Chu and Hayya's approach		The FPSO approach	
	Machine cell	Part families	Machine cell	Part families
Cell 1	M1,M5	P1,P4,P5	M1,M5	P1,P4,P5
Cell 2	M2,M6,M9	P2,P6,P9	M2,M6,M9	P2,P6,P9
Cell 3	M3,M4 M7,M8	P3,P7,P8	M3,M4 M7,M8	P3,P7,P8
$J(p)$	3.729600	3.859139	3.729600	3.859139

ISBN: 978-989-95079-6-8

The first example taken from Chu and Hayya [4] consists of nine machines and nine parts is illustrated in Table 1. Tables 2 and 3 show the membership matrix values for cells and part families, respectively. The final membership matrix values indicates the degree of membership of each machine associated with the machine cell (MC) and can be configured as: MC (1) = {M1, M5}, MC (2) = {M2, M6, M9}, and MC (3) = {M3, M4, M7, M8}. Similarity, the part family (PF) can be configured as: PF (1) = {P1, P4, P5}, PF (2) = {P2, P6, P9}, and PF (3) = {P3, P7, P8}. Table 4 compares Chu and Hayya's approach with the FPSO approach results. As illustrated in the table, no changes are observed between machine cells and part families.

Table 5: Data from Susanto et al. [22]

Machine	Parts									
	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	0	0	0	0	0	0
2	1	1	1	1	0	0	0	0	0	0
3	1	1	1	0	0	0	0	0	0	0
4	1	1	0	0	0	0	0	0	0	0
5	0	0	0	0	1	1	1	0	0	0
6	0	0	0	0	1	1	1	0	0	0
7	0	0	0	1	1	1	0	0	0	0
8	0	0	0	0	0	0	0	1	1	1
9	0	0	0	0	0	0	0	1	1	1
10	0	0	0	0	0	0	0	1	1	0
11	1	1	0	0	1	1	0	0	0	0
12	1	1	0	0	1	1	0	0	0	0

The second example is taken from Susanto et al. [22] as shown in Table 5. The obtained results are illustrated in Tables 6 and 7 and can be configured as: MC (1) = {M5, M6, M7, M11, M12}, MC (2) = {M8, M9, M10}, and MC (3) = {M1, M2, M3, M4}, $J_{(p)} = 4.844414$. Part families are as follows: PF (1) = {P5, P6, P7}, PF (2) = {P8, P9, P10}, and PF (3) = {P1, P2, P3, P4}, J (p) = 5.020256.

Table 6: The membership matrix for cells for second example

Machines	Degree of membership		
	1	2	3
1	.0788	.0701	.8511
2	.0788	.0701	.8511
3	.0842	.0730	.8428
4	.1233	.1082	.7685
5	.9017	.0460	.0522
6	.9017	.0460	.0522
7	.6359	.1578	.2062
8	.0541	.8940	.0519
9	.0541	.8940	.0519
10	.01159	.7661	.1180
11	.4311	.1561	.4128
12	.4311	.1561	.4128

Table 7: The membership matrix for part families for second example

Parts	Degree of membership		
	1	2	3
1	0.9265	0.0315	0.0420
2	0.9265	0.0315	0.0420
3	0.6430	0.1975	0.1595
4	0.4342	0.2743	0.2916
5	0.0238	0.0236	0.9525
6	0.0238	0.0236	0.9525
7	0.1765	0.2902	0.5333
8	0.0218	0.9534	0.0248
9	0.0218	0.9534	0.0248
10	0.0702	0.8483	0.0815

Table 8: Solution to Al-Ahmari

Cell	Machine Cells	$J(p)$ FCM	CPU Time FCM	$J(p)$ FPSO	CPU Time FPSO
2	{1,3,4,9,10,13,14,16,17,20,21,22},{2,5,6,7,8,11,12,15,18,19,23,24}	54.583351	0.300	54.583351	0.190
3	{1,3,13,16,20,21,22},{6,7,9,10,14,17,23,24},{2,4,5,8,11,12,15,18,19}	36.388919	0.641	36.388919	0.170
4	{2,3,4,6,9,10,11,12,15,16,17,18},{1,8,14,20,23,24},{7,13,21,22},{5,19}	22.515392	0.490	22.515392	0.431
5	{1,13,21,22},{2,5,11,19},{3,4,7,14,16,20,23,24},{9,10,17},{6,8,12,15,18}	14.691166	0.852	14.691166	0.280
6	{3,20},{1,13,21,22},{2,5,11,19},{7,14,23,24},{4,6,8,12,15,16,18},{9,10,17}	9.046538	0.571	9.046538	0.350
7	{4,16},{9,10,17},{6,8,12,15,18},{3,20},{2,5,11,19},{1,13,21,22},{7,14,23,24}	6.172838	0.611	6.172838	0.581
8	{1,13,21,22},{7,23,24},{4,16},{14},{9,10,17},{6,8,12,15,18},{3,20},{2,5,11,19}	4.139911	0.651	4.139911	0.591
9	{9,10},{14},{17},{6,8,12,15,18},{3,20},{2,5,11,19},{1,13,21,22},{7,23,24},{4,16}	2.772339	0.671	2.772339	0.411
10	{4,16},{6,7,23,24},{17},{8,12,15,18},{2,5,11,19},{1,14},{13,21,22},{6},{3,20},{9,10}	1.694722	0.551	1.694722	0.541

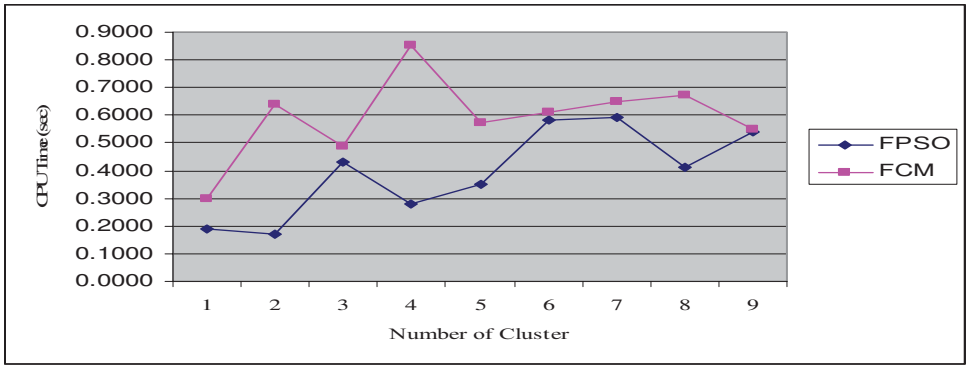


Figure 1: CPU Time comparison between FPSO and FCM algorithms

ISBN: 978-989-95079-6-8

The proposed approach is also tested on a large-sized problem taken from Al-Ahmari [5]. In the example, the problem with 24 machines and 40 parts is considered. Different numbers of manufacturing cells are used (i.e., 2, 3, 4, 5, 6, 7, 8, 9, and 10). The results summarized in table 8 and Figure 1. This example demonstrates the possibility of using the approach for large-scale CF problems, and evaluates the obtained results using $J(P)$.

6 Conclusion

Group technology (GT) is a useful way to increase productivity with high quality in flexible manufacturing systems; and CF is a key step in GT. It is used to design a good cellular manufacturing that uses the similarities of parts related to machines so that it can identify part families and machine groups. The cell formation problem is NP-complete and different -heuristic methods have been used to solve it. Particle swarm optimization is one of them. On the other hand, the crisp models fail to fully reflect the complex nature of part features or routing data, where boundaries between groups are fuzzy. Fuzzy clustering has been successfully adapted to solve the cell formation problem, but when the problem becomes larger, the fuzzy clustering algorithms may result uneven distribution of parts/machines. In the previous works, fuzzy clustering and particle swarm optimization have been applied in CF individually. We have designed a fuzzy particle swarm optimization clustering algorithm (FPSO) to solve the part–machine grouping problem, which is a hard combinatorial problem. The presented numerical examples confirm the effectiveness of the proposed approach. It is found that this algorithm provides a good solution to CF problems at reasonable time and allows the user in formulating the required size of machine cells and part families.

References

[1] A. Ballakur and H.J. Steudel, "A Within Cell Utilization Based Heuristic for Designing Cellular Manufacturing Systems", *Int. J. of Prod. Res.*, 25: 639-55, 1987.

[2] Y. Yin and K. Yasuda, "Similarity Coefficient Methods Applied to the Cell Formation Problem: A Taxonomy and Review", *Int. J. Prod. Economics,* 101: 329-352, 2006.

[3] H. Xu and H.P. Wang, "Part Family Formation for GT Applications Based on Fuzzy Mathematics", *Int. J. of Prod. Research,* 27: 1637-1651, 1989.

[4] C.H. Chu and J.C. Hayya, "A Fuzzy Clustering Approach to Manufacturing Cell Formation", *Int. J. of Prod. Research*, 29: 1475-1487, 1991.

[5] A.M.A. Al-Ahmari, "A Fuzzy Analysis Approach for Part-Machine Grouping in Cellular Manufacturing Systems", *Integrated Manufacturing System,* 13:7: 489-497, 2002.

[6] M.S. Yang, W.L. Hung and F.C. Cheng, "Mixed-Variable Fuzzy Clustering Approach to Part Family and Machine Cell Formation for GT Applications", *Int. J. Prod. Economics,* 103: 185-198, 2006.

[7] C.Y. Chen and F. Ye, "Adaptive Hyper-Fuzzy Partition Particle Swarm Optimization Clustering Algorithm", *Cybernetics and Systems,* 37:463-479, 2006.

[8] J. Li, C.H Chu, Y. Wang and W. Yan, "An Improved Fuzzy Clustering Method for Cellular Manufacturing*", Int. J. of Prod.. Research,* 45: 1049-1062, 2007.

[9] F.F. Boctor, "A Linear Formulation of the Machine-Part Cell Formation Problem", *Int. J. Prod. Res.,* 29: 343-56, 1991.

[10] W. Chen and B. Srivastava, "Simulated Annealing Procedures for Forming Machine Cells in Group Technology", *Eur. J. Oper. Res.*, 75: 100-111, 1994.

[11] M. Kazerooni, L. Luong and K. Abhary," Cell Formation using Genetic Algorithms. *Int. J. of Flexible Autom. Integr Manuf*, 3: 283-299, 1995.

[12] E.C. Brown and R.T. Sumichrast, "A Grouping Genetic Algorithm for the Cell Formation Problem", *Int. J. Prod. Res.,* 39: 3651-3669, 2001.

[13] K.S. Ravichandran, K. Chandra, S. Rao and R. Saravanan, "The Role of Fuzzy and Genetic Algorithms in Part Family Formation and Sequence Optimization for Flexible Manufacturing Systems", *Int. J. of Adv. Manuf. Technology,* 19: 879-888, 2002.

[14] N. Aljaber, W. Baek and C. L. Chen, "A Tabu Search Approach to the Cell Formation Problem", *Computers and Industrial Eng.*, 32: 169-185, 1997.

[15] S. Lozano, B. Adenso-Dı, I. Eguia, and L. Onieva, "A One-Step Tabu Search Algorithm for Manufacturing Cell Design", *J. of Oper. Res. Soc.,* 50: 509-516, 1999.

[16] A.A. Islier, "Group Technology by an Ant System Algorithm", *Int. J. of Prod. Research,* 43: 913-932, 2005.

[17] G. Zhao, P. Jiang, M. Zheng, G. Zhou, and B. Chen, "Using Swarm Intelligence Algorithm to Solve a Part Clustering Model Based on Weighted Directed Graph", *Int. J. of Prod. Res.,* 2007. A. Andres and S. Lozano, "A Particle Swarm Optimization Algorithm for Part–Machine Grouping", *Robotics and Computer-Integrated Manufacturing,* 22: 468-474, 2006.

[18] J.C. Bezdek, "*Pattern Recognition with Fuzzy Objective Function Algorithms*", Plenum Press: New York, 1981.32

[19] J.G. Klir and B. Yuan, "*Fuzzy Sets and Fuzzy Logic, Theory and Applications"*, Prentice Hall of India, 257-262, 2003.

[20] J. Kennedy and R.C. Eberhart, "Particle Swarm Optimization", In: *Proceedings of IEEE Int. Conf. on Neural Networks,* Perth, Australia, 1942-1948, 1995.

[21] S. Susanto, R.D. Kennedy and J.W.H. Price, "A New Fuzzy-*C*-Means and Assignment Technique-Based Cell Formation Algorithm to Perform Part-Type Clusters and Machine-Type Clusters Separately", *Prod. Planning & Control,* 10: 375-388, 1999.

ISBN: 978-989-95079-6-8

Multiple Negations in Fuzzy Interval Logic

Eunjin Kim[1*], Ladislav J. Kohout[2]

1.Dept. of Computer Science, University of North Dakota
Grand Forks, ND 58202-9015, USA
2.Dept. of Computer Science, Florida Sate University
Tallahassee, FL 32306-4530, USA
Email: ejkim@cs.und.edu, kohout@cs.fsu.edu

Abstract— *This paper continues our study in fuzzy interval logic based on the Checklist Paradigm(CP) semantics of Bandler and Kohout. We investigate the fuzzy interval system of negation which was defined by the Sheffer(*NAND*), the Nicod(*NOR*) and the implication connectives of m_1 interval system in depth. The top-bottom(*TOP-BOT*) pair of fuzzy negation interval shows non-involutive property; however, it shows various interesting properties as the negation is iterated to each* TOP *and* BOT *fuzzy negation connective. As the iteration reaches to infinity, the* TOP-BOT *pair of iterated negation shows nearly involutive property.*

Keywords— Checklist Paradigm, Fuzzy Interval, Fuzzy Negation, Non-Involution, Gap-Theorem.

1 Introduction

The convention that fuzzy logics default into 2-valued logic for the values 0 and 1 is generally accepted. The question how connectives of the same type differ outside the boundary points, within the open interval $(0,1)$ has been well researched. For example the implications such as Lukasiewicz($\overset{L}{\rightarrow}$), Kleene-Dienes($\overset{KD}{\rightarrow}$), Reichenbach($\overset{KDL}{\rightarrow}$), Goguen($\overset{G43}{\rightarrow}$), G$\ddot{o}$del($\overset{S*}{\rightarrow}$), Early-Zadeh($\overset{EZ}{\rightarrow}$) and Wilmott($\overset{W}{\rightarrow}$) are interrelated by the following inequalities [1]:

$$a \overset{S*}{\rightarrow} b \leq a \overset{G43}{\rightarrow} b \leq a \overset{L}{\rightarrow} b, \text{ and}$$

$$a \overset{W}{\rightarrow} b \leq a \overset{EZ}{\rightarrow} b \leq a \overset{KD}{\rightarrow} b \leq a \overset{KDL}{\rightarrow} b \leq a \overset{L}{\rightarrow} b, \text{ where}$$

$a \overset{L}{\rightarrow} b = min(1, 1-a+b)$; $a \overset{KDL}{\rightarrow} b = min(1, 1-a+ab)$; $a \overset{KD}{\rightarrow} b = max(1-a, b)$, etc[1]. However, much less is known about inter-relations of interval systems of connectives.

In 1979 Bandler and Kohout derived five interesting systems of fuzzy logic, $m_1, m_2, \ldots, m_5$, based on the Checklist paradigm [2]. Since then, the systems of connectives that can be generated from implicational intervals by group transformations have been investigated systematically; in particular, the m_1 interval logic system of 16 connectives has been investigated in depth. Among them, ten 2-argument connectives such as $\wedge, \vee, \rightarrow, \ldots \equiv$, *oplus*(XOR) yield the interval pairs of connectives ($conbot \leq m_1 \leq contop$) where its implication($\rightarrow$) yields Lukasiewicz and Kleene-Dienes implication for its TOP-BOT pair of interval, in particular. However, a unary connective, a negation, did not yield interval but just singleton: i.e. $\neg a = 1-a$. From this question, we proposed the alternative negations in [3] by the Sheffer(NAND), the Nicod(NOR) and a pair of <implication, 0>, which fuzzifies $\neg a$, generating the intervals like the fuzzified 2-ary connectives. The TOP-BOT pair of generated negation interval lacks of involutive property on the surface, but a question still remains if they show a nearly involutive property converging to a certain value as an iteration of negation proceeds.
Thus, we further investigate this non-involutive property of TOP-BOT pair of interval negations as well as the relationship between three definitions of negation which are equivalent in two-valued logic but bifurcate into different negations in interval many-valued logic in this paper.

[1]For the rest of definitions, refer to [1],[2]

2 Interval logic system generated by the checklist paradigm

The structure of five fuzzy interval systems m_1 - m_5, based on the Checklist paradigm by Bandler and Kohout in [2] is generated by a distinct measure that performs the *summarization* of the information contained in certain well-defined binary structures called *fine structures*. The interval produced by a measure m_i pair of connectives of one type can be generically characterized by the following inequality:

$$CON_{Bot} \leq m_i \leq CON_{Top}, \quad i \in \{1, 2, \ldots, 5\}$$

For the details of the checklist paradigm and its various uses, see the papers [2],[4],[5],[6],[7],[8].

2.1 Five Implication Operator Based Interval Systems of Bandler and Kohout

The following five inequalities linking the interval bounds for implication operators $[\rightarrow_{bot}, \rightarrow_{top}]$ with corresponding measures, m_i, $i = \{1,2,3,4,5\}$ have been listed in [2].

1. The Kleene-Dienes implication(KD) and Łukasiewicz implication (Ł) respectively, are attainable lower and upper bounds of m_1:
   $max(1-a, b) \leq m_1(\rightarrow) \leq min(1, 1-a+b)$.

2. A certain new function of (a,b) and the Goguen-Gaines (G43) implication (the left-hand side) are respectively attainable lower and upper bounds of m_2:
   $max(0, (a+b-1)/a) \leq m_2(\rightarrow) \leq min(1, b/a)$.

3. Another function of (a,b) and the Early Zadeh implication (EZ) are respectively attainable lower and upper bounds of m_3:
   $max(a+b-1, 1-a) \leq m_3(\rightarrow) \leq max[min(a,b), 1-a]$.

ISBN: 978-989-95079-6-8

4. Still another function of (a,b) and the Wilmott implication (W) respectively, are attainable lower and upper bounds of m_4:
$min[max(a+b-1,1-a),max(b,1-a-b)] \leq m_4(\rightarrow)$
$\leq min[max(1-a,b),(max(a,(1-b),min(b,1-a)))]$.

5. Yet another function of (a,b) and one of G43 respectively, are attainable lower and upper bounds of m_5:
$max[(a+b-1)/a,1-a] \leq m_5(\rightarrow) \leq max[min(1,b/a),1-a]$.

In m_1 system, a Kleene-Dinese logic system forms a a lower bound of the interval($m1(\rightarrow_{bot})$) while a Łukasiewicz logic system forms a upper bound of the interval ($m1(\rightarrow_{top})$).

From this implicational interval pair $[m1(\rightarrow_{bot}), m1(\rightarrow_{top})]$, one can generate all 16 pairs of connectives of corresponding interval systems by Kohout-Bandler group of logic transformations. 10 of these pairs of connectives are genuine interval pairs and remaining 6 connectives collapse into a single point, not an interval. It has been investigated systematically over the years in the series of papers [9], [10]. These 16 connectives of measure $m_1, \ldots, m_5$ can be generated by group transformation. 8 of these interval pair of connectives of m_1 system are shown in section 3.

3 Characterization of logics by group transformation

Definition 1 *Group Transformations in Logic:*
Let f be any one of the 10 two-argument propositional connectives of a logic system and $\neg$ be an involutive negation. Then, we define the following transformations over f:

1. $I(f) = f(x,y)$: *Identity transformation*
2. $D(f) = \neg f(\neg x, \neg y)$: *Dual transformation*
3. $C(f) = f(\neg x, \neg y)$: *Contradual transformation*
4. $N(f) = \neg f(x,y)$: *Negation transformation*

It is well known that for the crisp (2-valued) logic these $\mathcal{T} = \{I,D,C,N\}$ transformations determine the Piaget group which is a realization of the abstract Klein 4-element group.

The new 4 non-symmetrical transformations below which are added to the above 4 basic transformations by Bandler and Kohout in 1979 [11],[12] enriches the algebraic structure of logical transformations.

Definition 2 *8-element Group Transformations: Bandler-Kohout*

5. $LC(f) = f(\neg x, y)$: *Left Contradual*
6. $RC(f) = f(x, \neg y)$: *Right Contradual*
7. $LD(f) = \neg f(\neg x, y)$: *Left Dual*
8. $RD(f) = \neg f(x, \neg y)$: *Right Dual*

This enlarged set of transformations $\mathcal{T} = \{I, D, C, N, LC, RC, LD, RD\}$ forms 8-element group $\{T, \circ\}$ called $\mathcal{S}_{2\times2\times2}$ group. The equation, $N^2 = C^2 = D^2 = LC^2 = LD^2 = RC^2 = RD^2 = I$, provides sufficient information to identify this group. Its group operations are shown in [13] in detail.

When $\mathcal{T} = \{I, D, C, N, LC, RC, LD, RD\}$ are applied to Łukasiewicz implication($\rightarrow_{top}$) and to Kleene-Dienese implication($\rightarrow_{bot}$) of m_1 system of logic, respectively, they yield the closed set of connectives as below [13],[10].

The definitions of group transformation use an involutive negation $\neg a = 1-a$ which satisfies all of four axioms below. An involutive negation, $\neg$, can be generated by a pair of Łukasiewicz implication($\rightarrow_{Top}$) and a constant 0 or by that of Kleene-Dienese implication($\rightarrow_{Bot}$) and constant(0):

$$\neg a = \begin{cases} a \rightarrow_{Top} 0 = a \xrightarrow{Ł} 0 = min(1, 1-a+0) = 1-a \\ a \rightarrow_{Bot} 0 = a \xrightarrow{KD} 0 = max(1-a, 0) = 1-a \end{cases}$$

However, both of them fall into a single point; there is no genuine interval of negation.

Table 1: Group transformation of m_1 system: $<\rightarrow_{top}, \rightarrow_{bot}>$.

Transformation of Connective	Type of Interval Bound
$g_{1t} = I(\rightarrow_{Top}) = min(1, 1-a+b)$	$\rightarrow_{Top}$
$g_{2t} = C(\rightarrow_{Top}) = min(1, 1+a-b)$	$\leftarrow_{Top}$
$g_{3t} = D(\rightarrow_{Top}) = max(0, b-a)$	$\not\leftarrow_{Bot}$
$g_{4t} = N(\rightarrow_{Top}) = max(0, a-b)$	$\not\rightarrow_{Bot}$
$g_{5t} = LC(\rightarrow_{Top}) = min(1, a+b)$	$\vee_{Top}$
$g_{6t} = LD(\rightarrow_{Top}) = max(0, 1-a-b)$	$\downarrow_{Bot}$
$g_{7t} = RC(\rightarrow_{Top}) = min(1, 2-a-b)$	$\mid_{Top}$
$g_{8t} = RD(\rightarrow_{Top}) = max(0, a+b-1)$	$\wedge_{Bot}$
$g_{1b} = I(\rightarrow_{Bot}) = max(1-a, b)$	$\rightarrow_{Bot}$
$g_{2b} = C(\rightarrow_{Bot}) = max(a, 1-b)$	$\leftarrow_{Bot}$
$g_{3b} = D(\rightarrow_{Bot}) = min(1-a, b)$	$\not\leftarrow_{Top}$
$g_{4b} = N(\rightarrow_{Bot}) = min(a, 1-b)$	$\not\rightarrow_{Top}$
$g_{5b} = LC(\rightarrow_{Bot}) = max(a, b)$	$\vee_{Bot}$
$g_{6b} = LD(\rightarrow_{Bot}) = min(1-a, 1-b)$	$\downarrow_{Top}$
$g_{7b} = RC(\rightarrow_{Bot}) = max(1-a, 1-b)$	$\mid_{Bot}$
$g_{8b} = RD(\rightarrow_{Bot}) = min(a, b)$.	$\wedge_{Top}$

Axiom 3 *Axioms of Fuzzy Negation [1]:*

Let a negation $\neg$ a be defined by a function

$$\neg : [0,1] \rightarrow [0,1]$$

A function $\neg$ should satisfy at least two of axiomatic requirements below:

A1. $\neg 0 = 1$ *and* $\neg 1 = 0$. : *boundary condition.*
A2. *For all* $a, b \in [0,1]$, *if* $a \leq b$, *then* $\neg a \geq \neg b$. : *monotonicity.*
A3. $\neg$ *is a continuous function.* : *continuity.*
A4. $\neg$ *is involutive, i.e.* $\neg(\neg a) = a$, $\forall a \in [0,1]$. : *involution.*

In the next section, we define various negations which yield intervals and investigate their properties.

4 Interval Negations in Many-Valued Logics

In [3], we have shown that the negations can also be defined by the Sheffer or the Nicod connectives as well as by the implication:

Definition 4 *Negation [3]:*

1. $\neg_S a = a \mid a$: by Sheffer
2. $\neg_N a = a \downarrow a$: by Nicod
3. $\neg_{PLY} a = a \rightarrow 0$: by Implication

In classical logic, negations by above definitions are all equal: $\neg a = \neg_S a = \neg_N a = \neg_{PLY} a = 1-a$. In many_valued logic, however, they do not yield the equal results – $\neg_S a \neq \neg_N a \neq \neg_{PLY} a$ – but these negations in checklist paradigm m_1 system generate a genuine interval negation [3].

ISBN: 978-989-95079-6-8

4.1 Negation on m_1 Defined by the Sheffer Connective

Since the Sheffer connective, $a \mid b$, of the interval system m_1 appears a TOP-BOT pair by means of RD transformation of $<\rightarrow_{Top}, \rightarrow_{Bot}>$ as it is shown at the table in section 3,

$$a \mid b = \begin{cases} a \mid_{Top} b = min(1, 2-a-b) \\ a \mid_{Bot} b = max(1-a, 1-b), \end{cases}$$

the connective type $\neg_S a$ also also appears in two forms:

$$\neg_S a = \begin{cases} a \mid_{Top} a = min(1, 2(1-a)) \\ a \mid_{Bot} a = 1-a \end{cases}$$

Thus, $\neg_S a$ defined by *Sheffer* generates the interval

$$[\, \mid_{Bot},\ \mid_{Top}\,] = [\, 1-a,\ \ min(1, 2(1-a))\,].$$

4.2 Negation in m_1 Defined by the Nicod Connective

Since the Nicod connective, $a \downarrow b$, appears as a TOP-BOT pair of connectives in the interval system m_1 , similarly

$$a \downarrow b = \begin{cases} a \downarrow_{Top} b = min(1-a, 1-b) \\ a \downarrow_{Bot} b = max(0, 1-a-b), \end{cases}$$

the connective type $\neg_N a$ also also appears in two forms:

$$\neg_N a = \begin{cases} a \downarrow_{Top} a = 1-a & = D(a \mid_{Bot} a) \\ a \downarrow_{Bot} a = max(0, 1-2a) & = D(a \mid_{Top} a). \end{cases}$$

Thus, the interval of $\neg_N a$ by *Nicod* is:

$$[\downarrow_{Bot},\ \downarrow_{Top}] = [max(0, 1-2a),\ 1-a].$$

4.3 Negation in m_1 Defined by the Implication Connective

Since $a \overset{KD}{\rightarrow} b \leq m_1(\rightarrow) \leq a \overset{Ł}{\rightarrow} b$, in $a \rightarrow b$ of m_1 system, i.e. the BOT connective is the Kleene-Dienes implication operator while the TOP id Łukasiewicz, we have:

$$\neg_{PLY} a = \begin{cases} a \rightarrow_T 0 = a \overset{Ł}{\rightarrow} 0 = min(1, 1-a+0) = 1-a \\ a \rightarrow_B 0 = a \overset{KD}{\rightarrow} 0 = max(1-a, 0) = 1-a \end{cases}$$

4.4 Interval of Negation in m_1 logic system

Two intervals of negation defined by the *Sheffer* connective and by the *Nicod* connective may be combined to yield a genuine interval of negation since $1-a$ is the lower bound of *Sheffer* negation and the upper bound of *Nicod* negation, respectively. These two intervals are concatenated as below:

$$\begin{aligned} [\neg_{BOT},\ \neg_{TOP}] &= [\downarrow_{Bot},\ \downarrow_{Top}] \cup [\, \mid_{Bot},\ \mid_{Top}\,] \\ &= [\ \max(0,\ 1\text{-}2a),\ \ \min(1,\ 2(1\text{-}a))\ ] \\ &= \begin{cases} [\, 0,\ 2(1-a)\,], & 1 \geq a \geq .5 \\ [\, 1-2a,\ 1\,] & .5 > a \geq 0 \end{cases} \\ \neg_{MID} &= \neg_{PLY} a = a \downarrow_{Top} a = a \mid_{Bot} a = \ 1-a. \end{aligned}$$

Thus, *Sheffer*($\mid$) and *Nicod*($\downarrow$) form the TOP system of negation and the BOT system of negation, respectively. The $\neg_{PLY}$ $a = 1-a$, is a median value of the interval negation, $\neg_{Mid}$. Since both $\neg_{Top}$ and $\neg_{Bot}$ satisfy A1 - A3 of *Axioms of fuzzy negation* except A4 of sec.3, they are non-involutive TOP-BOT pair of fuzzy interval negations while $\neg_{Mid}$ is an involutive fuzzy negation.

Theorem 5 *Gap Theorem 1. (Bandler and Kohout [6])*

$$\begin{aligned} a \text{ AND}_{Top}\ b - a \text{ AND}_{Bot}\ b &= a \text{ OR}_{Top}\ b - a \text{ OR}_{Bot}\ b \\ &= a \text{ PLY}_{Top}\ b - a \text{ PLY}_{Bot}\ b \\ &= min(\varphi a,\ \varphi b). \\ a \text{ IFF}_{Top}\ b - a \text{ IFF}_{Bot}\ b &= a \text{ EOR}_{Top}\ b - a \text{ EOR}_{Bot}\ b \\ &= 2min(\varphi a,\ \varphi b). \end{aligned}$$

Similarly, we can describe Gap Theorem for a TOP-BOT pair of negation.

Theorem 6 *Gap Theorem 2.*

$$\neg_{TOP}\ a - \neg_{BOT}\ a = min(2a, 2(1-a)) = 2\varphi a$$

Hence, the margins of imprecision can be directly measured by the degree of fuzziness φ where $\varphi a = min(a, 1-a)$.

To further investigate a non-involutive properties of $[\neg_{Bot}, \neg_{Top}]$ pair, we apply multiple nagation. It leads to the iterative negations, interrupting mathematical properties and mathematical limit, but yields an interesting sequence of values of interval. In the following sections, we use the symbols $\neg_T$ and $\neg_B$ for $\neg_{Top}$ and $\neg_{Bot}$, respectively, for simplicity.

5 Multiple negations

As both $\neg_{Bot}(= \neg_B)$ and $\neg_{Top}(= \neg_T)$ are non-involutive, a question arise what would happen if a negation is iterated to $\neg_B$ and $\neg_T$, respectively: namely, a multiple negation.

Definition 7 *Multiple Negations.*

1. $\neg^1 a = \neg\, a.$
2. $\neg^n a = (\neg\, a)^n = \neg\, (\neg^{n-1}\, a)$
$= \neg(\neg(\neg^{n-2}\, a))$
$= \underbrace{\neg(\neg(\neg \ldots, (\neg}_{n}\, a))), \quad 1 \leq n$

Thus, multiple negations will be applied to $\neg_T$ and $\neg_B$, respectively in the following subsections in order to explore their properties from the computational results. The following two subsections summarize the computational results of multiple negations of $\neg_{Top}a$ and $\neg_{Bot}a$, respectively .

5.1 Multiple negation of $\neg_{TOP}$

$$\begin{aligned} \neg_T a = a \mid_T a &= min(1, 2(1-a)) \\ &= \begin{cases} 2(1-a), & a \in [\frac{1}{2},\ 1] \\ 1 & a \in [0, \frac{1}{2}) \end{cases} \end{aligned} \tag{1}$$

$$\begin{aligned} \neg_T^2 a = \neg_T(\neg_T a) &= min(1,\ 2(1-\neg_T a)) \\ &= \begin{cases} 1, & a \in (\frac{3}{4}, 1] \\ 2(2a-1), & a \in [\frac{1}{2},\ \frac{3}{4}] \\ 0 & a \in [0,\ \frac{1}{2}) \end{cases} \end{aligned} \tag{2}$$

$$\begin{aligned} \neg_T^3 a = \neg_T(\neg_T^2 a) &= min(1,\ 2(1-\neg_T^2 a))) \\ &= \begin{cases} 0, & a \in (\frac{3}{4},\ 1] \\ 2(3-4a), & a \in [\frac{5}{8},\ \frac{3}{4}] \\ 1 & a \in [0,\ \frac{5}{8}) \end{cases} \end{aligned} \tag{3}$$

$$\begin{aligned} \neg_T^4 a = \neg_T(\neg_T^3 a) &= min(1,\ 2(1-\neg_T^3 a))) \\ &= \begin{cases} 1, & a \in (\frac{11}{16},\ 1] \\ 2(8a-5), & a \in [\frac{5}{8},\ \frac{11}{16}] \\ 0 & a \in [0,\ \frac{5}{8}) \end{cases} \end{aligned} \tag{4}$$

ISBN: 978-989-95079-6-8

$$\neg_T^5 a = \neg_T(\neg_T^4 a) \quad = min(1,\ 2(1-\neg_T^4 a))) = \begin{cases} 0, & a \in (\frac{11}{16},\ 1] \\ 2(11-16a), & a \in [\frac{21}{32},\ \frac{11}{16}] \\ 1 & a \in [0,\ \frac{21}{32}] \end{cases} \tag{5}$$

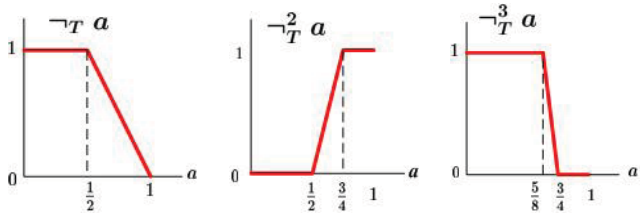


Figure 1: $\neg_T\ a,\ \neg_T^2\ a,\ \neg_T^3\ a$

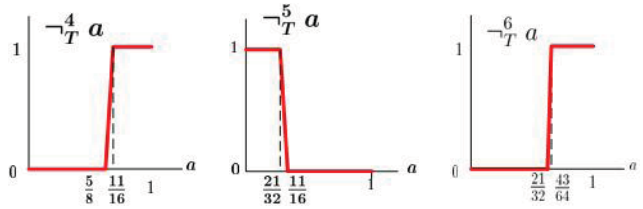


Figure 2: $\neg_T^4\ a,\ \neg_T^5\ a,\ \neg_T^6\ a$

If we further generalize it to $\neg_T^n\ a$ both for $n = 2k$ and for $n = 2k+1, \forall k \geq 1$, respectively, we get the following formulas (6)-(7).

$$\neg_T^{2k}\ a = \neg_T(\neg_T^{2k-1} a) \quad = \begin{cases} 1, & a \in (u_1,\ 1] \\ v_0, & a \in [u_2,\ u_1] \\ 0 & a \in [0,\ u_2) \end{cases} \tag{6}$$

where $v_0 = 2^{2k}a - \frac{2}{3}(2^{2k}-1)$,

$u_1 = \frac{2\cdot 4^k+1}{3\cdot 4^k}$,

$u_2 = \frac{2(4^k-1)}{3\cdot 4^k}$,

and

$$\neg_T^{2k+1}\ a = \neg_T(\neg_T^{2k} a) \quad = \begin{cases} 0, & a \in (u_3,\ 1] \\ v_1 & a \in [u_4,\ u_3] \\ 1 & a \in [0,\ u_4) \end{cases} \tag{7}$$

where $v_1 = \frac{2}{3}(2^{2k+1}+1) - 2^{2k+1}a$,

$u_3 = u_1 = \frac{2\cdot 4^k+1}{3\cdot 4^k}$,

$u_4 = \frac{4^{k+1}-1}{6\cdot 4^k}$.

Thus, any multiple negations in the odd sequence of $\neg_T^{2k+1}$, $\forall k \geq 0$, i.e. $< \neg_T a, \neg_T^3 a, \ldots, \neg_T^{2k+1} a, \ldots >$, are monotonically decreasing(**Axiom A2**) while those in the even sequence of $\neg_T^{2k}$, $\forall k \geq 1$, $< \neg_T^2 a, \neg_T^4\ a, \ldots, \neg_T^{2k}\ a, \ldots >$, are monotonically increasing(**A2**). All of multiple negations satisfy the boundary condition(**A1**) and continuity(**A3**) .

5.2 *Multiple negation of* $\neg_{BOT}$

$$\neg_{bot} a = a \downarrow_B a \quad = max(0, 1-2a) = \begin{cases} 0, & a \in (\frac{1}{2},\ 1] \\ 1-2a & a \in [0,\ \frac{1}{2}] \end{cases} \tag{8}$$

$$\neg_B^2 a = \neg_B(\neg_B a) \quad = max(0, 1-2\cdot\neg_B a) = \begin{cases} 1, & a \in (\frac{1}{2},\ 1] \\ 4a-1, & \in [\frac{1}{4},\ \frac{1}{2}] \\ 0 & a \in [0,\ \frac{1}{4}) \end{cases} \tag{9}$$

$$\neg_B^3 a = \neg_B(\neg_B^2\ a) \quad = max(0, 1-2\cdot\neg_B^2\ a) = \begin{cases} 0, & a \in (\frac{3}{8},\ 1] \\ 3-8a, & a \in [\frac{1}{4},\ \frac{3}{8}] \\ 1 & a \in [0,\ \frac{1}{4}) \end{cases} \tag{10}$$

$$\neg_B^4 a = \neg_B(\neg_B^3\ a) \quad = max(0,\ 1-2\cdot\neg_B^3 a) = \begin{cases} 1, & a \in (\frac{3}{8},\ 1] \\ 16a-5, & a \in [\frac{5}{16},\ \frac{3}{8}] \\ 0 & a \in [0,\ \frac{5}{16}) \end{cases} \tag{11}$$

$$\neg_B^5 a = \neg_B(\neg_B^4\ a) \quad = max(0,\ 1-2\cdot\neg_B^4 a) = \begin{cases} 0, & a \in (\frac{11}{32}\ 1] \\ 11-32a, & a \in [\frac{5}{16},\ \frac{11}{32}] \\ 1 & a \in [0,\ \frac{5}{16}) \end{cases} \tag{12}$$

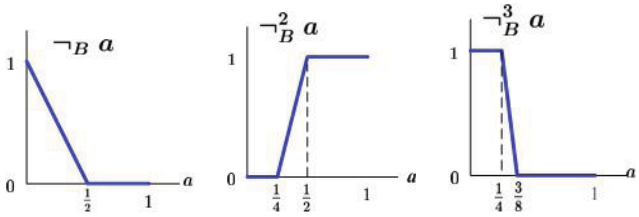


Figure 3: $\neg_B\ a,\ \neg_B^2\ a,\ \neg_B^3\ a$

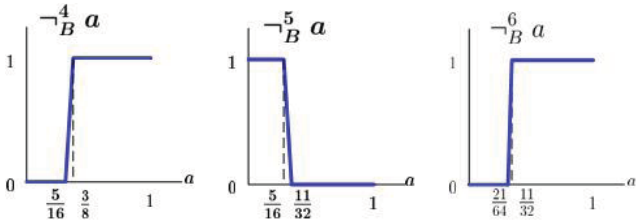


Figure 4: $\neg_B^4\ a,\ \neg_B^5\ a,\ \neg_B^6\ a$

If we further generalize it to $\neg_T^n\ a$ both for $n = 2k$ and for $n = 2k+1, \forall k \geq 1$, respectively, we get the following formulas (13)-(14).

$$\neg_B^{2k}\ a = \neg_B(\neg_B^{2k-1} a) \quad = \begin{cases} 1, & a \in (w_1,\ 1] \\ y_0, & a \in [w_2,\ w_1] \\ 0 & a \in [0,\ w_2) \end{cases} \tag{13}$$

where $y_0 = 2^{2k}a - \frac{1}{3}(2^{2k}-1)$,

$w_1 = \frac{4^k+2}{3\cdot 4^k}$,

$w_2 = \frac{4^k-1}{3\cdot 4^k}$,

ISBN: 978-989-95079-6-8

and

$$\neg_B^{2k+1} a = \neg_B(\neg_B^{2k} a) = \begin{cases} 0, & a \in (w_3,\ 1] \\ y_1 & a \in [w_4,\ w_3] \\ 1 & a \in [0,\ w_4) \end{cases} \tag{14}$$

where $y_1 = \frac{1}{3}(2^{2k+1}+1) - 2^{2k+1}a$,

$w_3 = \frac{2\cdot 4^k+1}{6\cdot 4^k}$,

$w_4 = w_2 = \frac{4^k-1}{3\cdot 4^k}$.

Similar to TOP system of negation in sec. 5.1., any multiple negations in the odd sequence are monotonically decreasing while those in the even sequence are monotonically increasing. All of them satisfy the boundary condition and continuity. Both TOP-BOT systems of negation also show the following property in the interval of a.

Property 8 *The width of interval of a:*
The width of interval of a in eq.(6)-(7) and eq.(13)-(14) is:

1. *For* $n = 2k$: $\neg_T^{2k} a$ *and* $\neg_B^{2k} a$,
$$| u_1 - u_2 | = | w_1 - w_2 | = \tfrac{1}{2^{2k}} = \tfrac{1}{2^n},$$

2. *For* $n = 2k+1$: $\neg_T^{2k+1} a$ *and* $\neg_B^{2k+1} a$,
$$| u_3 - u_4 | = | w_3 - w_4 | = \tfrac{1}{2^{2k+1}} = \tfrac{1}{2^n}.$$

Theorem 9 *A relationship between* $\neg^n$ *and* $\neg^{n+1}$*:*

(1) $1 - \neg_T^n a = \frac{1}{2}\neg_T^{n+1} a,\ \forall n \in N,\ a \in [max(u_2,u_4),\ u_1]$ *in eq.(6)-(7).*

(2) $1 - \neg_B^{n+1} a = 2 \cdot \neg_B^n a,\ \forall n \in N,\ a \in [w_2,\ min(w_1,w_3)]$ *in eq.(13)-(14).*

Proof:

(1) (*a*) Let $n = 2k,\ \forall k \geq 1$.
$$\begin{aligned} 1 - \neg_T^{2k} a &= 1 - (2^{2k}a - \tfrac{2}{3}(2^{2k}-1)) \\ &= \tfrac{1}{3}(2^{2k+1}+1) - 2^{2k}a \\ &= \tfrac{1}{2}(\tfrac{2}{3}(2^{2k+1}+1) - 2^{2k+1}a) \\ &= \tfrac{1}{2}\cdot\neg_T^{2k+1} a \end{aligned}$$
(*b*) Let $n = 2k+1, \forall k \geq 0$.
$$\begin{aligned} 1 - \neg_T^{2k+1} a &= 1 - (\tfrac{2}{3}(2^{2k+1}+1) - 2^{2k+1}a) \\ &= 2^{2k+1}a - \tfrac{1}{3}(2^{2k+2}-1) \\ &= \tfrac{1}{2}(2^{2k+2}a - \tfrac{2}{3}(2^{2k+2}-1)) \\ &= \tfrac{1}{2}\neg_T^{2k+2} a \end{aligned}$$
Therefore, $1 - \neg_T^n a = \frac{1}{2}\neg_T^{n+1} a,\ ,\ \forall n \in N$. □

(2) (*a*) $n = 2k-1, \forall k \geq 1$.
$$\begin{aligned} 1 - \neg_B^{2k} a &= 1 - (2^{2k}a - \tfrac{1}{3}(2^{2k}-1)) \\ &= \tfrac{2}{3}(2^{2k-1}+1) - 2^{2k}a \\ &= 2(\tfrac{1}{3}(2^{2k-1}+1) - 2^{2k-1}a) \\ &= 2\cdot\neg_B^{2k-1} a \end{aligned}$$
(*b*) $n = 2k, \forall k \geq 1$.
$$\begin{aligned} 1 - \neg_B^{2k+1} a &= 1 - (\tfrac{1}{3}(2^{2k+1}+1) - 2^{2k+1}a) \\ &= 2^{2k+1}a - \tfrac{2}{3}(2^{2k}-1) \\ &= 2(2^{2k}a - \tfrac{1}{3}(2^{2k}-1)) \\ &= 2\cdot\neg_B^{2k} a \end{aligned}$$
Therefore, $1 - \neg_B^{n+1} a = 2 \cdot \neg_B^n a,\ \forall n \in N$. □

5.3 *Convergence of multiple negations*

In order to investigate a convergence of multiple negations, let us apply lim to $\neg_T^{2k} a$, $\neg_T^{2k+1} a$, $\neg_B^{2k} a$ and to $\neg_B^{2k+1} a$ as $k \to \infty$, respectively. It derives the following formula of convergence in each case:

$$\lim_{k\to\infty} \neg_T^{2k} a = \begin{cases} 1, & a \in (\frac{2}{3}^+,\ 1] \\ z_0, & a \in [\frac{2}{3}^-,\ \frac{2}{3}^+] \\ 0 & a \in [0,\ \frac{2}{3}^-) \end{cases} \tag{15}$$

where $z_0 = \lim_{k\to\infty} v_0 = \lim_{k\to\infty} 2^{2k}a - \frac{2}{3}(2^{2k}-1)$, in eq.(6)
$z_0 = \frac{2}{3}$ if $a = \frac{2}{3}$, in particular,
and $\frac{2}{3}^+ = \lim_{k\to\infty} u_1 = \frac{2}{3} + \varepsilon_1$,
$\frac{2}{3}^- = \lim_{k\to\infty} u_2 = \frac{2}{3} - \varepsilon_1$, for a very small $\varepsilon_1 > 0$.

$$\lim_{k\to\infty} \neg_T^{2k+1} a = \begin{cases} 0, & a \in (\frac{2}{3}^+,\ 1] \\ z_1 & a \in [\frac{2}{3}^-,\ \frac{2}{3}^+] \\ 1 & a \in [0,\ \frac{2}{3}^-) \end{cases} \tag{16}$$

where $z_1 = \lim_{k\to\infty} v_1 = \lim_{k\to\infty} \frac{2}{3}(2^{2k+1}+1) - 2^{2k+1}a$, in eq.(7)
$z_1 = \frac{2}{3}$ if $a = \frac{2}{3}$, in particular,
and $\frac{2}{3}^+ = \lim_{k\to\infty} u_3 = \frac{2}{3} + \varepsilon_2$,
$\frac{2}{3}^- = \lim_{k\to\infty} u_4 = \frac{2}{3} - \varepsilon_2$, for a very small $\varepsilon_2 > 0$
where $\varepsilon_2 < \varepsilon_1$.

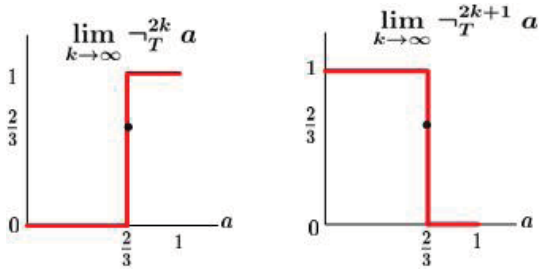


Figure 5: $\lim_{k\to\infty} \neg_T^{2k} a$ and $\lim_{k\to\infty} \neg_T^{2k+1} a$

As $k \to \infty$. both lim $\neg_T^{2k} a$ and lim $\neg_T^{2k+1} a$ converge either to 0 or to 1, depending on the values of a in the most of intervals, except $a \in [\frac{2}{3} - \varepsilon_i,\ \frac{2}{3} + \varepsilon_i]$ for $i = 1, 2$ where ε_i is in eq.(15)-(16). For an a in the interval $[\frac{2}{3} - \varepsilon_i,\ \frac{2}{3} + \varepsilon_i]$, the values of $\lim_{k\to\infty} \neg_T^{2k} a$ or those of $\lim_{k\to\infty} \neg_T^{2k+1} a$ increases or decreases drastically in monotonicity, respectively.

We can investigate the limit negations of $\neg_B^{2k} a$ and $\neg_B^{2k+1} a$ of BOT system, similarly.

$$\lim_{k\to\infty} \neg_B^{2k} a = \begin{cases} 1, & a \in (\frac{1}{3}^+,\ 1] \\ z_2, & a \in [\frac{1}{3}^-,\ \frac{1}{3}^+] \\ 0 & a \in [0,\ \frac{1}{3}^-) \end{cases} \tag{17}$$

where $z_2 = \lim_{k\to\infty} y_0 = \lim_{k\to\infty} 2^{2k}a - \frac{1}{3}(2^{2k}-1)$, in eq.(13)
$z_2 = \frac{1}{3}$ if $a = \frac{1}{3}$, in particular,
and $\frac{1}{3}^+ = \lim_{k\to\infty} w_1 = \frac{1}{3} + \delta_1$,
$\frac{1}{3}^- = \lim_{k\to\infty} w_2 = \frac{1}{3} - \delta_1$, for a very small $\delta_1 > 0$.

ISBN: 978-989-95079-6-8

$$\lim_{k\to\infty} \neg_B^{2k+1} a = \begin{cases} 0, & a \in (\frac{1}{3}^+,\ 1] \\ z_3, & a \in [\frac{1}{3}^-,\ \frac{1}{3}^+] \\ 1 & a \in [0,\ \frac{1}{3}^-) \end{cases} \quad (18)$$

where $z_3 = \lim_{k\to\infty} y_1 = \lim_{k\to\infty} \frac{1}{3}(2^{2k+1}+1) - 2^{2k+1}a$, in eq.(14)
$z_3 = \frac{1}{3}$ if $a = \frac{1}{3}$, in particular,
and $\frac{1}{3}^+ = \lim_{k\to\infty} w_3 = \frac{1}{3} + \delta_2$,
$\frac{1}{3}^- = \lim_{k\to\infty} w_4 = \frac{1}{3} - \delta_2$, for a very small $\delta_2 > 0$
where $\delta_2 < \delta_1$.

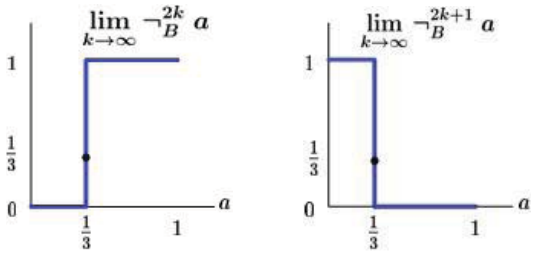


Figure 6: $\lim_{k\to\infty} \neg_B^{2k} a$ and $\lim_{k\to\infty} \neg_B^{2k+1} a$

As $k \to \infty$. both $\lim \neg_B^{2k} a$ and $\lim \neg_B^{2k+1} a$ converge either to 0 or to 1 for the most of $a \in [0,\ 1]$, except $a \in [\frac{1}{3} - \delta_i,\ \frac{1}{3} + \delta_i]$, for $i = 1, 2$ where δ_i is in eq.(17)-(18). For an a in $[\frac{1}{3} - \delta_i,\ \frac{1}{3} + \delta_i]$, the values of $\lim_{k\to\infty} \neg_B^{2k} a$ or those of $\lim_{k\to\infty} \neg_B^{2k+1} a$ are in the drastic monotonic increase or decrease, respectively. Thus, we can further describe the following properties.

Property 10 *For $n \in N$,*

P1. $\lim_{n\to\infty} \neg_T^{n+1} a \approx 1 - \lim_{n\to\infty} \neg_T^n a, \quad a \in [0,\ 1]$.

P2. $\lim_{n\to\infty} \neg_B^{n+1} a \approx 1 - \lim_{n\to\infty} \neg_B^n a, \quad a \in [0,\ 1]$.

P3. $\lim_{k\to\infty} \neg_T^n a \approx \lim_{n\to\infty} \neg_B^n (a + \frac{1}{3}), \quad a \in [0,\ 1]$.

P4. $\lim_{n\to\infty} \neg_B^n a \approx \lim_{k\to\infty} \neg_T^n (a - \frac{1}{3}), \quad a \in [0,\ 1]$.

From the above **P1** and **P2**, we can notice that both $\neg_T$ and $\neg_B$ are *nearly limit involutive* because $\lim_{n\to\infty} \neg_T^{n+1} a \approx 1 - \lim_{n\to\infty} \neg_T^n a \approx 1 - (1 - \lim_{n\to\infty} \neg_T^{n-1} a) \approx \lim_{n\to\infty} \neg_T^{n-1} a, \quad a \in [0,\ 1]$, i.e. $\neg_T(\neg_T(\lim_{n\to\infty} \neg_T^{n-1}) \approx \lim_{n\to\infty} \neg_T^{n-1}$. It holds for $\neg_B$, similarly.

6 Conclusions

The involutive fuzzy negation, $\neg a = 1 - a$, has been collapsed into a single point while ten of binary connectives yield TOP-BOT interval pairs in m_1 logic system of 16 connectives by group of logic transformations since Bandler-Kohout proposed five checklist paradigm based logic systems $m_1 - m_5$. A BOT-TOP pair of connectives of fuzzy negation interval $[\neg_{BOT},\ \neg_{TOP}]$, however, could be successfully generated from the interval of *Nicod* system $[\downarrow_{top},\ \downarrow_{bot}]$ and that of *Sheffer* system $[|_{bot},\ |_{top}]$. These pairs of negation connective are non-involutive negation pairs by themselves. When a negation is applied iteratively, those multiple negations, $\neg_T^n a, \neg_B^n a$ change their values from 0 to 1 or 1 to 0 drastically within a very small interval whose width is $\frac{1}{2^n}$ and reveals a property of nearly limit involution.

References

[1] G.J. Klir and B. Yuan Fuzzy sets and fuzzy logic: Theory and applications. Prentice Hall, 1995.

[2] W. Bandler and L.J. Kohout. Semantics of implication operators and fuzzy relational products. *Internat. Journal of Man-Machine Studies*, 12:89–116, 1980. Reprinted in Mamdani, E.H. and Gaines, B.R. eds. *Fuzzy Reasoning and its Applications*. Academic Press, London, 1981, pp. 219-246.

[3] E. Kim and L.J. Kohout. Interval Negation in Fuzzy Logics. In *Proc. of NAFIPS (North American Fuzzy Information Processing Society)*, pp. 537-542, San Diego, CA, June 24-27 2007. IEEE.

[4] W. Bandler and L.J. Kohout. Unified theory of multiple-valued logical operators in the light of the checklist paradigm. In *Proc. of the 1984 IEEE Conference on Systems, Man and Cybernetics*, pp. 356–364, New York, 1984. IEEE.

[5] W. Bandler and L.J. Kohout. The interrelations of the principal fuzzy logical operators. In M.M. Gupta, A. Kandel, W. Bandler, and J.B. Kiszka, eds., *Approximate Reasoning in Expert Systems*, pp. 767–780. North-Holland, Amsterdam, 1985.

[6] W. Bandler and L.J. Kohout. The use of checklist paradigm in inference systems. In C.V. Negoita and H. Prade, eds., *Fuzzy Logic in Knowledge Engineering*, chapter 7, pp. 95–111. Verlag TÜV Rheinland, Köln, 1986.

[7] P. Hájek and L.J. Kohout. Fuzzy implications and generalized quantifiers. *Internat. Journal of Uncertainty, Fuzziness and Knowledge Based Systems*, 4(3):225-233, 1996.

[8] L.J. Kohout and E. Kim. On the emergence of fuzzy logics from elementary percepts: the checklist paradigm and the theory of perceptions. In *Forging New Frontiers: Fuzzy Pioneers I. Studies in Fuzziness and Soft Computing*, 217:437-460, Springer, 2007.

[9] L.J. Kohout and E. Kim. Group transformations of systems of logic connectives. In *Proc. of IEEE-FUZZ'97*, 1:157–162, IEEE, New York, July 1997.

[10] L.J. Kohout and E. Kim. Characterization of interval fuzzy logic systems of connectives by group transformations. *Reliable Computing*, 10:299–334, 2004.

[11] L.J. Kohout and W. Bandler. Checklist paradigm and group transformations. *Technical Note EES-MMS-ckl91.2*, 1979. Dept. of Electrical Engineering, University of Essex. U.K.

[12] L.J. Kohout and W. Bandler. How the checklist paradigm elucidates the semantics of fuzzy inference. In *Proc. of the IEEE Internat. Conference on Fuzzy Systems 1992*, pp. 571–578. IEEE, New York, 1992.

[13] L.J. Kohout and E. Kim. Global characterization of fuzzy logic systems with para-consistent and gray set features. In P. Wang, ed., *Proc. 3rd Joint Conf. on Information Sciences JCIS'97 (5th Int. Conf. on Fuzzy Theory and Technology)*, pp. 238–241, Research Triangle Park, NC, March 1997. Duke University. In volume 1: Fuzzy Logic, Intelligent Control and Genetic Algorithms.

ISBN: 978-989-95079-6-8

Development of Multiple Linguistic Equation Models with Takagi-Sugeno Type Fuzzy Models

Esko K. Juuso
Control Engineering Laboratory
Department of Process and Environmental Engineering
University of Oulu, Finland
Email: esko.juuso@oulu.fi

Abstract— *Multimodel approaches are widely used with linear submodels, but border areas around submodels are problematic. Special cases of fuzzy linguistic equation models, which can be understood as linguistic Takagi-Sugeno (LTS) type fuzzy models, can be used to solve these problems in many cases. These models use a special nonlinear scaling approach for both inputs and outputs. The LTS models are robust solutions for applications where the same variables can be used for defining operating areas and also in the submodels. No special smoothing algorithms are needed.*

Keywords— linguistic equations, nonlinear systems, multimodels, Takagi-Sugeno fuzzy models.

1 Introduction

Fuzzy set systems enable the use of expert system techniques in uncertain and vague systems [1]. The traditions of physical modelling on the basis of understanding system behaviour are maintained with fuzzy rules and membership functions, which can represent not only gradually changing nonlinear mappings but also abrupt changes [2]. Various approaches using either expertise or data are used in constructing these mappings, but as the complexity of the application increases more and more combined approaches are required. These approaches are presented in the upper right-hand corner of Fig. 1. Heuristic knowledge and know-how can be introduced to fuzzy set systems with a trial and error based approach. Data-based approaches rely usually on automatic generation of rules from predefined simple sets of membership functions. Both knowledge and data need to be used together to develop practical applications.

Linguistic fuzzy models [3] are mainly used in the knowledge-based approach, whereas Takagi-Sugeno (TS) type fuzzy models [4] and fuzzy relational models [5] are mainly used for data-driven methods. TS models are constructed by combining supervised and unsupervised learning. The antecedent membership functions can be initialised by grid partioning, iterative search, or fuzzy clustering [6]. The neuro-fuzzy ANFIS method [7] is widely used in tuning. The structure and the parameters can also be updated recursively when new data become available [8]. The trade-off between global model accuracy and interpretability of local models as the linearisations of a nonlinear system is important in the development of a TS model. To restrict the freedom of the parameters a multiobjective identification for dynamic TS models is presented in [9].

This paper presents a set of fuzzy linguistic equation models, which combine nonlinear scaling and Takagi-Sugeno type fuzzy models.

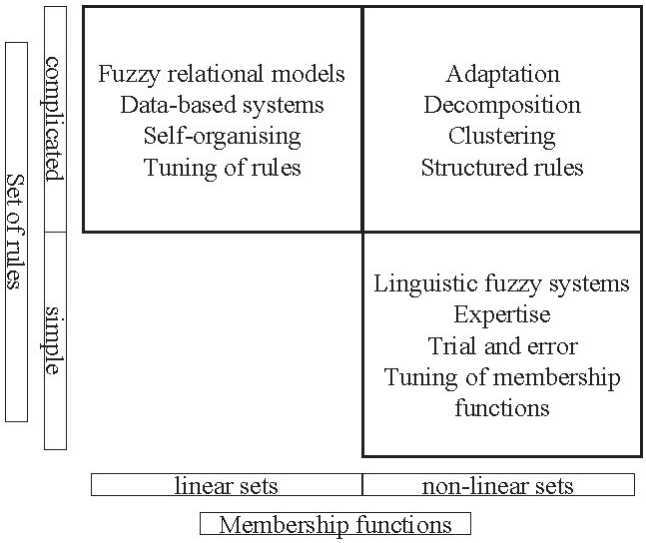


Figure 1: Classification of fuzzy set systems [2].

2 Methodologies

2.1 Linguistic equations

The linguistic equation (LE) approach originates from fuzzy set systems: rule sets are replaced with equations, and the effects of membership functions are handled with scaling [10]. For nonlinear models, the scaling technique must be nonlinear as the model equations are linear. The scaling functions are called membership definitions as they are closely connected to the membership functions used in fuzzy set systems. [2]

Nonlinear scaling. Nonlinear scaling is carried out using functions which denote membership definitions. The mapping function is performed by converting variable values from the variable range into a range [-2, 2], known as the *linguistic range*. The new range describes the distribution of variable values over the original range fairly accurately. Membership definitions are presented by a variable specific function

$$x_j = f(X_j) \,\forall\, \min(x_j) \le x_j \le \max(x_j),\ X_j \in [-2, 2], \tag{1}$$

where x_j is the value of variable j, and X_j is the corresponding value within the range $[-2,\ 2]$, which includes the normal operation range $[-1,\ 1]$ and areas for warnings and alarms. The values X_j are called *linguistic values* because the scaling function is based on the membership functions of fuzzy set systems: values -2, -1, 0, 1 and 2 can be associated to the

ISBN: 978-989-95079-6-8

linguistic labels, e.g.

$$\{very\,low,\ low,\ normal,\ high,\ very\,high\} \qquad (2)$$

are defined with membership functions (Fig. 2). The number of membership functions is not limited to five: the values between these integers correspond to finer partitions of the fuzzy set system. Early applications of linguistic equations only used integer values [10].

In the case of polynomial membership definitions,

$$\begin{array}{ll} f_j^- = a_j^- X_j^2 + b_j^- X_j + c_j, & X_j \in [-2, 0), \\ f_j^+ = a_j^+ X_j^2 + b_j^+ X_j + c_j, & X_j \in [0, 2], \end{array} \quad . \qquad (3)$$

the linguistic level of the input variable j is calculated according to the equation

$$X_j = \begin{cases} 2 & \text{with } x_j \geq \max x_j \\ \frac{-b_j^+ + \sqrt{b_j^{+2} - 4a_j^+ (c_j - x_j)}}{2a_j^+} & \text{with } c_j \leq x_j \leq \max(x_j) \\ \frac{-b_j^- + \sqrt{b_j^{-2} - 4a_j^- (c_j - x_j)}}{2a_j^-} & \text{with } \min(x_j) \leq x_j \leq c_j \\ -2 & \text{with } x_j \leq \min(x_j). \end{cases} \qquad (4)$$

where a_j^-, b_j^-, a_j^+, and b_j^+ are coefficients of the polynomials (3), c_j is the real value corresponding to the linguistic value 0, and x_j is the real value. $\min(x_j)$ and $\max(x_j)$ are the minimum and maximum values of the real data corresponding to the linguistic values -2 and 2.

After the linguistic level of the model output, X_{out}, is calculated using the linguistic equation model, it is converted into a real value of output, x_{out}, using the following equation:

$$x_{out} = \begin{cases} a_{out}^- {X_{out}}^2 + b_{out}^- X_{out} + c_{out} \\ \qquad \text{with } X_{out} < 0 \\ a_{out}^+ {X_{out}}^2 + b_{out}^+ X_{out} + c_{out} \\ \qquad \text{with } X_{out} \geq 0 \end{cases} \qquad (5)$$

where a_{out}^-, b_{out}^-, a_{out}^+ and b_{out}^+ are coefficients of the polynomials (3), and c_{out} is the real value corresponding to the linguistic value 0.

The coefficients of the polynomials can be represented by

$$\begin{array}{rcl} a_j^- & = & \frac{1}{2}(1 - \alpha_j^-)\,\Delta c_j^-, \\ b_j^- & = & \frac{1}{2}(3 - \alpha_j^-)\,\Delta c_j^-, \\ a_j^+ & = & \frac{1}{2}(\alpha_j^+ - 1)\,\Delta c_j^+, \\ b_j^+ & = & \frac{1}{2}(3 - \alpha_j^+)\,\Delta c_j^+, \end{array} \qquad (6)$$

where $\Delta c_j^- = c_j - (c_l)_j$ and $\Delta c_j^+ = (c_h)_j - c_j$. Membership definitions may contain linear parts if some coefficients α_j^- or α_j^+ equal one.

Membership definitions are determined by the centre point c_j, and the core and support areas, which guarantee that the resulting membership definitions are monotonously increasing functions. An easier way is to define the centre point, the core $[(c_l)_j,\ (c_h)_j]$ and the ratios

$$\begin{array}{l} \alpha_j^- = \frac{(c_l)_j - \min(x_j)}{c_j - (c_l)_j}, \\ \alpha_j^+ = \frac{\max(x_j) - (c_h)_j}{(c_h)_j - c_j}, \end{array} \qquad (7)$$

from the range $\frac{1}{3} \ldots 3$, and calculate the support $[\min(x_j),\ \max(x_j)]$. The membership definitions of each variable are configured with five parameters, which can presented in three consistent sets. The working point (centre point) c_j belongs to all these sets, and the others are:

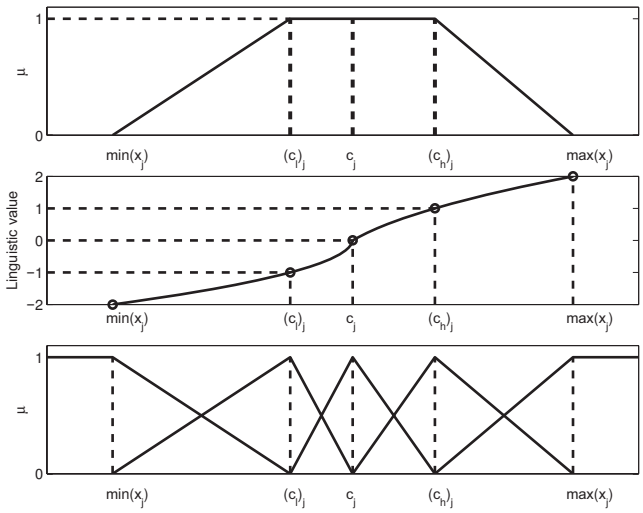


Figure 2: Feasible range, membership definitions and membership functions [2].

- the corner points $\{\min(x_j),\ (c_l)_j,\ (c_h)_j,\ \max(x_j)\}$ are good for visualisation;
- the parameters $\{\alpha_j^-,\ \Delta c_j^-,\ \alpha_j^+,\ \Delta c_j^+\}$ are suitable for tuning;
- the coefficients $\{a_j^-,\ b_j^-,\ a_j^+,\ b_j^+\}$ are used in calculations.

The upper and the lower parts of the scaling functions can be convex or concave, independent of each other. Simplified functions can also be used: a linear membership definition only requires two parameters: c_j and $b_j = b_j^+ = b_j^-$ or $\Delta c_j = \Delta c_j^+ = \Delta c_j^-$, since $\alpha_j^+ = \alpha_j^- = 1$ and $a_j^+ = a_j^- = 0$; an asymmetrical linear definition has $\Delta c_j^+ \neq \Delta c_j^-$ and $b_j^+ \neq b_j^-$.

Interactions. The basic element of a linguistic equation (LE) model is a compact equation

$$\sum_{j=1}^{m} A_{ij} X_j + B_i = 0, \qquad (8)$$

where X_j is a linguistic level of the variable j, $j = 1 \ldots m$. The direction of the interaction is represented by the interaction coefficients A_{ij}. The bias term B_i was introduced for fault diagnosis systems. A LE model with several equations is represented as a matrix equation

$$AX + B = 0, \qquad (9)$$

where the interaction matrix A contains all the coefficients A_{ij} and the bias vector B all the bias terms B_i.

The model is represented by

$$x_{out} = f_{out}\left(-\frac{1}{A_{i\,out}} \Big(\sum_{j=1, j \neq out}^{m} A_{ij}\, f_j^{-1}(x_j) + B_i \Big) \right), \qquad (10)$$

ISBN: 978-989-95079-6-8

where the functions f_j and f_{out} are membership definitions. In the general case, the weight factors

$$w_{ij} = -\frac{A_{ij}}{A_{i\,out}}, \tag{11}$$

and the bias term

$$B_i = -\frac{B_i}{A_{i\,out}}, \tag{12}$$

Altogether, there are $n_i + 1$ parameters. As the scaling functions of each variable require three additional parameters for every two parameters needed for normalisation, the total number of additional parameters is $4 + 4n_i$ for n_i input variables.

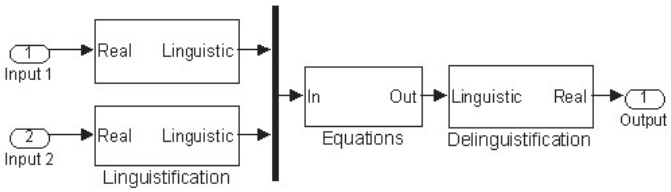


Figure 3: A LE model.

Nonlinear steady-state models can be constructed with linguistic equations, and then extended to dynamic systems using dynamic structures. Case-based systems can include both steady state and dynamic models.

2.2 Takagi-Sugeno type fuzzy systems

Takagi-Sugeno (TS) fuzzy model [4], in which the consequent is a crisp function of the antecedent variables, can be interpreted in terms of local models, see Fig. 4(a). A TS model with a common consequent structure can be understood as a global linear model with input-dependent parameters. For widely used linear functions, the standard weighted mean inference is

$$y = \frac{\sum_{i=1}^{K} \beta_i(\mathbf{x}) y_i}{\sum_{i=1}^{K} \beta_i(\mathbf{x})}, \tag{13}$$

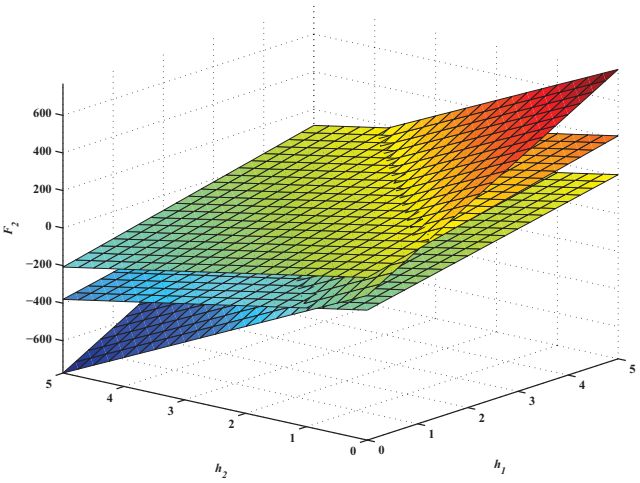
where the degree of fulfillment, $\beta_i(\mathbf{x})$, is the membership degree of the input vector $\mathbf{x}$ in the antecedent of the rule i, see Fig. 4(b). The models surface (Fig. 4(c)) is constructed from the values y_i which are calculated from the local models

$$y_i = \mathbf{a}_i^T \mathbf{x},\ i = 1, 2, \ldots, K. \tag{14}$$

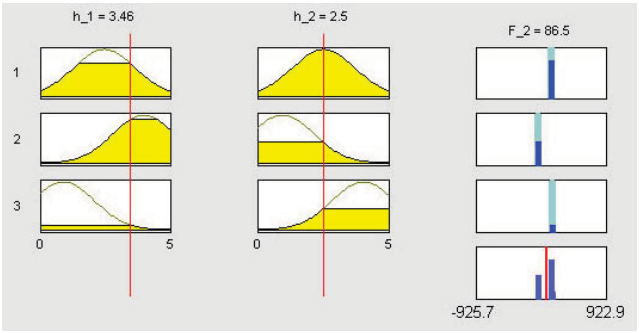
The inference method (13) introduces some undesirable properties in the border areas of the local models if the local models intersect within the border area. These drawbacks can be partly removed by using crisp transitions, $y = \max(y_1, y_2)$ and $y = -\max(-y_1, y_2)$ for a convex and a concave case, respectively. However, the slopes of the local models should differ drastically, and anyway the fuzzy inference is lost and the result is a piece-wise linear approximation. The inference (13) results are much better if the local models intersect outside the border area.

The smoothness of the model is directly dependent on the smoothness of the antecedent membership functions, e.g. the frequently used trapezoidal functions result in nonsmooth ouputs [3].

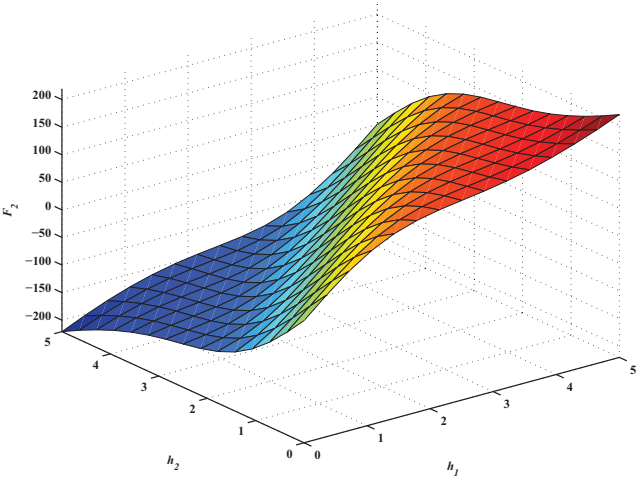
Takagi-Sugeno (TS) type fuzzy models are widely used for the identification of nonlinear systems, since the neuro-fuzzy ANFIS method [7] provides an efficient tuning method for these models. Fitting results are very good with strongly overlapping models, but process insight is lost as the individual models do not have any meaning. The ANFIS tuning increases the overlap of clusters and destroys the meanings of the individual linear models, e.g. the role of some submodels may transform into a part of a smoothing algorithm.



(a) Consequent surfaces.



(b) Fuzzy reasoning.



(c) Model surface.

Figure 4: A Takagi-Sugeno type fuzzy model.

The strong overlap may also result in a steep increase or decrease in the borders of the operating area as can be seen in Fig. 4(c). In dynamic simulation, these steep changes can be alleviated by limiting the range of the output variable, in

ISBN: 978-989-95079-6-8

addition, the operating area of rule 1 should be narrowed down from the solution provided by ANFIS tuning.

Steady-state TS models created with subtractive clustering [11] and ANFIS have proven to be very accurate in the fed-batch enzyme fermentation process [12], however dynamic simulation turned out to be too demanding [13]. Dimension reduction with clustering is necessary when the number of inputs is large. TS models based on grid partitioning require a large number of membership functions and rules which are difficult to tune in practice, e.g. models of the Kappa number in a continuous digester already resulted eight local models with three input variables when two membership functions were used for each variable [14].

The interpolation properties of the TS models can be improved by replacing the weighted mean by mechanisms which result in a piece-wise convex or concave interpolation surface, in which [3]: the gradients are bound by the gradients of the rule consequent functions, and the surface is smooth with continuous derivatives of a sufficient order. Then the model approximates the function more accurately than the surface generated by (13).

For example, the smoothing maximum function [15] connects two consequent hyperplanes, y_1 and y_2, by a smooth convex or concave surface, i.e.

$$y = y_1 + s_\gamma(y_2 - y_1), \tag{15}$$

where $s_\gamma(z)$ is a piece-wise polynomial

$$s_\gamma = \begin{cases} 0, & z \le -\gamma, \\ \frac{z+\gamma)^2}{4\gamma} & -\gamma < z < \gamma, \\ z, & z \ge \gamma \end{cases} \tag{16}$$

The interpolant starts to deviate from the consequent y_1 when $y_2 - y_1 = \gamma$. The smoothing parameter γ, which is a positive real number, is determined for each pair of adjacent rules.

Sharper borders require nonlinear consequent models, i.e. smoothing should be a part of each individual local model.

3 Multimodel LE system

A *multimodel approach* based on fuzzy LE models has been developed for combining specialised submodels [16]. The approach is aimed for systems that cannot be sufficiently described with a single set of membership definitions due to very strong nonlinearities. Additional properties can be achieved because equations and delays can also vary between different submodels. In the multimodel approach, the working area is defined by a separate working point model. The submodels are developed using the case-based modelling approach.

A multimodel system contains several submodels and a fuzzy decision system for selecting a suitable model for each situation using several working point variables. If several inputs are combined into a single working point index, the fuzzy set system is reduced to a fuzzification block (Fig. 5). Linguistic equation (LE) models have been used in several applications [17, 18, 19, 20, 21].

Linguistic Takagi-Sugeno fuzzy models (LTS) belong to this class of models, however with one limitation: the fuzzy partition is defined with same variables as the models. As *LE* models are nonlinear, the local models are also nonlinear.

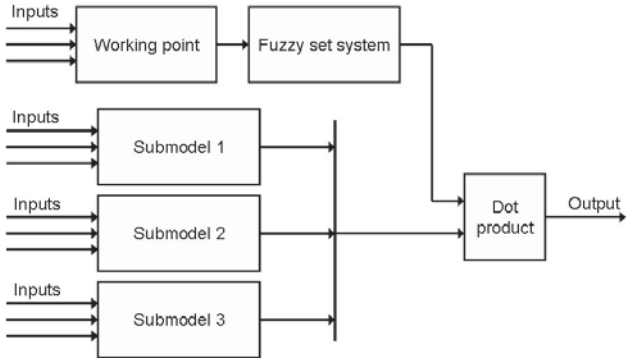


Figure 5: Multimodel LE system with fuzzy decision module.

LTS models can be developed and tuned with the same methods as the normal TS models. The only difference is that the variable values are scaled with the nonlinear scaling functions presented above.

The only difference to the normal LE model is that the equation part is handled with a fuzzy set system (Fig. 6). Nonlinear scaling is done with the same variable specific functions in each local model.

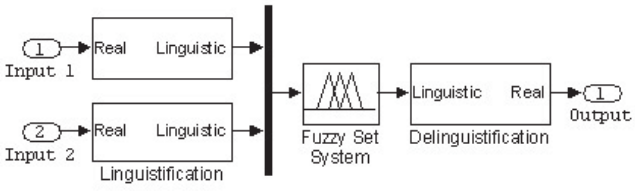


Figure 6: A fuzzy LE simulator.

4 Example: a tank system

The training environment consists of tank systems which include interactive parts where the inflow is divided between two tanks and the flow between the tanks depends on the levels of these tanks, h_1 and h_2, respectively (Fig. 7). The flow F_2 depends on the level of the tanks 1 and 2 (h_1 and h_2):

$$F_2 = c_1\sqrt{h_1 - h_2}, \tag{17}$$

$$F_3 = c_2\sqrt{h_2}. \tag{18}$$

The working point is set with the valve properties presented by coefficients c_1 and c_2.

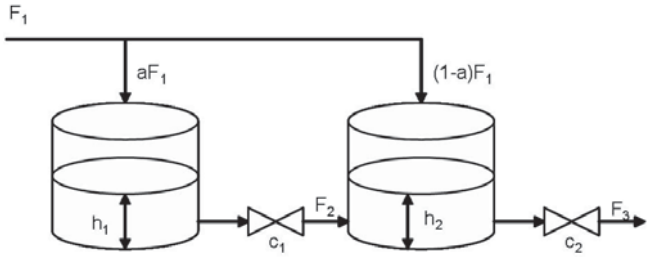
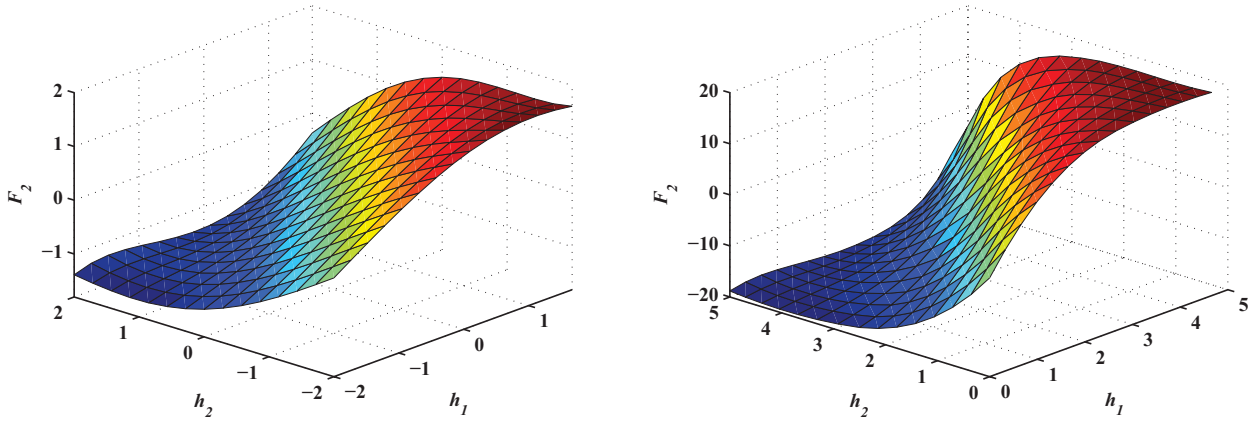


Figure 7: Interactive tank system.

ISBN: 978-989-95079-6-8



(a) The model in the linguistic range.

(b) The model in the real scale.

Figure 8: The model surface of the fuzzy LE model.

TS model. A good fitting can be achieved with a TS model

$$y_1 = 154.2h_1 - 153.9h_2 - 1.98, \quad (19)$$
$$y_2 = 57.61h_1 - 59.86h_2 - 80.08, \quad (20)$$
$$y_3 = 64.35h_1 - 60.18h_2 + 93.15. \quad (21)$$

This model shown in Fig. 4 was tuned with ANFIS resulting in a fairly extensive overlap between the submodels. Strong differences between the submodels can still be observed, but the detection of the models on the basis of their performance is not very clear. These problems were discussed in Section 2.2.

LTS model. A good fitting can be achieved with a LTS model

$$\widetilde{y_1} = 1.704\widetilde{h_1} + 1.785\widetilde{h_2}, \quad (22)$$
$$\widetilde{y_2} = 1.775\widetilde{h_1} + 3.144\widetilde{h_2}, \quad (23)$$
$$\widetilde{y_3} = 1.93\widetilde{h_1} + 5.013\widetilde{h_2}, \quad (24)$$

where the variables h_1 and h_2 are scaled with the nonlinear functions: $\widetilde{h_1} = f_1^{-1}(h_1)$ and $\widetilde{h_2} = f_2^{-1}(h_2)$. The output is first constructed in the linguistic range by using (13) and the submodels $\widetilde{y_1}$, $\widetilde{y_2}$ and $\widetilde{y_3}$. The result is then converted into a real value F_2 by (5) with the scaling parameters of the output variable: a_{out}^-, b_{out}^-, a_{out}^+, b_{out}^+ and c_{out}.

The operating areas are clearly observed, and the operation is smooth (Fig. 8(a)). The first consequent model is rather steep and operates when the levels of two tanks are close to each other (Fig. 8(b)). The second model corresponds to negative flow, i.e. $h_2 > h_1$, and the third one to positive flow, $h_1 > h_2$. No smoothing algorithms are needed, and there are no problems in the boundaries of the operating area (Fig. 8(b)). Performance evaluation also provides a clear indication of the active submodels.

5 Applications

LTS models provide good results if the same variables can be used for defining the operating areas and in the submodels. There are clear differences between the applications in this sense.

5.1 Single LE models

A single LE model proved to be a reliable solution for forecasting the Kappa number in continuous cooking with much less parameters than TS models [14]. Also, the steady-state LE model developed in an early lime kiln application provided better results than a TS model since the weaknesses of the data could be clearly observed and corrected in the LE approach [22].

A single LE model can capture nonlinear behaviour so well that the multimodel structure is not needed. The gas furnace data [23] was modelled using a single LE model, which provided very accurate results even in dynamic simulation [24]. The dynamic model of a flotation unit calculates the outlet turbidity from the properties of incoming water, chemical dosages, and previously calculated turbidity. This model is also used as an indicator of water quality [25]. The dynamic model of fluidised bed granulation consists of three interactive models: temperature, humidity, and granule size. All submodels are based on single LE models. [20]

5.2 Multimodels

Dynamic LE multimodels are used for the control design of two application: a lime kiln [10] and a solar collector field [26]. Smooth changes between the submodels were controlled by a fuzzy decision module as in Fig. 5. The lime kiln model is based on six operating areas defined by the production level and the trend of the fuel feed [16]. The interaction matrices are similar, and the differences between the submodels are introduced by scaling functions. The model has been used in a fuel quality indicator, which had an essential part in the development of a successful control system [10]. The model of a solar collector field consists of four specialised LE models, which are specified with interaction coefficients and scaling function parameters. The fuzzy decision module contains a working point model, which is based partly on different variables [24]. The simulator represents the field operation very accurately, even oscillatory conditions are handled correctly [26].

There are forecasting applications for batch cooking [19] and fed-batch fermentation [21]. In batch cooking, specific submodels are needed due to variations in the quality of the

ISBN: 978-989-95079-6-8

chips and the properties of the incoming cooking liquor. Multimodel aspects are taken into account by adapting the model to different operating conditions by selecting an appropriate speed factor depending on the H-factor and alkali level [19]. The dynamic model of fed-batch fermentation consists of three interactive models: carbon dioxide concentration, oxygen transfer rate, and dissolved oxygen concentration [21]. Transitions between the three phases, lag, exponential growth, and steady state, are defined on the basis of time, oxygen transfer rate, and glucose feed rate. The primary aim of the simulator is to detect fluctuations in the process control.

6 Conclusions

Special cases of fuzzy linguistic equation models can be understood as linguistic Takagi-Sugeno (LTS) type fuzzy models, in which nonlinear scaling is used for both inputs and outputs. The LTS models are robust solutions to applications where the same variables can be used for defining the operating areas and also in the submodels. No special smoothing algorithms are needed near the borders of the submodels.

References

[1] L. A. Zadeh. Knowledge Representation in Fuzzy Logic. *IEEE Transactions on Knowledge and Data Engineering*, 1(1):89–100, 1989.

[2] E. K. Juuso. Integration of intelligent systems in development of smart adaptive systems. *International Journal of Approximate Reasoning*, 35:307–337, 2004.

[3] R. Babuška. *Fuzzy Modeling and Identification*. Kluwer Academic Publisher, Boston, 1998.

[4] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, 15(1):116–132, 1985.

[5] W. Pedrycz. An identification algorithm in fuzzy relational systems. *Fuzzy Sets and Systems*, 13:153–167, 1984.

[6] R. Babuška and H. Verbruggen. Neuro-fuzzy methods for nonlinear system identification. *Annual Reviews in Control*, 27:73–85, 2003.

[7] J.-S. R. Jang. Anfis: Adaptive-network-based fuzzy inference systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(3):665–685, 1993.

[8] P. P. Angelov and D. P. Filev. An approach to online identification of Takagi-Sugeno fuzzy models. *IEEE Transactions on Systems, Man, & Cybernetics*, 34(1):484–498, 2004.

[9] T. A. Johansen and R. Babuška. Multiobjective identification of Takagi-Sugeno fuzzy models. *IEEE Transactions on Fuzzy Systems*, 11(6):847–860, 2003.

[10] E. K. Juuso. Fuzzy control in process industry: The linguistic equation approach. In H. B. Verbruggen, H.-J. Zimmermann, and R. Babuska, editors, *Fuzzy Algorithms for Control, International Series in Intelligent Technologies*, pp. 243–300. Kluwer, Boston, 1999.

[11] S. Chiu. Fuzzy Model Identification Based on Clustering Estimation. *Journal of Intelligent & Fuzzy Systems*, 2(3):Sept., 1994.

[12] U. Saarela, K. Leiviskä, E. Juuso, and A. Kosola. Modelling of a fed-batch enzyme fermentation process. In *IFAC International Conference on Intelligent Control Systems and Signal Processing. Faro, Portugal*. IFAC, 2003.

[13] E. K. Juuso. Dynamic simulation of a fed-batch enzyme fermentation process. In *Proceedings of SIMS 2005, 46th Conference on Simulation and Modeling, Trondheim, Norway*, pp. 117–124. Tapir Academic Press, Trondheim, 2005.

[14] K. Leiviskä, E. Juuso, and A. Isokangas. Intelligent modelling of continuous pulp cooking. In K. Leiviskä, editor, *Industrial Applications of Soft Computing*, pp. 147–158. Springer, Heidelberg, 2001.

[15] I. Zhang. A smoothing-out technique for min-max optimization. *Math. Prog.*, 19:61–77, 1980.

[16] E. K. Juuso. Intelligent dynamic simulation of a lime kiln with linguistic equations. In *ESM'99: Modelling and Simulation: A tool for the Next Millenium, 13th European Simulation Multiconference, Warsaw, Poland*, pp. 395–400, Delft, The Netherlands, 1999. SCS.

[17] E. K. Juuso. Modelling and simulation with intelligent methods. In *White paper of the Virtual Institute for Simulation (SimServ): www.sim-serv.com*. Sim-Serv, 2004. 17 pp.

[18] E. K. Juuso. Applications of smart adaptive system in pulp and paper industry. In *Proceedings of Eunite 2004 - European Symposium on Intelligent Technologies, Hybrid Systems and their implementation on Smart Adaptive Systems, Aachen, Germany*, pp. 21–33. Wissenschaftsverlag Mainz, Aachen, 2004.

[19] E. K. Juuso. Forecasting batch cooking results with intelligent dynamic simulation. In *Proceedings of the 6th EUROSIM Congress on Modelling and Simulation, Ljubljana, Slovenia*, volume 2, page 8 pp. University of Ljubljana, Ljubljana, Slovenia, 2007.

[20] E. K. Juuso. Intelligent modelling of a fluidised bed granulator used in production of pharmaceuticals. In *Conference Proceedings of SIMS 2007 - The 48th Scandinavian Conference on Simulation and Modeling, Göteborg (Särö)*, pp. 101–108. Linkping University Electronic Press, Linkping, Sweden, 2007.

[21] E. K. Juuso. Intelligent dynamic simulation of a fed-batch enzyme fermentation process. In *Tenth International Conference on Computer Modelling and Simulation, EUROSIM/UKSim, Cambridge, UK*, pp. 301–306. The Institute of Electrical and Electronics Engineers IEEE, 2008.

[22] E. K Juuso, T. Ahola, and K. Leiviskä. Fuzzy modelling of a rotary lime kiln using linguistic equations and neuro-fuzzy methods. In *Proceedings of the 3rd IFAC Symposium on Intelligent Components and Instruments for Control Applications - SICICA'97, Annecy, France*, pp. 579–584, 1997.

[23] G. E. P. Box and G. M. Jenkins. *Time Series Analysis, Forecasting and Control*. Holden Day, San Francisco, 1970.

[24] E. K. Juuso. Intelligent systems design with linguistic equations. In *9th Workshop Fuzzy Control des GMA-FA 5.22 am 4/5.11.1999, Dortmund, Deutschland. Forschungsberict Nr. 0449*, pp. 177–196, Dortmund, 1999. Universitt Dortmund, Facultt fur Electrotechnik.

[25] I. Joensuu, M. Piironen, and E. Juuso. Dynamic simulator for dosing of water treatment chemicals. In *Proceedings of European Symposium on Computer Aided Process Engineering-15 (Escape-15), Barcelona, Spain, Computer-aided chemical engineering, 20A*, pp. 301–306. Elsevier, Amsterdam, 2005.

[26] E. K. Juuso. Intelligent dynamic simulation of a solar collector field. In *Simulation in Industry, 15th European Simulation Symposium ESS 2003*, pp. 443–449. SCS, Gruner Druck, Erlangen, Germany, 2003.

ISBN: 978-989-95079-6-8

Knowledge Discovery in the Prediction of Bankruptcy

R. J. Almeida[1] S. Vieira[2] V. Milea[1] U. Kaymak[1] J. M. C. Sousa[2]

1.Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
2.Technical University of Lisbon, Instituto Superior Técnico, Dept. Mechanical Engineering/CIS-IDMEC
Av. Rovisco Pais 1, 1049-001 Lisbon
Email: {rjalmeida,kaymak,milea}@ese.eur.nl, {susana,jsousa}@dem.ist.utl.pt

Abstract— *Knowledge discovery in databases (KDD) is the process of discovering interesting knowledge from large amounts of data. However, real-world datasets have problems such as incompleteness, redundancy, inconsistency, noise, etc. All these problems affect the performance of data mining algorithms. Thus, preprocessing techniques are essential in allowing knowledge to be extracted from data. This work presents a real world application of knowledge discovery in databases, with the objective of prediction of bankruptcy. For this task fuzzy classification models based on fuzzy clustering are used, which are developed solely from numerical data. This data set has missing values, extreme values and also presents a much smaller bankruptcy class than the not bankruptcy class, which makes it a challenging problem in the scope of KDD.*

Keywords— Knowledge discovery in databases, feature selection, missing data, noisy data, prediction of bankruptcy, fuzzy classification.

1 Introduction

With the increase of economic globalization and evolution of information technology, financial data are being generated and accumulated at an exponential rate. It is used to keep track of companies, business performance, monitor market changes, and support financial decision-making. This rapidly growing volume of data triggered the need for automated approaches that allow effective and efficient utilization of massive financial data to support companies and individuals in strategic planning and investment decision-making.

Knowledge discovery can contribute to solving business problems in finance by finding patterns, causalities, and correlations in business information and market prices that are not immediately apparent to managers because the volume of data is too large or is generated too quickly to be screened by experts. Knowledge discovery has already been applied to a number of financial applications, including development of trading models, investment selection, loan assessment, portfolio optimization, fraud detection and bankruptcy prediction, amongst others. The prediction of bankruptcy has been been previously investigated in terms of the likelihood of success for the introduction of fuzzy systems for decision support [1]

The prediction of corporate failure or bankruptcy has been characterized as one of the most important problems facing business and government [2]. It also is a problem that affects the economy of every country. The number of failing firms is important for the economy of a country and it can be considered as an index of the development and robustness of the economy. The high individual, economic, and social costs encountered in corporate failures or bankruptcies make this problem very important to parties such as auditors, management, government policy makers, and investors [3].

There is a long history of research attempting to develop bankruptcy prediction models based on financial variables and other indicators of financial distress, using a wide variety of techniques. The pioneer in predicting business failure ratios is considered to be [4]. The predictive accuracy of the initial approaches has varied from around 65% [5] to around 90% [6]. Higher predictive accuracy is often achieved by using samples concentrated in a few industries, using samples with widely varying bankruptcy/non-bankruptcy company sizes or making inappropriate assumptions about real world bankruptcy/non-bankruptcy frequencies.

The data set used in this work, concerning the bankruptcy, is different from all of the above mentioned, so no direct comparison of results can be made. Also, other works never mentioned that data sets had missing values and extreme values, as in this work.

In this work we present a full KDD process, applied in a real-world problem: the prediction of bankruptcy. The data set used is quite challenging as it has missing values and extreme values. It also presents a much smaller bankruptcy class than the not bankruptcy class. Several possibilities were tested in each step of the KDD process, such as data preparation, feature selection and fuzzy classification, and we discuss them briefly although we give more focus on the best results obtained. Note that the use of fuzzy systems for classification, besides building a numeric prediction model also represents the model behaviour in terms of linguistic rules, making it possible to interpret, which is an important final step in the KDD process.

The outline of the paper is as follows. In Section 2 we briefly present the KDD steps used to obtain a bankruptcy classifier. In Section 3 we present techniques used for data preparation. Feature selection is presented in Section 4 and Section 5 describes the procedures used to derive Takagi-Sugeno fuzzy models by means of fuzzy clustering. The data used in the work and the results are presented in Section 6. Finally, conclusions and future work are given in Section 7.

2 KDD process

The search for knowledge in large data sets, with the use of different hypothesis spaces, is the central and necessary phase within the discovery process. A large number of methods have been developed that handle many search tasks, but hypotheses

ISBN: 978-989-95079-6-8

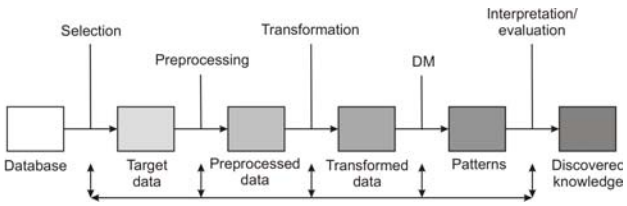


Figure 1: Phases in the KDD process, adapted from [8]

inference and verification is only a part of the whole process of knowledge discovery. As any other process, it has its environment, its phases, and runs under certain assumptions and constraints. The process undertakes many phases, namely [7]:

1. Definition and analysis of the problem;
2. Understanding and preparation of data;
3. Setup of the search for knowledge;
4. The actual search for knowledge;
5. Interpreting mined patterns;
6. Deployment and practical evaluation of the solutions.

The KDD process, activities and phases, is shown in Figure 1.

Compared to the traditional manual analysis, KDD provides a much higher degree of system autonomy, especially in processing large hypotheses spaces. However, at the current state of the art, a human analyst still makes many decisions in the course of a discovery process.

The KDD process starts from specification of a given problem and data understanding, and ends with actionable conclusions from the discovered knowledge. The output of DM is, in general, a set of patterns, some of which possibly represent discovered knowledge. In the next sections we will briefly explain each step of the KDD process.

3 Data Preparation

Real-world databases are highly susceptible to noisy, missing, and inconsistent data due to their typically huge size. Thus, data preparation consumes most of the time needed to mine data [9], and can substantially improve the overall quality of the patterns mined and/or the time required for the actual mining [10].

This section is a brief overview of some of the concepts regarding data preparation to yield the best possible model. For more details refer to [9, 10].

3.1 Missing Data

When applying data analysis methods to real problems, we often find that the data sets contain many missing elements. There are two forms of randomly missing data: missing completely at random (MCAR) and missing at random (MAR) [11]. Missing values MCAR, behave like a random sample and their probability does not depend on the observed data or the unobserved data [12, 13]. MAR exists when missing values are not randomly distributed across all observations but are randomly distributed within one or more subsamples.

A possible approach to deal with the missing values is to discard all incomplete data, and then execute the data analysis method on the remaining data. However, if missing values are frequent, the data set size may be considerably reduced, yielding unreliable or distorted results. A way to minimize this extreme data reduction problem is presented in [14]. If values are missing completely at random they can be imputed. Widely used imputation methods use the variables mean, median or the most probable value as a replacement [11, 10].

3.2 Noisy Data

Noise is considered to be a random error or variance in a measured variable [10]. A specific type of noise, which the data in this study has, is extreme values, which in some cases can be labelled as outliers. In this work numerically sensitive mining algorithms are used, and it is recommended to normalize the individual variable distributions [15].

Outliers are defined as a large deviation from the mean value of the rest of the data. Usually outliers can be thrown out of the data set, as they bias the analytical results. Distribution normalization deals with the problems presented by valid outliers. An outlier is valid if it represents an accurate measurement and still falls well outside the range of the majority of values. These should not be discarded.

The distribution normalization can be achieved using transformations. In cases where the data have strong asymmetry, many outliers or batches at different levels with widely differing spreads a power transformation may alleviate this problem, without violating the necessary transformation properties [16].

4 Feature Selection

One of the great challenges in classification is selecting the important input variables from all possible input variables. Classification problems involve a large number of potential inputs. The number of inputs actually used by the model must be reduced to the necessary minimum, especially when dealing with fuzzy models that are, presumably, nonlinear and contain many parameters. Therefore, it is necessary to select carefully the variables that are relevant for the feature class.

Feature selection is a process that chooses a subset of M features from the original set of n features ($M \leq n$), so that the feature space is optimally reduced according to a certain criteria.

Even when a good criteria exists for model selection, there is no guarantee that a model based on a given set of variables is optimal unless all possible combinations of variables have been explored. The problem is known to be NP-hard [17]. Hence, finding an optimal solution requires building a model for each possible combination of input variables, which becomes computationally prohibitive for problems involving even a moderate number of candidate input variables.

Feature selection algorithms, essentially divide into wrappers and filters [18]. In this work we use wrappers that make use of specific learning algorithms to evaluate variables in the context of the learning problem, rather than independently. Wrappers share strengths and weaknesses of the learning algorithms and have the advantage of using the actual hypothesis accuracy as a measure of subset quality. Furthermore, wrapper methods do tend to outperform filter methods [18].

ISBN: 978-989-95079-6-8

The variable selection procedure can be used with various performance criteria for model selection. In real-world databases sometimes one of the classes is more difficult to classify then the others. This can happen, for instance, when one of the classes is much bigger than the other or the interest of the problem is a specific class. To cope with this problem, we use a criterion, that assigns specific weights to each class in the model evaluation of the feature selection algorithm [19].

In this work we compare the well known sequential forward selection (SFS) and the sequential backward elimination (SBE) against the newly proposed ant feature selection algorithm (AFS). The SFS and SBE search algorithms may not be the best search methods, nor guarantee an optimal solution, but they are popular because they are simple, fast, provide a very reasonable solution and are much more efficient than exhaustive search.

Sequential forward selection and sequential backward elimination were first used in the context of feature selection for pattern classification. SFS was first used in [20] and was later used in [21]. It has also been used to determine input variables for fuzzy models in [22] and [23]. SBE was first used in [20] and in [24] it was used with simple linear models to provide a first-round elimination of the input variables for a fuzzy model.

4.1 Ant Feature Selection

Ant algorithms were first proposed by Dorigo [25] as a multi-agent approach to difficult combinatorial optimization problems, such as traveling salesman problem, quadratic assignment problem or supply chain management [26]. Here we present an implementation of ACO applied to feature selection, where the best number of features is determined automatically.

In this approach, two objectives are considered: minimizing the number of features and minimizing the error classification. Two cooperative ant colonies optimize each objective. The first colony determines the number (cardinality) of features and the second selects the features based on the cardinality given by the first colony. Thus, two pheromone matrices and two different heuristics are used. The heuristic value is computed using the Fisher discriminant criterion for feature selection [27], which ranks the features giving them a given relative importance.

The best number of features is called *features cardinality* N_f. The determination of the *features cardinality* is addressed in the first colony sharing the same minimization cost function J^τ with the second colony, which in this case aggregates both the maximization of the classification accuracy and the minimization of the features cardinality. Hence, the first colony determines the size of the subsets of the ants in the second colony, and the second colony selects the features that will be part of the subsets.

The objective function of this optimization algorithm aggregate both criteria, the minimization of the classification error rate and the minimization of the features cardinality:

$$J^\tau = w_1 \frac{N_e^\tau}{N_n} + w_2 \frac{N_f^\tau}{n} \tag{1}$$

where $\tau = 1, \ldots, g$, g being the number of ants, w_1 and w_2 are weights, N_n is the number of used data samples, n is the total number of features, N_e is the number of errors produced by the solution and N_f is the features cardinality.

To evaluate the classification error, a fuzzy classifier is built for each solution following the procedure described in Section 5. This approach was presented in [28].

5 Fuzzy Classification

In this section we outline the basics of the adopted fuzzy reasoning scheme for pattern classification problems. Let us consider a n-dimensional classification problem for which N patterns $\vec{x}_p = (x_p^1, \ldots, x_p^n)$, $p = 1, 2, \ldots, N$ are given from κ classes $C_1, C_2, \ldots, C_\kappa$. The task of a pattern classifier is to assign a given pattern $\vec{x}$ to one of the κ possible classes based on its features values. Thus, a classification task can be represented as a mapping $\psi : X \subset \Re^n \rightarrow \{0,1\}^\kappa$ where $\psi(\vec{x}) = \vec{c} = (c_1, ..., c_\kappa)$ such that $c_k = 1$ and $c_j = 0 (j = 1, \ldots, \kappa, j \neq k)$.

Assuming that there is an arbitrary ordering of the classes, one way to solve this classification problem is to consider classifiers with a continuous output, e.g., a Takagi-Sugeno affine system [29]. The output of an affine Takagi-Sugeno fuzzy rule is

$$y_k = \vec{a}\vec{x} + b_k \tag{2}$$

where y_k is the output for rule k, $\vec{a}_k$ is a parameter vector and b_k is a scalar offset. In classification problems the output should be a discrete value corresponding to one of the classes to be identified. So a threshold T_l can be used on the output y_k, to decide which class that it belongs to, as $x_k \in C_l$ if $y_k \in T_l$.

To form the fuzzy system model from the data set with N data samples, given by $X = [\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N]^T$, $Y = [y1, y2, \ldots, yN]^T$ where each data sample has a dimension of n ($N >> n$), the structure is first determined and afterwards the parameters of the structure are identified. The number of rules characterizes the structure of a fuzzy system. Fuzzy clustering in the Cartesian product-space $X \times Y$ is applied to partition the training data. The partitions correspond to the characteristic regions where the systems behavior is approximated by local linear models in the multidimensional space. Given the training data X_T and the number of clusters K, a suitable clustering algorithm is applied.

The fuzzy clustering algorithms used in this work are based on the optimization of an objective function. In particular we use the fuzzy c-means (FCM) [15], the Gustafson-Kessel (GK) [30], the possibilistic c-means (PCM) [31], fuzzy possibilistic c-means (FPCM) [32] and the recent possibilistic fuzzy c-means (PFCM) [33].

The FCM functional uses a probabilistic constraint which states that the sum of membership degrees must equal one [15]. Problems arise in situations, where the total membership of a data point to all the clusters does not equal one, as in the presence of outliers. Clearly, one would like the memberships for representative feature points to be as high as possible, while unrepresentative points should have low membership in all clusters. The PCM objective function relaxes this constraint [31].

Gustafson-Kessel extended the standard fuzzy c-means algorithm by employing an adaptive distance norm, in order to detect clusters of different geometrical shapes in one data set [30]. As the GK algorithm is based on an adaptive distance

ISBN: 978-989-95079-6-8

measure, it is less sensitive to scaling (normalization, standardization) of the data.

Fuzzy-possibilistic c-means, simultaneously produces both memberships and possibilities. FPCM tries to solve the noise sensitivity defect of FCM, and also overcomes the coincident clusters problem of PCM. Note that FCM and FPCM will not generate the same membership values, even if both algorithms are started with the same initialization [32].

FPCM imposes a constraint on the typicality values. PFCM relaxes this constraint but retains the column constraint on the membership values. The PFCM functional has two constants that define the relative importance of fuzzy membership and typicality values.

Memberships and typicalities are both considered important for correct interpretation of data substructure. When the objective is to classify a data point, membership may be a better choice as it is natural to assign a point to that cluster whose representative vector is closest to the data point. On the other hand, when seeking the clusters, i.e., while estimating the centroids, typicality is an important means for alleviating the undesirable effects of outliers.

Depending on the clustering algorithm used in this work, a fuzzy partition matrix $U = [\mu_{ik}]$ and/or the typicality matrix $T = [t_{ik}]$ will be obtained. The fuzzy sets in the antecedent of the rules are identified by means of the matrix U and T which have dimensions $[N \times K]$. One dimensional fuzzy sets A_{ij} are obtained from the multidimensional fuzzy sets by projections onto the space of the input variables x_j. This is expressed by the point-wise projection operator of the form $\mu_{A_{ij}}(x_{jk}) = \text{proj}_j(\mu_{ik})$ The antecedent membership functions can now be obtained from the fuzzy partition matrix or from the typicality matrix. The point-wise defined fuzzy sets A_{ij} are then approximated by appropriate parametric functions. The consequent parameters for each rule are obtained by means of linear least square estimation, which concludes the identification of the classification system.

6 KDD Applied to Prediction of Bankruptcy

The data set used in this work contains data from 1817 companies, each one described by 52 features (including the class feature), containing financial, behavioral and qualitative features (as perceived by the account manager).

Some of the features contain extreme values (EV) as their value is 10000 larger then the other values. In this case, the extreme values are believed to be different values for different type and size of companies, containing valuable information.

The data set contains about 10% of missing values. Only 18.38% (334) of the companies have all the features complete and there are only 5 features without missing values, one of those features being the status of the company.

Each company in the data set has two possible status (classes): status 0 (bankrupt) and status 1 (not bankrupt). The distribution of the classes is uneven: only 4.3% (78) of the companies have the status 0, and the remaining 95.7% (1739) have a status 1. This distribution skewness is common in bankruptcy data.

The KDD steps taken in this work to obtain a compact fuzzy classification model are the following:

1. Manual selection of relevant data from the available data.
2. Preprocessing of the data to deal with extreme values, using a power transformation followed by a linear transformation.
3. Preprocessing of the data to deal with missing values, by replacement of missing values using the most probable value.
4. Searching and selecting the relevant features using search algorithms.
5. Obtaining a fuzzy classification model by using only the selected features in the previous step.

The following computational protocols were used: $\varepsilon = 0.0001$, maximum number of iterations 100, fuzzy exponent $m = 2$ and $\eta = 2$, and the Euclidean norm is used. For both PCM and PFCM we first run FCM to termination. All trials terminated with the convergence criteria after a few iterations.

6.1 Data Preprocessing

When analysing the data we found that seven features can be discarded because they contain company descriptive features that may not be relevant to this research.

Extreme Values Linear transformations of re-expressed data present little additional difficulty in interpretation. Since power transformations are monotonic for positive data values, we transform the data so that it has only positive values, maintaining the missing values, using the linear transformation,

$$z_{jk} = z^*_{jk} + |\min z_j| \tag{3}$$

where the asterisk denotes the unscaled data. After this linear transformation, which does not alter the shape of the data, we apply a power transformation $T(x) = \log(x)$, because this transformation alters the distribution of the data and bring the extreme values to a value closer to the other values [16]. We chose a matched transformation, presented as:

$$z^1 = a \; + \; bT(x), \tag{4}$$

where a and b are chosen, by using a point x_0 and require that:

$$z_0 = a \; + \; bT(x) = x_0, \tag{5}$$

and, furthermore, that the derivative of z with respect to x, evaluated at x_0, to be 1. That is,

$$\left.\frac{dz}{dz}\right|_{x_0} = \frac{d[a + bT(x)]}{dx} = b \left.\frac{dT(x)}{dx}\right|_{x_0} = 1. \tag{6}$$

This method relies on the linearity of the transformation near the center of the range x. For sake of simplicity, we chose x_0 as the median value of each feature. Other point could have been chosen. Furthermore, we normalized the obtained values of the matched transformation so that all the features are contained between the interval $[0, 1]$, obtaining the data set z^2.

Missing Values Almost every company has missing values for their features, and so discarding the companies with missing values cannot be considered a feasible approach for this research. Filling in the missing value during the data preprocess, was used. After the data transformation, we imputed the missing values using probable value, calculated using the Expectation-Maximization (EM) algorithm [34].

ISBN: 978-989-95079-6-8

6.2 Fuzzy Classification

One of the most important advantages of the, rather complex, transformation described, can be seen in Table 1. This table shows the obtained accuracy obtained with the data before the transformation, using 6 clusters, and the obtained results after the transformation with only 3 clusters. Since the class 0 is so small, we used 4-fold cross validation. As can be seen the results obtained with the data before transformation are more disperse. This does not happen with the transformed data. The GK algorithm could not be used in the original data because of the extreme values. For this reason we decided not to included it in this comparison. Furthermore the converge of the clustering algorithms is improved and less number of clusters are needed to obtain good results.

Table 1: Accuracy obtained for bankruptcy data with 6 clusters (Raw data) and 3 clusters (Processed data).

	Raw data			Processed data		
Alg.Cl.	Max.	Mean	Min.	Max.	Mean	Min.
FCM	0.928	0.885	0.800	0.972	0.946	0.919
PCM	0.925	0.894	0.733	0.969	0.945	0.919
FPCM-U	0.924	0.877	0.800	0.953	0.944	0.939
PFCM-U	0.926	0.894	0.733	0.955	0.941	0.927
FPCM-T	0.925	0.891	0.800	0.939	0.939	0.939
PFCM-T	0.925	0.891	0.800	0.939	0.925	0.919

6.3 Feature Selection

After the fuzzy classification models with all the features were extracted, we look to further simplify the model, using feature selection. We applied each algorithm of feature selection (SFS,SBE and AFS), using the weighted accuracy for each feature class present [19], and all of the clustering algorithms in study (GK,FCM, FFCA, PCM, FPCM and PFCM), to both the data normalized and not normalized.

In this case, if no weights were used, the algorithms would choose the features that would maximize the class no bankruptcy as this is the dominant class. Usually after one or two iterations the process would stop. In this case, without the use of the weights the results are highly biased towards the bigger class.

The features which were selected throughout our tests varied between 2 with the sequential forward search and 37 with the sequential backward search. This result was expected as different types of features selection algorithms choose features in a different manner, and we tested a number of different clustering algorithms. Good results were obtained using the ants feature selection. The number of features chosen was 15, 3 rules were derived and the obtained accuracy was 78.9% for companies that are bankrupt and 94.9% for companies that are not bankrupt. The fuzzy clustering algorithm used was PCM.

The use of fuzzy systems for classification, besides building a numeric prediction model also represents the model behaviour in terms of linguistic rules, which is very natural for human to understand. Interpretability is considered to be the main advantage of fuzzy systems over other non-fuzzy alternatives like statistical models or neural networks.

The simplest classification model obtained with the data normalized, only has 2 features and 3 rules and was derived using the PCM algorithm. The obtained accuracy is 57.9% for the companies that are bankrupt, whereas the percentage of companies that are not bankrupt is 97.7%. If we compare the possibility of interpretability of this model against the model derived in [3], that had 70 rules and 35 features then it can be considered that these results are good. Also bare in mind that this data set has missing values, extreme values and one class which is much smaller than the other.

The rule base model is relatively simple, as only 3 rules and 2 features are used, the understanding of the consequents of the model is not a simple task. The obtained rules, for this simple model are:

1. **If** $PROFIAT$ is Low **and** $CFEQ$ is Low **then** $y_1 = 0.35PROFIAT + 0.16CFEQ - 0.16$
2. **If** $PROFIAT$ is $Medium$ **and** $CFEQ$ is $Medium$ **then** $y_2 = -24.52PROFIAT - 18.65CFEQ + 39.12$
3. **If** $PROFIAT$ is $High$ **and** $CFEQ$ is $High$ **then** $y_3 = 5.54PROFIAT + 8.49CFEQ - 11.85$

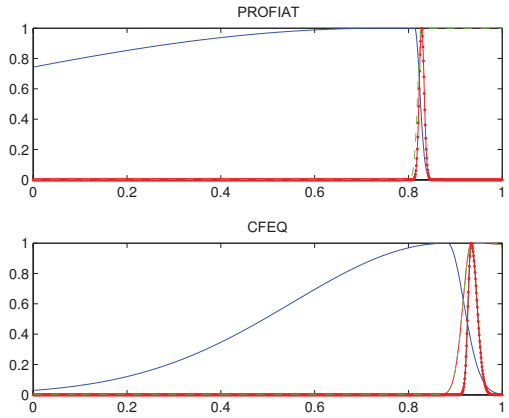


Figure 2: Membership Functions bankruptcy Model. solid - Low, dotted - Medium, dashed - High.

The obtained membership functions are shown in Fig. 2. The features we have used are the profit after tax (PROFIAT), and cash flow to equity (CFEQ). Note that the former is the company's net operating after tax profit for investors, while the latter is the cash that can be paid to the equity shareholders after the company expenses. It is interesting to note that the membership functions are skewed towards the higher values, which indicates that the difference between company defaulting or not is quite small. According to our simple model, a small value on either of these features is an indication that the company is performing badly and it is likely to default. It is interesting to note that the *medium* membership function are located in high values. Therefore, even companies with high values of both features can still default.

7 Conclusions

In this work we present a real world application of knowledge discovery for prediction of bankruptcy using a databases that has noisy, missing, and inconsistent data. We present for each step of the KDD process several possibilities that we tested in order to obtain good fuzzy classification models. Before fuzzy models could be extracted from the data, it is necessary to represent the real-world objects of interest in the data

ISBN: 978-989-95079-6-8

in a way that this specific method can access the data. With data preparation, although time consuming, it was possible to derive compact fuzzy models with only a few features that predicted the bankruptcy with an high accuracy rate.

Acknowledgments

This work is partially supported by the Portuguese Government and FEDER under the programs: Programa de financiamento Plurianual das Unidades de I&D da FCT (POCTI-SFA-10-46-IDMEC) and by the FCT grant SFRH/25381/2005, Fundação para a Ciência e a Tecnologia, Ministério do Ensino Superior, da Ciência e da Tecnologia, Portugal.

References

[1] M. Setnes, U. Kaymak, and H.R. van Nauta Lemke. Fuzzy target selection in direct marketing. *IEEE Transactions on Fuzzy Systems*, 7(1):368–369, 1999.

[2] D.E. O'Leary. Using neural networks to predict corporate failure. *International Journal of Intelligent Systems in Accounting, Finance and Management*, 7(3):187–197, 1998.

[3] J. Stefanowski and S. Wilk. Evaluating business credit risk by means of approach-integrating decision rules and case-based learning. *International Journal of Intelligent Systems in Accounting, Finance and Management*, 10(2):97–114, June 2001.

[4] E. I. Altman. Financial ratios, discriminate analysis and the prediction of corporate bankruptcy. *Journal of Finance*, 23(4):589–609, 1968.

[5] D.H. Lindsay and A. Campbell. A chaos approach to bankruptcy prediction. *Journal of Applied Business Research*, 12(4):1–9, 1996.

[6] T.B. Bell, G.S. Ribar, and J. Verchio. Neural nets versus logistic regression: a comparison of each models ability to predict commercial bank failures. In *Proceedings of the 1990 Deloitte and Touche University of Kansas Symposium on Auditing Problems*, pages 29–54, 1990.

[7] G. Piatetsky-Shapiro U. Fayyad and P. Smyth. The kdd process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11):27–34, 1996.

[8] P. Smyth U.M. Fayyad, G. Piatetsky-Shapiro and R. Uthurusamy. *Advances in Knowledge Discovery and Data Mining*. MIT Press, Cambridge, MA, USA, 1996.

[9] D. Pyle. *Data Preparation for Data Mining*. Morgan Kaufmann, San Francisco, CA, 1999.

[10] J. Han and M. Kamber. *Data Mining Concepts and Techniques*. Morgan Kaufmann Publishers Inc, San Francisco, CA, 2000.

[11] H. Timm, C. Döring, and R. Kruse. Differentiated treatment of missing values in fuzzy clustering. In Taner Bilgiç, Bernard De Baets, and Okyay Kaynak, editors, *Proceedings of the 10th International Fuzzy Systems Association World Congress (IFSA 2003), Lecture Notes in Artifical Intelligence*, volume 2715.

[12] R. J. A. Little and D. B. Rubin. *Statistical Analysis with Missing Data, Second Edition*. Wiley and Sons, New York, 2002.

[13] J. L. Schafer. *Analysis of Incomplete Multivariate Data*. Chapman & Hall, London, 1997.

[14] M. Setnes and U. Kaymak. Fuzzy modeling of a client preference from large data sets: an application to target selection in direct marketing. *IEEE Transactions on Fuzzy Systems*, 9(1):153–163, 2001.

[15] J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York, 1981.

[16] M.A. Stoto J.D. Emerson. Transforming data. In F. Mosteller D.C. Hoaglin and J.W. Tukey, editors, *Understanding Robust and Exploratory Data Analysis*, chapter 4. Wiley-Interscience, New York, London, Sydney, Toronto, 2000.

[17] E. Amaldi and V. Kann. On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems. *Theoretical Computer Science*, 209(1–2):237–260, 1998.

[18] R. Kohavi and G.H. John. Wrappers for feature subset selection. *Artificial Intelligence*, 97(1-2):273–324, 1997.

[19] C. A. Silva R.J. Almeida and J.M.C. Sousa. A new feature selection criterion for fuzzy classification. In *Proc. of the 2006 IEEE International Conference on Fuzzy Systems, WCCI 06*, pages Session Classification, pp. 1513–1520, Vancouver, Canada, July 2006.

[20] T. Marill and D.M. Green. On the effectiveness of receptors in recognition systems. *IEEE Transactions on Information Theory*, 9(1):11–17, 1963.

[21] A.W. Whitney. A direct method of nonparametric measurement selection. *IEEE Transactions on Computers*, 20(9):1100–1103, 1971.

[22] M. Sugeno and G.T. Kang. Structure identification of fuzzy model. *Fuzzy Sets and Systems*, 28(1):15–33, 1988.

[23] M. Sugeno and T. Yasukawa. A fuzzy-logic-based approach to qualitative modeling. *IEEE Transactions on Fuzzy Systems*, 1(1):7–31, February 1993.

[24] H. Watanabe K. Tanaka, M. Sano. Modeling and control of carbon monoxide concentration using a neurofuzzy technique. *IEEE Transactions on Fuzzy Systems*, 3(3):271–279, 1992.

[25] Marco Dorigo, M. Birattari, and T. Stützle. Ant colony optimization. *IEEE Computational Intelligence Magazine*, 1(4):28–39, 2006.

[26] C. A. Silva, J. M. C. Sousa, T. A. Runkler, and J. M. G. Sá da Costa. Distributed optimization of a logistic system and its suppliers using ant colony optimization. *International Journal of Systems Science*, 37(8):503–512, 2006.

[27] R. O. Duda, P. E. Hart, and D. G Stork. *Pattern Classification, Second Edition*. Wiley–Interscience Publication, 2001.

[28] S. M. Vieira, J. M. C. Sousa, and T. A. Runkler. Fuzzy classification in ant feature selection. In *Proc. of 2008 IEEE World Congress on Computational Intelligence, WCCI 2008*, pages 1763–1769, Hong Kong, China, June 2008.

[29] T. Takagi and M. Sugeno. Fuzzy identification of systems and its aplication to modeling and control. *IEEE Transactions on Systems, Man and Cybernetics*, 15(1):116–132, 1985.

[30] D. Gustfson and W. Kessel. Fuzzy clustering with a fuzzy covariance matrix. In *Proceedings of IEEE Conference on Decision and Control, CDC*, pages 761–766, San Diego, CA, USA, 1979.

[31] R. Krishnapuram and J.M. Keller. A possibilistic approach to clustering. *IEEE Transactions on Fuzzy Systems*, 1(2):98–110, May 1993.

[32] N.R Pal, K. Pal, and J.C Bezdek. A mixed c-means clustering model. In *Proceedings of the Sixth IEEE International Conference on Fuzzy Systems*, volume 1, pages 11–21, Barcelona, Catalonia, July 1997.

[33] J.M. Keller N.R. Pal, K. Pal and J.C.Bezdek. A possibilistic fuzzy c-means clustering algorithm. *IEEE Transactions of Fuzzy Systems*, 13(4):517–530, August 2005.

[34] T. Schneider. Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values. *Journal of Climate*, 14:853–871, October 2001.

Generalizations of Fuzzy C-Means Algorithm to Granular Feature Spaces, based on Underlying Metrics: Issues and Related Works

Vasile Georgescu

Faculty of Economics and Business Administration, University of Craiova
Craiova, Romania
Email: vgeo@central.ucv.ro, v_geo@yahoo.com

Abstract—*This paper considers dissimilarity measures and clustering techniques for two special cases of set-defined objects: fuzzy granules and subsequence time series. To deal with clustering of such kind of objects, we propose two implementations that generalize the Fuzzy C-Means algorithm to granular feature spaces. Granular computing is a paradigm oriented towards capturing and processing meaningful pieces of information, the so-called information granules. In a granular feature space, such as a space populated with p-dimensional fuzzy granules, we are concerning with both granular data samples and granular centroids (center of clusters). In order to accommodate clustering algorithms to work in a granular environment we have to choose and/or define appropriate metrics and descriptors. Either a crisp distance between granules or the defuzzified value of a fuzzy distance has to be chosen. On the other hand, subsequence time series clustering requires a generalization of Fuzzy C-Means algorithm in a similar way. It involves a set-defined centroid and appropriate dissimilarity measures to determine the degree to which time sequences are different from their centroid. Furthermore, we discuss related work in granular clustering and subsequence time series clustering.*

Keywords—Granular clustering, Metrics in granular feature spaces, Fuzzy C-Means clustering, Agglomerative granular clustering algorithms, Subsequence time series clustering.

1 *P*-dimensional fuzzy granules: representation and cardinality measure

In what follows, the attention will be restricted to the class Φ^p of normal fuzzy convex granules on $\Re^p$, whose α-level sets are nonempty compact convex sets for all $\alpha > 0$. Each fuzzy granule $M \in \Phi^p$ is uniquely characterized by its support function $s_M(u,\alpha) = \sup\{ < u, x > | \, x \in M^\alpha \}$, $u \in S^{p-1}, \alpha \in (0,1]$, where S^{p-1} is the (p-1)-dimensional unit sphere of $\Re^p$ (i.e. $\|u\| = 1$) and $< \cdot, \cdot >$ is the inner product in $\Re^p$. The support function is a mapping from the class of fuzzy sets Φ^p into the space of functions $L\left(S^{p-1} \times [0,1]\right)$, which preserves addition and multiplication with non-negative scalars.

A p-dimensional fuzzy granule $A = A^1 \times \cdots \times A^p$ is defined on the product space $X = X_1 \times \cdots \times X_p$. Representation and cardinality measure for A can be given either in terms of membership functions:

$$\mu_A : \Re^p \to [0,1],$$
$$\mu_A(x) = \min\left(\mu_{A^1}(x_1), \ldots, \mu_{A^p}(x_p) \right), \quad \forall\, x \in X$$

$$|A| = \int_{\xi \in Supp(A)} \mu_A(x)\, dx =$$
$$= \int_{x_1 \in Supp(A^1)} \cdots \int_{x_p \in Supp(A^p)} \min\left(\mu_{A^1}(x_1), \ldots, \mu_{A^p}(x_p)\right) dx_1 \ldots dx_p$$

or in terms of α – level intervals:

$$A(\alpha) = A^1(\alpha) \times \cdots \times A^p(\alpha); \qquad \alpha \in [0,1]$$

$$|A| = \int_0^1 |A(\alpha)|\, d\alpha = \int_0^1 \prod_{j=1}^{p} \left|A^j(\alpha)\right| \, d\alpha.$$

In particular, let A^j be an LR-fuzzy set. We have:

$$A^j(\alpha) = \left[A^{jL}(\alpha),\, A^{jR}(\alpha)\right]_{LR}$$

$$|A| = \int_0^1 \prod_{j=1}^{p} \left| A^{jR}(\alpha) - A^{jL}(\alpha) \right| d\alpha.$$

The latest integral can be evaluated numerically by means of a quadrature formula (e.g. the adaptive Simpson quadrature).

2 Metrics in granular feature spaces

2.1 A few examples

We focus on distances in granular feature spaces populated with fuzzy granules. There is an enormous literature that covers this topic (at least in case of one-dimensional fuzzy spaces). There is also a large diversity of approaches, among which we can distinguish:

a) Membership focused distances (vertical)
b) Spatially focused distances (horizontal)
c) Mix of spatial and membership distances (tolerance)
d) Feature distances (low or high dimensional representations)
e) Morphological (mixed focus)

Most of these metrics are defined as crisp distances. As a typical example, we can consider an L_2-metric on the space of normal compact convex fuzzy sets Φ^p using the L_2-metric on the Hilbert space of square-integrable functions

ISBN: 978-989-95079-6-8

$L_2\left(S^{p-1}\times[0,1]\right)$. This space is equipped with the inner product

$$< M, N >_\lambda = p\cdot \int_{[0,1]} \int_{S^{p-1}} s_M(u,\alpha)\ s_N(u,\alpha)\ \mu(du)\ \lambda(d\alpha).$$

where μ is normalized Lebesgue measure on S^{p-1} (i.e. $\mu(S^{p-1})=1$) and λ is normalized Lebesgue measure on $[0,1]$. Now, the corresponding L_2-metric results of the following form:

$$\delta_2^{(\lambda)}(M,N)=\|s_M-s_N\|_2 = \left(p\cdot \int_{[0,1]}\int_{S^{p-1}} |\, s_M(u,\alpha)-s_N(u,\alpha)\,|^2 \mu(du)\,\lambda(d\alpha)\right)^{1/2}.$$

Set-defined distances (such as fuzzy distances) have been also proposed ([1]).

For example, a *fuzzy distance* of two non-empty one-dimensional fuzzy sets A and B is defined as the fuzzy set $D_f(A,B)=\left(\Re^+,\ \mu_{D_f}(A,B)\right)$ with the membership function

$$\mu_{D_f(A,B)}(y)=\begin{cases}\sup\{\alpha\in[0,1]\mid A_\alpha, B_\alpha\neq\varnothing;\ \rho(A_\alpha,B_\alpha)\le y\} \\ \quad for\ y\le \lim\limits_{\beta\to\min\{h(M),h(N)\}_-}\rho(A_\beta,B_\beta) \\ 0 \qquad\qquad otherwise\end{cases}$$

A generalization of a fuzzy distance to the *p*-dimensional feature space is easiest if we define fuzzy granules as a Cartesian product of fuzzy sets, where the expression for the fuzzy distance can be simplified using distances in each dimension. A simplification can be obtained by choosing some particular kinds of fuzzy sets.

2.2 *A fuzzy distance between p-dimensional fuzzy granules*

In [5] we introduced a fuzzy distance between *p*-dimensional fuzzy granules defined as a Cartesian product of LR-fuzzy sets. Let us consider a p-dimensional feature space $X=X_1\times\cdots\times X_p$ and a configuration of n objects, each *one described by a p-dimensional fuzzy granule. Two arbitrary objects u and v are imprecisely located in the feature space by means of two fuzzy granules:* $A_u=A_{u1}\times\cdots\times A_{up}$ *and* $A_v=A_{v1}\times\cdots\times A_{vp}$. *Assume the components* A_{uk} *and* A_{vk} $(k=1,\ldots,p)$ *belonging to the* k^{th} *dimension of X are LR-fuzzy sets, i.e.* $A_{uk}(\alpha)=(A^L_{uk}(\alpha),\ A^R_{uk}(\alpha))$ *and* $A_{vk}(\alpha)=(A^L_{vk}(\alpha),\ A^R_{vk}(\alpha))$. *Therefore, the following inequalities hold:*

$$A^L_{uk}(\alpha)\le x_{uk}(\alpha)\le A^R_{uk}(\alpha);\quad A^L_{vk}(\alpha)\le x_{vk}(\alpha)\le A^R_{vk}(\alpha).$$

By subtracting $x_{vk}(\alpha)$ from $x_{uk}(\alpha)$, we obtain

$$A^L_{uk}(\alpha)-A^R_{vk}(\alpha)\le x_{uk}(\alpha)-x_{vk}(\alpha)\le A^R_{uk}(\alpha)-A^L_{vk}(\alpha)$$

and then, using well-known properties such as $\min\limits_{w\le x\le z}|\,x\,|=\max(0,w,-z)$ and $\max\limits_{w\le x\le z}|\,x\,|=\max(w,-z)$, respectively, the range of $|x_{uk}(\alpha)-x_{vk}(\alpha)|$, for any α in $[0,1]$, can be written as follows:

$$|x_{uk}(\alpha)-x_{vk}(\alpha)|\in \Big[\max\Big(0,\ A^L_{uk}(\alpha)-A^R_{vk}(\alpha),\ A^L_{vk}(\alpha)-A^R_{uk}(\alpha)\Big), \max\Big(A^R_{uk}(\alpha)-A^L_{vk}(\alpha),\ A^R_{vk}(\alpha)-A^L_{uk}(\alpha)\Big)\Big].$$

Now, the identity

$$\max(w,z)=\frac{1}{2}\left(w+z+|\,w-z\,|\right),\ \forall w,z\in\Re,$$

allows us to derive the left and right $\alpha-$bounds of a granular distance, namely the fuzzy distance. We have:

$$d^{\min}_{uv}(\alpha)=\frac{1}{4}\Bigg(\sum_{k=1}^{p}\Big[A^L_{uk}(\alpha)-A^R_{vk}(\alpha)+A^L_{vk}(\alpha)-A^R_{uk}(\alpha)+ +|\,A^L_{uk}(\alpha)-A^R_{vk}(\alpha)-A^L_{vk}(\alpha)+A^R_{uk}(\alpha)\,|+ +\Big|A^L_{uk}(\alpha)-A^R_{vk}(\alpha)+A^L_{vk}(\alpha)-A^R_{uk}(\alpha)+ +|\,A^L_{uk}(\alpha)-A^R_{vk}(\alpha)-A^L_{vk}(\alpha)+A^R_{uk}(\alpha)\,|\Big|\Big]^2\Bigg)^{\frac{1}{2}}$$

$$d^{\max}_{uv}(\alpha)=\frac{1}{2}\Bigg(\sum_{k=1}^{p}\Big[A^R_{uk}(\alpha)-A^L_{vk}(\alpha)+A^R_{vk}(\alpha)-A^L_{uk}(\alpha)+ +|\,A^R_{uk}(\alpha)-A^L_{vk}(\alpha)-A^R_{vk}(\alpha)+A^L_{uk}(\alpha)\,|\Big]^2\Bigg)^{\frac{1}{2}}.$$

Finally, the fuzzy distance between the two p-dimensional fuzzy granules A_u and A_v results in the form of a LR-fuzzy set:

$$d_{uv}(\alpha)=\left(d^{\min}_{uv}(\alpha),\ d^{\max}_{uv}(\alpha)\right)=\left(d^L_{uv}(\alpha),\ d^R_{uv}(\alpha)\right).$$

A point-wise distance can then be obtained by defuzzifying the fuzzy distance d_{uv}, e.g. by computing its centroid. An alternative way may be that of averaging the α-level intervals $[d^L_{uv}(\alpha),\ d^R_{uv}(\alpha)]$ over $[0,1]$:

$$Defuzz(d_{uv})=E\left[d_{uv}(\alpha)\right]=\frac{1}{2}\int_{[0,1]}\left|d^R_{uv}(\alpha)+d^L_{uv}(\alpha)\right|\lambda(d\alpha).$$

where λ is a normalized Lebesgue measure on $[0,1]$, e.g.

$$\lambda([0,1])=\int_0^1 w(\alpha)\,d\alpha=1,\ \text{where}\ w(\alpha)=2\alpha.$$

Remark: The fuzzy distance between two p-dimensional fuzzy granules A_u and A_v with empty intersection, whose components A_{uk} and A_{vk} are trapezoidal fuzzy sets for all $k=1,\ldots,p$, is still a trapezoidal fuzzy set. This is because, in

ISBN: 978-989-95079-6-8

such a case, $d_{uv}^{L}(\alpha)$ is an increasing function that interpolates linearly between $d_{uv}^{L}(0)$ and $d_{uv}^{L}(1)$ whereas $d_{uv}^{R}(\alpha)$ is a decreasing function that interpolates linearly between $d_{uv}^{R}(1)$ and $d_{uv}^{R}(0)$ (see Fig. 1).

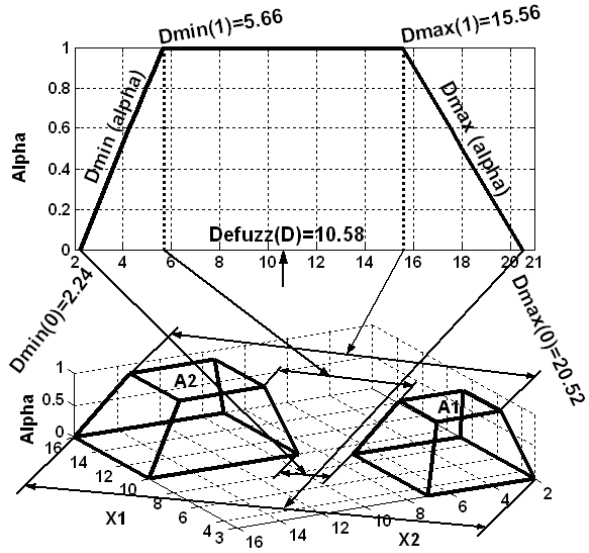


Figure 1: Two-dimensional fuzzy granules, the fuzzy distance between them and its defuzzified value.

The fuzzy distance between pairs of fuzzy granules with non-empty intersection, is linear-shaped, but not necessarily trapezoidal-shaped (see Fig. 2).

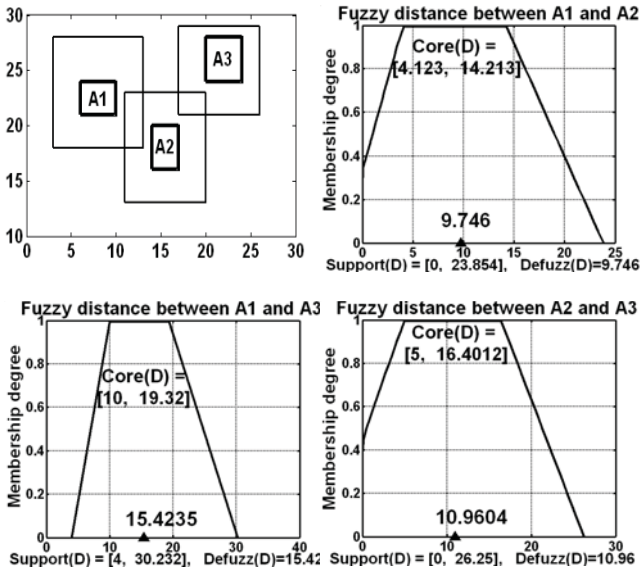


Figure 2: Two-dimensional fuzzy granules with non-empty intersection and the pair-wise fuzzy distances.

Obviously, our formal definition of fuzzy distance introduced above applies similarly to fuzzy granules of any finite dimension p. Fig. 3 illustrates the case of three-dimensional fuzzy granules.

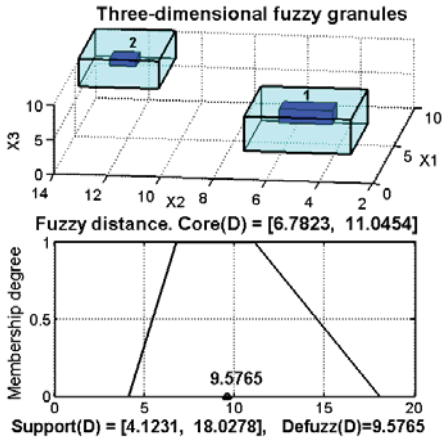


Figure 3: Three-dimensional fuzzy granules, the fuzzy distance and its defuzzified value.

3 Granular Clustering

3.1 A generalization of fuzzy c-means algorithm to granular feature spaces

We propose a generalization of Fuzzy C-Means Algorithm to fuzzy feature spaces, which leads to adopting a well defined distance between p-dimensional fuzzy granules (either a crisp distance, or the defuzzified value of the fuzzy distance defined above).

The Fuzzy C-means clustering algorithm is based on the minimization of an objective function called *C-means functional*:

$$J(X;U,V)=\sum_{i=1}^{c}\sum_{k=1}^{N}(u_{ik})^{m}\left\|x_k-v_i\right\|_A^2$$

where

$$V=(v_1,\ldots,v_c),\qquad v_i\in\Re^n$$

is a vector of *cluster prototypes* (centers), which have to be determined, and

$$D_{ikA}^2=\left\|x_k-v_i\right\|_A^2=(x_k-v_i)'A(x_k-v_i)$$

is a squared inner-product distance norm.

In a granular feature space, such as a space populated with p-dimensional fuzzy granules, x_k are granular samples and v_i are granular centroids.

Remark. The granular centroid of the set $\{A_u\}_{u=1,\ldots,n}$ of LR-type p-dimensional fuzzy granules is still an LR-type p-dimensional fuzzy granule.

Fig. 4 shows a set $\{A_1, A_2, A_3\}$ of 3-dimensional fuzzy granules, their granular centroid G, as well as the three pair-wise granular centroids G_{12}, G_{13} and G_{23}, respectively.

Furthermore, the distance D_{ikA}^2 must be replaced with a well defined square distance between p-dimensional fuzzy granules. Since the fuzzy distance introduced above have been defined as a square distance, we can use the square value of the fuzzy distance in order to replace D_{ikA}^2.

ISBN: 978-989-95079-6-8

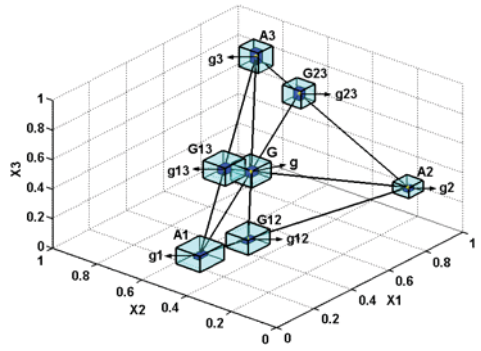


Figure 4. Granular and point-wise weighted centroids of 3D fuzzy granules (the granular version of median theorem).

For testing the granular clustering algorithm we started from a set of crisp data (Fig. 5).

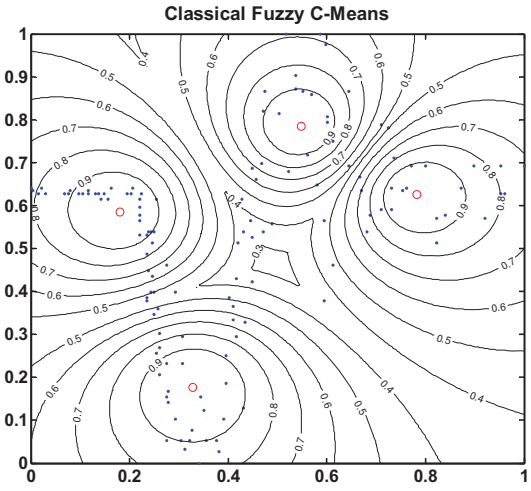


Figure 5: Fuzzy C-means clustering for crisp data.

Then, we transformed these initial data into two-dimensional fuzzy granules, i.e. as a Cartesian product of trapezoidal shaped fuzzy sets on each dimension. Fig. 6 shows 4 fuzzy clusters populated with two-dimensional fuzzy granules

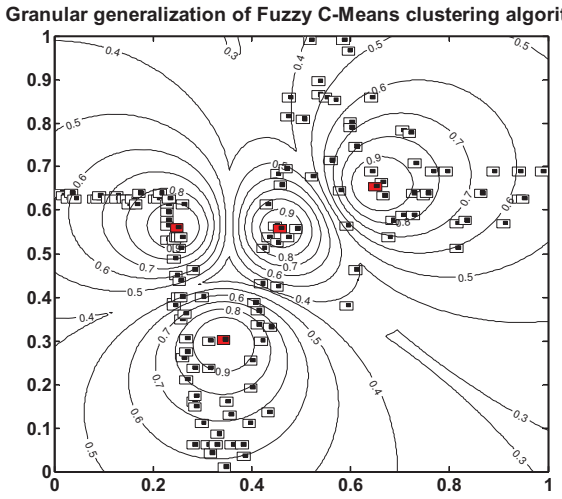


Figure 6: Granular fuzzy C-means clustering.

3.2 *Related work*

In ([4], [5]) we proposed an agglomerative granular clustering method that generalizes the method in [12], originally proposed by Pedrycz and Bargiela. Since resorting to a different granulation framework (fuzzy set theory instead of interval mathematics) involves a refinement of the metric formalism, the fuzzy distance has been used for allowing fuzzy granules (rather than hyperboxes) to be processed during the cluster growing process. The main step of the iterative process is finding the two closest information granules in order to aggregate them into a more comprehensive one. Let $C = agg(A, B)$ be the resulting granule. In terms of p-dimensional LR-fuzzy granules, the aggregation process can be carried out as follows:

$$C = C_1 \times \cdots \times C_p$$

with $C_i^{\alpha} = \left(x_{C_i}^L - \ell_{C_i} \cdot L^{-1}(\alpha),\ x_{C_i}^R + r_{C_i} \cdot R^{-1}(\alpha)\right)$.

The core and the support of C_i is obtained as

$$C_i^1 = \left(x_{C_i}^L, x_{C_i}^R\right) = \left(\min\left(x_{A_i}^L, x_{B_i}^L\right),\ \max\left(x_{A_i}^R, x_{B_i}^R\right)\right)$$

$$C_i^0 = \left(x_{C_i}^L - \ell_{C_i}, x_{C_i}^R + r_{C_i}\right) =$$

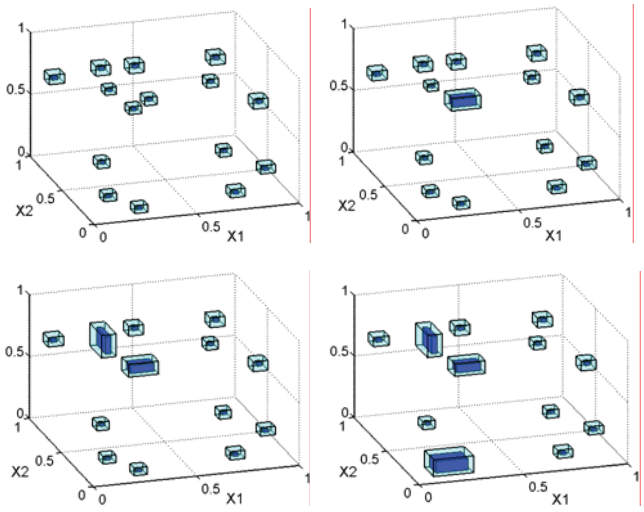
$$= \left(\min(x_{A_i}^L - \ell_{A_i}, x_{B_i}^L - \ell_{B_i}, \max(x_{A_i}^R + r_{A_i}, x_{B_i}^R + r_{B_i})\right).$$

The compatibility measure guiding the search for the two closest fuzzy granules can now be defined as

$$compat(A, B) = 1 - Defuzz\left(d(A, B)\right) \cdot e^{-\beta|C|}$$

where $|C| = \int_0^1 \left|C^{\alpha}\right| d\alpha = \int_0^1 \prod_{i=1}^{p} \left|C_i^{\alpha}\right| d\alpha$ is the cardinality of C and β is a tuning coefficient.

Maximizing the compatibility measure means that the pair of candidate fuzzy granules to be clustered should not only be close enough (i.e., the distance between them should be small), but the resulting granule should be compact (i.e., its expansion along every direction must be well-balanced). Fig. 7 shows the sequence of cluster growing over the granular clustering process.



ISBN: 978-989-95079-6-8

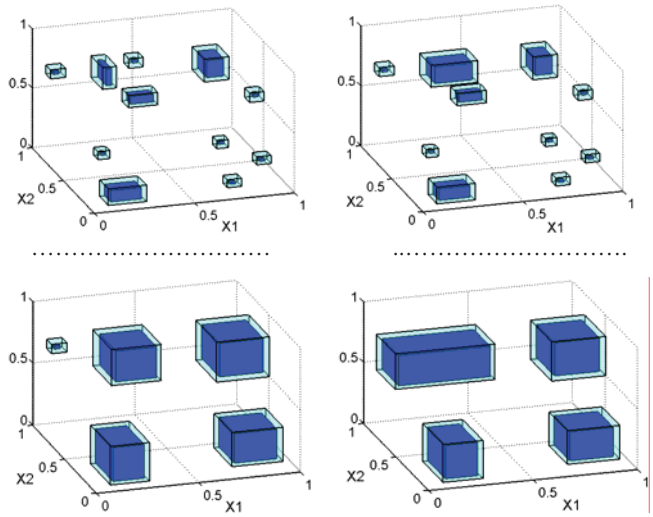


Figure 7: The sequence of cluster growing over the granular clustering process.

4 Time Series Clustering

4.1 Time series (dis)similarity

Roughly speaking, a granule is a set-defined object. In these terms, a subsequence time series can be viewed as a special kind of granule.

Let $S = y_m, \ldots, y_{m+w-1}$ be a subsequence with length w of time series $Y = y_1, \ldots, y_n$, where $1 \le m \le n - w + 1$. Subsequences will be represented as vectors in a w-dimensional vector space.

Clustering of subsequence time series requires the definition of a similarity or of a distance measure. Examples of dissimilarity measures are based on Euclidean norms, piecewise linear approximations, dynamic time warping, longest common subsequences, and probabilistic similarity models.

The usage of the Euclidean distance is subject to the constraint that both time subsequences are of the same length w. Thus we can define the dissimilarity between sequences S_1 and S_2 as $L_p(S_1, S_2)$, that is, the distance between the two $w-$dimensional vectors measured by the L_p norm (when p = 2, this reduces to the familiar Euclidean distance).

Some major disadvantages of the Euclidean distance are as follows: it does not allow for different *baselines* in the time sequences; it is very sensitive to *phase shifts* in time; it does not allow for *acceleration and deceleration* along the time axis.

Other specialized distance measures have been described for time series clustering, such as Dynamic Time Warping, DTW, and Longest Common Subsequence Similarity, LCSS.

Dynamic Time Warping (DTW) is an extensively used technique in speech recognition and allows acceleration–deceleration of signals along the time dimension.

Following [11], let us consider two sequences (of possibly different lengths) $Q = \{q_1, \ldots, q_n\}$ and $C = \{c_1, \ldots, , c_m\}$.

To align two sequences using DTW, we construct an n-by-m matrix where the (i^{th}, j^{th}) element of the matrix contains the distance $d(q_i , c_j)$ between the two points q_i and c_j (i.e. $d(q_i, c_j) = (q_i - c_j)^2$). Each matrix element (i, j) corresponds to the alignment between the points q_i and c_j. A warping path W is a contiguous set of matrix elements that defines a mapping between Q and C. The k^{th} element of W is defined as $w_k = (i, j)_k$, so we have:

$$W = w_1, w_2, \ldots, w_k, \ldots, w_K, \qquad \max(m,n) \le K < m+n-1.$$

The warping path is typically subject to several constraints.

- Boundary conditions: $w_1 = (1, 1)$ and $w_K = (m, n)$.
- Continuity: Given $w_k = (a, b)$, then $w_{k-1} = (a', b')$, where $a - a' \le 1$ and $b - b' \le 1$. This restricts the allowable steps in the warping path to adjacent cells.
- Monotonicity: Given $w_k = (a, b)$, then $w_{k-1} = (a', b')$, where $a - a' \ge 0$ and $b - b' \ge 0$. This forces the points in W to be monotonically spaced in time.

There are exponentially many warping paths that satisfy the above conditions, however we are interested only in the path which minimizes the warping cost:

$$DTW(Q, C) = \min\left(\sqrt{\sum_{i=1}^{K} w_k}\right) \Big/ K .$$

The K in the denominator is used to compensate for the fact that warping paths may have different lengths.

This path can be found efficiently using dynamic programming to evaluate the following recurrence, which defines the cumulative distance $\gamma(i, j)$ as the distance $d(i, j)$ found in the current cell and the minimum of the cumulative distances of the adjacent elements:

$$\gamma(i, j) = d(q_i, c_j) + \min\{\gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1)\}.$$

The Euclidean distance between two sequences can be seen as a special case of DTW where the k^{th} element of W is constrained such that $w_k = (i, j)_k$, $i = j = k$.

The warping path is also constrained in a global sense by limiting how far it may stray from the diagonal. The subset of the matrix that the warping path is allowed to visit is called the warping window. The two most common constraints in the literature are the Sakoe-Chiba band and the Itakura parallelogram. We can view a global or local constraint as constraining the indices of the warping path $w_k = (i, j)_k$, such that $j - r \le i \le j + r$, where r is a term defining the allowed range of warping, for a given point in a sequence. In the case of the Sakoe-Chiba band, r is independent of i; for the Itakura parallelogram, r is a function of i.

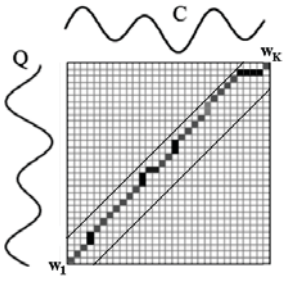


Figure 8: Optimal warping path with the Sakoe-Chiba band as global constraints.

ISBN: 978-989-95079-6-8

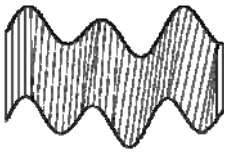


Figure 9: Aligning two time sequences using DTW.

4.2 Non-overlapping subsequence time series clustering

Our implementation of Fuzzy C-Means algorithm for non-overlapping subsequence time series clustering is based essentially on using the DTW distance, which is proved to be largely superior to the Euclidian distance.

We also report a real-world application of this method for clustering non-overlapping subsequences of length 30 from Bucharest Stock Exchange Bet Index. The dataset contains 2210 daily values. Fig. 10 shows ten clusters, which consist of several subsequence time series, grouped around a cluster centroid.

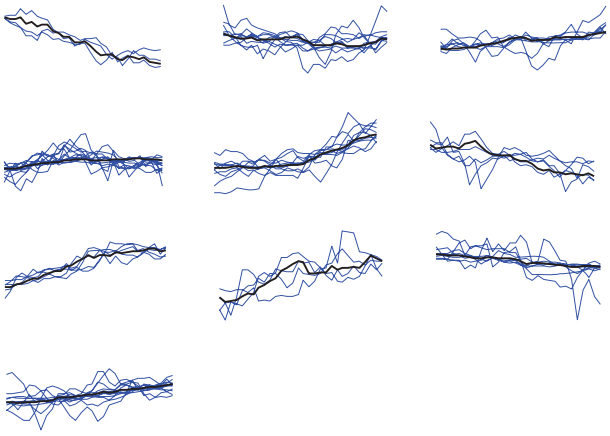


Figure 10: Non-overlapping subsequence time series of length 30: clusters and centroids obtained by fuzzy c-means.

4.3 Overlapping subsequence time series clustering

Subsequence Time Series (STS) clustering typically employs a clustering technique to the subsequences of a time series generated using a sliding window technique (Fig. 11). Although it is very popular, Keogh et al. [9] reported, for the first time, a surprising anomaly: cluster centers obtained using STS clustering closely resemble "sine waves", irrespective of the nature of original time series itself, and therefore, the results of STS clustering are meaningless.

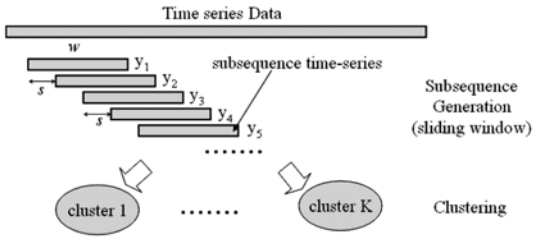


Figure 11: STS clustering using a sliding window technique.

According to the Keogh's criticism, superposition of slightly shifted subsequences causes the generation of sine waves. In order to explicitly exclude superposition, a motif-based clustering algorithm has been proposed in [10]. Another solution to avoid sine wave has been reported in [2], which consists of aligning the phase at the frequency with maximum spectrum power with respect to each time series subsequence, and then, to apply a clustering algorithm to them. Although these approaches are promising, the need of identifying solid mathematical foundations behind STS clustering steel remains.

References

[1] I. Bloch. On fuzzy distances and their use in image processing under imprecision. *Pattern Recognition*, 32:1873–1895, 1999.

[2] R. Fujimaki, S. Hirose and T. Nakata. Theoretical analysis of subsequence time-series clustering from a frequency-analysis viewpoint. *SIAM International Conference on Data Mining*, 506-517, 2008.

[3] V. Georgescu. Granular vs. point-wise metrics and statistics to accommodate machine learning algorithms in granular feature spaces, *Fuzzy Economic Review,* vol.XII, No.2, 45-74, 2007.

[4] V. Georgescu. A Generalization of Symbolic Data Analysis Allowing the Processing of Fuzzy Granules, *Lecture Notes in Artificial Intelligence,* vol. 3131, 215-226, 2004.

[5] V. Georgescu. A Formal Framework Allowing to Deal with Granular Clustering not only in Terms of Hyperboxes, but also in terms of Fuzzy Granules, *Proceedings of MS'2004-France*, Lyon, July 5–7, 2004.

[6] V. Georgescu. On the Foundations of Granular Computing Paradigm, *Fuzzy Economic Review,* vol.VIII, No.2, 73-105, 2003.

[7] V. Georgescu. Multivariate Fuzzy-termed Data Analysis: Issues and Methods, *Fuzzy Economic Review,* vol.VI, No.1, 19-48, 2001.

[8] D. Gunopulos and G. Das. Time series similarity measures and time series indexing. *ACM SIGMOD Conference*, Santa Barbara, CA., 2001.

[9] E. Keogh, J. Lin and W. Truppel. Clustering of time series subsequences is meaningless: implications for previous and future research. *Proceedings of the 3rd IEEE International Conference on Data* Mining, 115–122, 2003.

[10] E. Keogh, B. Chiu and S. Lonardi. Probabilistic discovery of time series motifs. *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2003.

[11] E. Keogh and C. A. Ratanamahatana. Exact indexing of dynamic time warping. *Knowledge and Information Systems*, 7: 358-386, 2005.

[12] W. Pedrycz and A. Bargiela Granular Clustering: a Granular Signature of Data, *IEEE Transactions on Systems, Man and Cybernetic.,* vol. 32, No.2, 212-224, 2002.

[13] W. Pedrycz. *Knowledge-Based Clustering. From Data to Information Granules*, John Wiley & Sons, 2005.

[14] W. Pedrycz. (ed.). *Granular Computing: An Emerging Paradigm*, Physica-Verlag, 2001.

[15] L. A. Zadeh. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, *Fuzzy Sets and Systems*, 90: 111-117, 1997.

ISBN: 978-989-95079-6-8

Enhancing OLAP Querying with the aid of H-IFS

Panagiotis Chountas[1] Ermir Rogova[1] Krassimir T. Atanassov[2]

[1]DKMG – SECS, University of Westminster, London, HA1 3TP, UK
[2]CLBME, Bulgarian Academy of Sciences, Bl. 105, Sofia-1113, BULGARIA
Email: chountp@wmin.ac.uk, krat@argo.bas.bg

Abstract— *Over the past years we have witnessed an increasing interest in expressing user or domain preferences or knowledge inside database queries. First, it appeared to be desirable property of a query system to offer more expressive query languages that can be more faithful to what a user intends to say. Second, a classical query in the sense of relational paradigm may also have a restricted answer or sometimes an empty set of answers, while a relaxed version of the query enhanced with background or domain knowledge might be matched by some items in the database.*

Keywords— Query answering, OLAP, IFS, multidimensional modelling.

1 Introduction

The need for flexible systems to manage value uncertainty has been the focus for database researchers [1], mainly at theoretical level and in the context of the relational model.

At the same time OLAP technology required [2] the extension of the relational systems with the inclusion of the data-cube and operators to operate over it. Alternatively, new models [3] were proposed to support OLAP based querying on top of multidimensional views. Both approaches support the organisation of data around several axes of analysis. In OLAP based systems, when it comes to the model level, support for value uncertainty will be required at the fact level as well at the level of dimensions with the support of non-rigid hierarchies [4]. Still [4] considers that facts and dimensions as in [5], [6] represent structural information [7].

Current research issues for OLAP systems can be summarised as follows

- Flexible models are required to support value uncertainty at fact level as well as at the dimension level with the provision of non rigid dimensions
- Flexibility should not be limited at the structural level. It should be allowed also at the query level. Users should be allowed to synthesise their own model of dimensions for analysis purposes based on existing structure. Dimensions may be based in either rigid or non-rigid hierarchies.

Recently in OLAP systems a need has been identified for enhancing the query scope with the aid of "kind-of" relation that describe knowledge as well as ordering of the elements of a domain or a hierarchical universe.

However, in our context, the terms of the hierarchy [8], and the relations between terms are not fuzzy and do not represent [9] only kind-of relations. These observations led us to introduce the concept of closure of the H-IFS which is a developed form defined on the whole hierarchy.

The rest of the paper is organised as follows; In Section two we define the basic properties of Intuitionistic Fuzzy sets and H-IFS. In Section 3 the semantics of the H-IFS cubic representation [14], [15] in contrast to the basic multidimensional-cubic structures are presented we also define the extended SQL-OLAP aggregators. In Section 4 we present the main concepts involved in the designing and implementation the 'IF-Oracle' OLAP utility and also demonstrate the potential of 'IF-Oracle' utility. Finally we conclude and compare our proposal with other school of thoughts.

2 Intuitionistic Fuzzy Sets A-IFS

2.1 Foundations

Each element of an Intuitionistic fuzzy [10, 11] set has degrees of membership or truth (μ) and non-membership or falsity (ν), which don't sum up to 1.0 thus leaving a degree of hesitation margin (π).

As opposed to the classical definition of a fuzzy set given by

A′= $\widetilde{A} = \{< x, \mu_A(x) > | x \in X\}$ where $\mu_A(x) \in [0,1]$ is the membership function of the fuzzy set A′, an Intuitionistic fuzzy set A is given by

$$A = \{< x, \mu_A(x), \nu_A(x) > | x \in X\}$$
$$\mu_A : X \to [0,1] \text{ and } \nu_A : X \to [0,1]$$

Such that

$$0 \le \mu_A(x) + \nu_A(x) \le 1$$

and $\mu_A : X \to [0,1], \nu_A : X \to [0,1]$ denote a degree of membership and a degree of non-membership of X∈A, respectively. Obviously, each fuzzy set may be represented by the following Intuitionistic fuzzy set;

$$A = \{< x, \mu_A(x), 1 - \mu_A(x) > | x \in X\}$$

For each Intuitionistic fuzzy set in X, we will call $\pi_A(x) = 1 - \mu_A(x) - \nu_A(x)$, Intuitionistic fuzzy index of $x \in A$ which expresses a lack of knowledge of whether x belongs to A or not. For each $x \in A$ $0 \le \pi_A(x) \le 1$.

Definition 1. Let A and B be two Intuitionistic fuzzy sets defined on a domain X. A is included in B (denoted A ⊆ B) if and only if their membership functions and non-membership functions satisfy the condition:

ISBN: 978-989-95079-6-8

$(\forall X \in X)\ (\mu_A(x) \leq \mu_B(x)\ \&\ \nu_A(x) \geq \nu_B(x))$

Two scalar measures are classically used in classical fuzzy pattern matching to evaluate the compatibility between an ill-known datum and a flexible query, known as

- a possibility degree of matching, **Π(Q/ D)**
- a necessity degree of matching, **N(Q/ D)**

Definition 2. Let Q and D be two Intuitionistic fuzzy sets defined on a domain X and representing, respectively, a flexible query and an ill-known datum.

The possibility degree of matching between Q and D, denoted Π(Q/D), is an "optimistic" degree of overlapping that measures the maximum compatibility between Q and D, and is defined by:

$$\Pi(Q/D) = \left\langle \sup_{x\in X} \min(1-\nu_Q(x), \nu_Q(x)), \inf_{x\in X} \max(1-\nu_D(x), \nu_D(x)) \right\rangle$$

The necessity degree of matching between Q and D, denoted N(Q/D), is a "pessimistic" degree of inclusion that estimates the extent to which it is certain that D is compatible with Q, and is defined by:

$$N(Q/D) = \left\langle \inf_{x\in X} \max(\mu_Q(x), 1-\mu_Q(x)), \sup_{x\in X} \min(\mu_D(x), 1-\mu_D(x)) \right\rangle$$

The problem occurring from defining Intuitionistic fuzzy sets based on the "kind-of" relation is that two different Intuitionistic fuzzy sets on the same hierarchy do not necessarily have the same definition domain, which means they cannot be compared using the classic comparison operations **Π**(Q/ D), N(Q/ D).

2.2 From IFS to H-IFS

The definition domains of the hierarchical fuzzy sets [12, 13] that we propose below are subsets of hierarchies composed of elements partially ordered by the "kind of" relation. An element l_i is more general than an element l_j (denoted $l_i \sim l_j$), if l_i is a predecessor of l_j in the partial order induced by the "kind-of" relation of the hierarchy. An example of such a hierarchy is given in 'Figure 1'.

Definition 3. Let F be a H-IFS defined on a subset D of the elements of a hierarchy L. It degree is denoted as <μ, ν>. The closure of F, denoted clos(F), is a H-IFS defined on the whole set of elements of L and its degree $<\mu, \nu>_{clos(F)}$ is defined as follows.

For each element **l** of L, let $S_L = \{l_1, \ldots, l_n\}$ be the set of the smallest super-elements in D.

If S_L is not empty,

$$<\mu, \nu>_{clos(F)}(S_L) = <\max_{1\leq i\leq n} \mu(L_i), \min_{1\leq i\leq n} \nu(L_i)>$$

else

$$<\mu, \nu>_{clos(F)}(S_L) = <0,0>$$

In other words, the closure of a H-IFS F is built according to the following rules. For each element l_1 of L:

- If l_I belongs to F, then l_I keeps the same degree in the closure of F (case where $S_L = \{l_I\}$).
- If l_I has a unique smallest super-element l_I in F, then the degree associated with l_I is propagated to L in the closure of F, $S_L = \{l_I\}$ with $l_I > l_I$)

If L has several smallest super-elements $\{l_1, \ldots, l_n\}$ in F, with different degrees, a choice has to be made concerning the degree that will be associated with l_I in the closure. The proposition put forward in definition 3, consists of choosing the maximum degree of validity μ and minimum degree of non validity v associated with $\{l_1, \ldots, l_n\}$. We refer to as the *Optimistic strategy*.

The proposition put forward in definition 3, consists of choosing the maximum degree of validity μ and minimum degree of non validity v associated with $\{l_1, \ldots, l_n\}$. We refer to as the *Optimistic strategy*.

We can also utilise a *Pessimistic strategy* which consists of choosing the minimum degree of validity μ and maximum degree of non validity v associated with $\{l_1, \ldots, l_n\}$.Alternatively, an *Average strategy* could be utilised, which consists of calculating the IF-Average and applying it to the degrees of validity μ and non-validity v.

It has been observed that two different H-IFSs, defined on the same hierarchy, can have the same closure, as in the following example.

The H-IFSs Q={Wine<1,0>, Red Wine<0.7,0.1>, Brown Wine<1,0>, White Wine <0.4,0.3>} and R={Wine<1,0>, Red Wine<0.7,0.1>, Brown Wine<1,0>, Pinot Noir <1.0,0.0>} have the same closure, represented 'Figure 1' below.

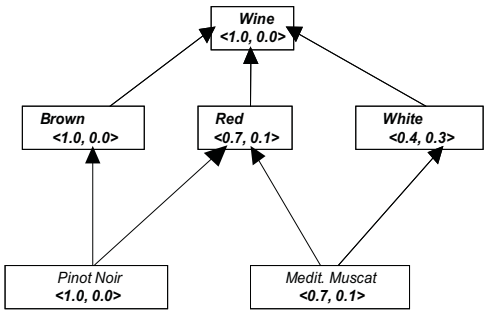


Figure 1: Common closure of the H-IFS's Q and R

Such H-IFS's form equivalence classes with respect to their closures.

Definition 4. Let F be a hierarchical fuzzy set, with dom(F) = $\{l_1, \ldots, l_n\}$, and F_{-k} the H-IFS resulting from the restriction of F to the domain dom(F) \ $\{l_k\}$. l_k is deducible in F if

$$<\mu, \nu>clos_{(F-k)}(l_k) = <\mu, \nu>clos_{(F)}(l_k)$$

As a first intuition, it can be said that removing a derivable element from a hierarchical fuzzy set allows one to eliminate redundant information. But, an element being derivable in F does not necessarily mean that removing it from F will have no consequence on the closure: removing k from F will not impact the degree associated with k itself in the closure, but it may impact the degrees of the sub-elements of k in the closure.

For instance, if the element Brown Wine is derivable in Q, according to *definition 4*, removing Brown Wine <1,0> from Q would not modify the degree of Brown Wine itself in the resulting closure, but it could modify the degree of its

ISBN: 978-989-95079-6-8

sub-element Pinot Noir. Thus, Brown Wine <1,0> cannot be derived or removed. This remark leads us to the following definition of a minimal hierarchical fuzzy set.

Definition 5. In a given equivalence class (that is, for a given closure C), a hierarchical fuzzy set is said to be **minimal** if its closure is C and if none of the elements of its domain is derivable**.**

Obtaining the Minimal H-IFS

Step 1: Assign Min-H-IFS ← ∅. Establish an order so that the sub-elements $\{l_1,\ldots,l_n\}$ of the hierarchy L are examined after its super-elements.

Step 2: Let l_1 be the first element & $(l_1)/\langle\mu, \nu\rangle \neq (l_1)/\langle 0, 0\rangle$ then add l_1 to Min-H-IFS and

$$\langle\mu, \nu\rangle_{clos(Min\text{-}HIFS)}\ (l_1)= (l_1)/\langle\mu, \nu\rangle.$$

Step 3: Let us assume that K elements of the hierarchy L satisfy the condition $\langle\mu, \nu\rangle_{clos(Min\text{-}HIFS)}\ (l_i)=(l_i)/\langle\mu, \nu\rangle$.
In this case the Min-H-IFS do not change. Otherwise go to next element l_{k+1} and execute Step 4.

Step 4: The $l_{k+1}/\langle\mu_{k+1}, \nu_{k+1}\rangle$ associated with l_{k+1}.
In this case l_{k+1} is added to Min-H-IFS with the corresponding $\langle\mu_{k+1}, \nu_{k+1}\rangle$.

Step 5: Repeat steps three and four until $clos_{(Min\text{-}HIFS)}=C$.

For instance S_1 and S_2 are **minimal** (none of their elements is derivable). They cannot be reduced further.

S_1= Wine<1,0>

S_2={Wine<1,0>, Red Wine<0.7,0.1>, Pinot Noir<1,0>, White Wine <0.4, 0.3>}

3 Representing H-IFS Inside Cubes

In this section we present the semantics of the H-IFS cubic representation [14], [15] in contrast to the basic multidimensional-cubic structures. The basic cubic operators are extended and enhanced with the aid of Intuitionistic Fuzzy Logic.

Since the emergence of the OLAP technology [16] different proposals have been made to give support to different types of data and application purposes. One of these is to extend the relational model (ROLAP) to support the structures and operations typical of OLAP. Further approaches [17], are based on extended relational systems to represent data-cubes and operate over them.

Nowadays, information and knowledge-based systems need to manage imprecision in the data and more flexible structures are needed to represent the analysis domain. New models have appeared to manage incomplete datacube [18], imprecision in the facts and the definition of fact using different levels in the dimensions.

Nevertheless, these models continue to use inflexible hierarchies thus making it difficult to merge reconcilable data from different sources with some incompatibilities in their schemata. These incompatibilities arise due to different perceptions-views about a particular modelling reality.

In addressing the problem of representing flexible hierarchies we propose a new multidimensional model that is able to treat with imprecision over conceptual hierarchies based on Intuitionistic Fuzzy logic. The use of conceptual hierarchies enables us to:

- define the structures of a dimension in a more perceptive way to the final user, thus allowing a more perceptive use of the system.
- query information from different sources or even use information or preferences given by experts to improve the description of hierarchies, thereby getting more knowledgeable query results. We outline a unique way for incorporating "H-IFS" relations, or conceptual imprecise hierarchies as dimensions with respect to the model proposed in [19].

3.1 Semantics of the IF-Cube in contrast to Crisp Cube

In this section we review the semantics of Multidimensional modelling and Intuitionistic Fuzzy Logic and based on these we propose a unique concept named as Intuitionistic Fuzzy Cube (IF-Cube). The IF-Cube is the basis for the representation of flexible hierarchies and thus flexible facts.

3.1.1 Overview of the Cube Model

A logical model that influences the database design and the query engines is the *multidimensional-cubic* view of data in the warehouse. In a multidimensional data model, there is a set of *numeric measures* that are the objects of analysis. Examples of such measures are sales, budget, etc. Each of the numeric measures depends on a set of *dimensions,* which provide the context for the measure. The attributes of a dimension may be related via a hierarchy of relationships. In the above example, the product name is related to its category and the industry attribute through a hierarchical relationship, see Fig. 2

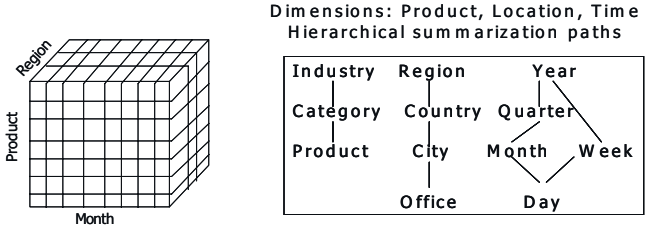


Figure 2: Cube_Sales, Rigid Hierarchies for Product, Location, Time Dimensions

According to [19] a cube structure is defined as a 4-tuple, <D, M, A, f> where the four components indicate the characteristics of the cube. These characteristics are:
a set of n dimensions $D = \{d_1, d_2, \ldots, d_n\}$ where each d_i is a dimension name, extracted from a domain $dom_{dim(i)}$. A set of k measures $M = \{m_1, m_2, \ldots, m_k\}$ where each m_i is a measure name, extracted from a domain $dom_{measure(i)}$. The set of dimension names and measures names are disjoint; i.e., $D \cap M = 0$. A set of t attributes $A = \{a_1, a_2, \ldots, a_t\}$ where each a_i is an attribute name, extracted from a domain $dom_{attr(i)}$. A one-to-many mapping $f : D \rightarrow A$, i.e. there exists, corresponding to each dimension, a set of attributes.

3.1.2 Semantics of the IF-Cube

ISBN: 978-989-95079-6-8

In contrast, an **IF-Cube** is an abstract structure that serves as the foundation for the multidimensional data cube model. Cube C is defined as a five-tuple *(D, l, F, O, H)* where:

- *D* is a set of dimensions
- *l* is a set of levels $l_1, ..., l_n$,
- A dimension $d_i = (l \leq O, l_\perp, l_\top)$ $dom(d_i)$ where
- $l=l_i\ i=1...n.$ l_i is a set of values and $l_i \cap l_j = \{\}$
- $\leq O$ is a partial order between the elements of *l.*
- To identify the level *l* of a dimension, as part of a hierarchy we note it as *dl.* $l_\perp$: base level $l_\top$: top levelfor each pair of levels l_i and l_j we have the relation :

$\mu_{ij} : l_i \times l_j \rightarrow [0,1]$ $\quad v_{ij} : l_i \times l_j \rightarrow [0,1]$ $0 < \mu_{ij} + v_{ij} < 1$

- *F* is a set of fact instances with schema $F = \{<x, \mu_F(x), v_F(x)> | x \in X\}$, where $x=<att_1, ...,att_n>$ is an ordered tuple belonging to a given universe *X,* $\mu_F(x)$ *and* $v_F(x)$ are the degree of membership and non-membership of *x* in the fact table *F* respectively.
- *H* is an object type history that corresponds to a cubic structure*(l, F, O, H')* which allows us to trace back the evolution of a cubic structure after performing a set of operators i.e. aggregation.

The example below provides a sample imprecise cube (*D, l, F, O, H) i.e. sales* and a conceptual non-rigid hierarchy product with reference to wine consisting of $l_i, ..., l_n$ levels with respective levels of membership and non membership < μ_{ij} v_{ij}, >.

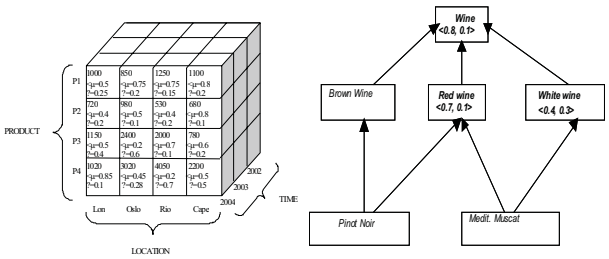


Figure 3: *Imprecise Cube Sales, IF Hierarchy 'Wine*

3.1.3 Cube Aggregation

Aggregation (A): An aggregation operator *A* is a function *A(G)* where $G = \{<x, \mu_F(x), v_F(x)> | x \in X\}$ where $x=<att_1, ...,att_n>$ is an ordered tuple belonging to a given universe.

X, $\{att_1, ..., att_n\}$ is the set of attributes of the elements of *X,* $\mu_F(x)$ *and* $v_F(x)$ are the degree of membership and non-membership of *x.*

The result is a bag of the type $\{<x', \mu_F(x'), v_F(x')> | x' \in X\}$. To this extent, the bag is a group of elements that can be duplicated and each one has a degree of μ and ν.

Input: C_i = *(D, l, F, O, H)* and the function *A(G)*

Output: C_o = (D, l_o, F_o, O_o, H_o)

The definition of the extended group operators allows us to define the extended group operators ***Drill up (Δ), and Drill Down (Ω).***

Drill up (Δ): The result of applying Drill up over dimension d_i at level dl_r using the aggregation operator A over a datacube $C_i = (D_i, l_i, F_i, O, H_i)$ is another datacube $C_o = (D_o, l_o, F_o, O, H_o)$.

Input: $C_i = (D_i, l_i, F_i, O, H_i)$

Output: $C_o = (D_o, l_o, F_o, O, H_o)$

An object of type history is a recursive structure. The structured history of the datacube allows us to keep all the information when applying *Drill up* and get it all back when *Drill Down* is performed. To be able to apply the operation of *Drill up* we need to make use of the IF_{SUM} aggregation operator.

Drill Down (Ω): This operator performs the opposite function of the *Drill up* operator. It is used to Drill Down from the higher levels of the hierarchy with a greater degree of generalization, to the leaves with the greater degree of precision. The result of applying *Drill Down* over a datacube C_i = *(D, l, F, O, H)* having H=*(l', D', A', H')* is another datacube C_o= *(D', l', F', O, H').*

Input: C_i=*(D, l, F, O, H)*

Output C_o=*(D',l',F',O,H')* where *F' → set of fact instances defined by operator A.*

The defined IF OLAP Cube model allows us to:

- accommodate imprecise facts
- utilise *conceptual hierarchies* used for aggregation purposes in the cases of roll-up and Drill Down operations.
- offer a unique feature such as keeping track of the history when we move between different levels of a hierarchical order.

In the next section we demonstrate the usefulness of the H-IFS notion and the extended aggregation operators for extending the query capabilities of Oracle10g. We developed an ad-hoc utility 'IF-Oracle' implemented on top of Oracle10g that allow us to:

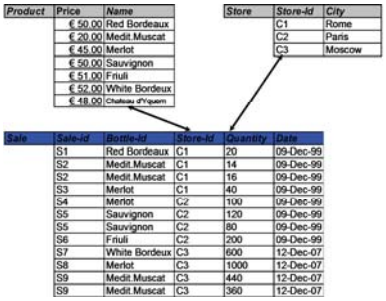
- Define an H-IFS hierarchy
- Incorporate hierarchical knowledge in the form of H-IFS as part of the standard OLAP queries.
- Enhance the scope of query answers against the Oracle10g standard query answers.

4 Embedding IF Cubes in Oracle10g

IF-Oracle has been developed using Visual Studio.Net as an ad-hoc utility that is attached to and enhances Oracle10g DBMS query capabilities.

For demonstrating the functionality of IF-Oracle let us consider a sample multidimensional model, Fig.4 in the form of a star schema that describes sales of Vitis Vinifera type wines.

ISBN: 978-989-95079-6-8



Product	Price	Name
	€ 50,00	Red Bordeaux
	€ 20,00	Medit.Muscat
	€ 45,00	Merlot
	€ 50,00	Sauvignon
	€ 51,00	Friuli
	€ 52,00	White Bordeux
	€ 48,00	Chateau d'Yquem

Store	Store-Id	City
	C1	Rome
	C2	Paris
	C3	Moscow

Sale	Sale-Id	Bottle-Id	Store-Id	Quantity	Date
	S1	Red Bordeaux	C1	20	09-Dec-99
	S2	Medit.Muscat	C1	14	09-Dec-99
	S2	Medit.Muscat	C1	16	09-Dec-99
	S3	Merlot	C1	40	09-Dec-99
	S4	Merlot	C2	100	09-Dec-99
	S5	Sauvignon	C2	120	09-Dec-99
	S5	Sauvignon	C2	80	09-Dec-99
	S6	Friuli	C2	200	09-Dec-99
	S7	White Bordeux	C3	600	12-Dec-07
	S8	Merlot	C3	1000	12-Dec-07
	S9	Medit.Muscat	C3	440	12-Dec-07
	S9	Medit.Muscat	C3	360	12-Dec-07

Figure 4**:** Sample of a Star Schema

After forming the structure and storing it as a concept relation in Oracle10g, we perform the calculation of the hierarchical closure of the H-IFS and its weights.
The user now has the choice of selecting three different strategies: ***Optimistic, Pessimistic*** or ***Average*** as defined on section 2.2.
Let's assume that the user's interest lays on finding information about Red wines. Fig.5 below shows the hierarchy after weights have been estimated and assigned reflecting the user's intent.

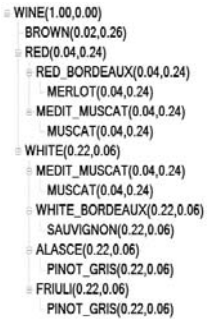


Figure 5: Vitis Vinifera sub-hierarchy view with weights

We can observe that the principle of the H-IFS closure *(see definition 3)* has been preserved when propagating the degree of validity μ and non-validity ν from super-elements to sub-elements by using the optimistic strategy.
The degree of validity and non-validity <μ, ν > are calculated as follows:

$$\mu = \frac{|c_l|}{|c_{l-1}|} \qquad \nu = \frac{|\neg c_l|}{|c_{l-1}|}$$

Where c_l corresponds to those elements from the fact table that absolutely satisfy the selection criteria with reference to a node in the hierarchy. C_{l-1} represents the elements children elements of that selection on a lower level that satisfy the selection condition to some extent. It is obvious that $\pi = 1-(\mu+\nu)$.
After adding the hierarchy into the repository and automatically calculating the weights for the requested nodes, the user can utilize the ad-hoc interface for execution of queries either in standard SQL or make use of the enhanced Select clause and features that IF-Oracle provides.

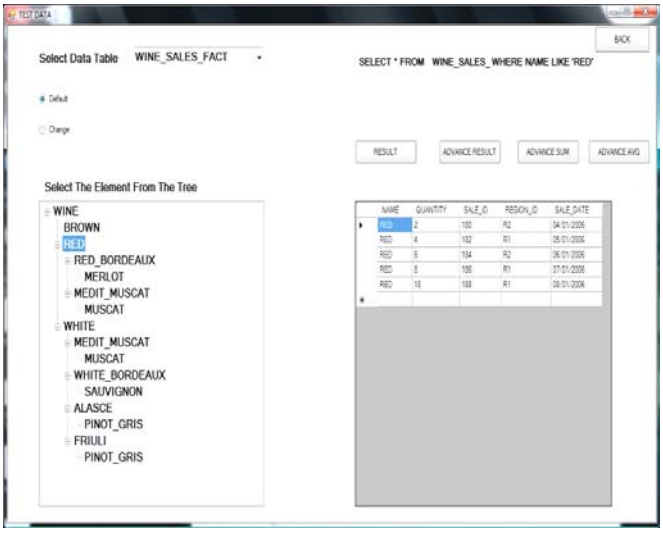


Figure 6**:** Standard SQL output for "Red" wine

Fig. 6 shows the results of a user request for "Red" wine executed in standard SQL provided by Oracle10g.
In contrast, Fig. 7 shows the output after executing the same query, but this time using the IF-Oracle utility.

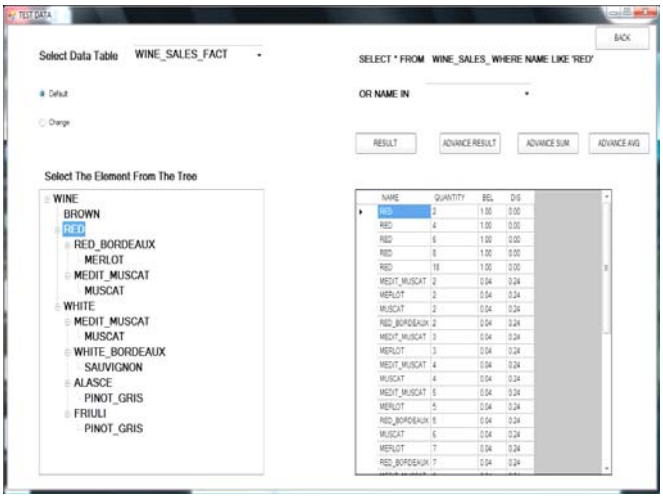


Figure 7: Enhanced SQL output for "Red" wine

By comparing the two figures, one can observe that IF-Oracle produces a knowledge-based answer instead of mindlessly matching the records against the word "Red".
The results show that IF-Oracle not only retrieves sales of "Red" bottles, but also sales of bottles that are classified as red wines by the knowledge represented in the H-IFS hierarchy as "Merlot", "Red Bordeaux", "Medit. Muscat", etc. with indicative degrees of <μ, ν> relevant to the user's preference.
At this point one may decide to further enhance the query capabilities of the IF-Oracle utility by allowing versions of hierarchies. In such case similarities [20] and dissimilarities [21] between different versions should be reflected in the query results, or in the common acceptable DWH schema.

ISBN: 978-989-95079-6-8

5 Conclusions & Comparisons

In this paper, we focus on integrating hierarchical users preferences or intent in OLAP queries with the aim on enhancing the OLAP scope and in return to get richer answer, closer to user requests. We provide a means of using background knowledge to re-engineer query processing and answering with the aid of H-IFS and Intuitionistic Fuzzy relational representation.

The hierarchical links defined on the basis of the H-IFS closure are representing hierarchical knowledge in different forms. A cubic OLAP model presented that allows dimensions to be defined as H-IFS.

In comparing our framework with the bipolar querying school of thought as presented in [22], [23], and [24] the reader may take into account the following:

- The bipolar querying school of thought is the main vehicle for allowing users to express their preferences as part of a query formulation. Allowing thus queries to come up with more knowledgeable answers.

However our framework is resolving a different problem, related specifically to OLAP in the following ways:

- It allows users to import knowledge from external sources in order to redefine the axis of analysis and eventually the cubic structure for different business scenarios.
- The IF-Oracle utility can operate using either cubic structures-MOLAP or relations-ROLAP.
- The IF-Oracle utility assumes that stored data are precise. However it can also cope with imprecise data. The importance here is not on the goodness of data, is mainly on how well the stored data do fit in different scenarios or changing dimensions.

Future research efforts will concentrate on incorporating knowledge arriving from external sources either semi structured or unstructured, considering the web as such source.

References

[1] S. Abiteboul, R. Agrawal, P. Bernstein, M. Carey, S. Ceri, B. Croft, D. Dewitt, M. Franklin, H. Garcia Molina, D. Gawlick, J.Gray, L. Haas, A. Halevy, J. Hellerstein, Y. Ioannidis, M. Kersten, M. Pazzani, M. Lesk, D. Maier, J. Naughton, H. Schek, T. Sellis, A. Silberschatz, M. Stonebraker, R. Snodgrass, J. Ullman, G. Weikum, J. Widom, S. Zdonik, "The Lowell Database Research Self Assessment", Communications of the ACM, 2005, Volume 48, Number 5, pp 111-118.

[2] T.B. Pedersen, C.S. Jensen, Multidimensional "Data Modeling for Complex Data", In Proc. of 15th ICDE, IEEE Computer Society, 1999, pp.336-345

[3] J. Gray, A. Bosworth, A. Layman, H. Pirahesh, "Data cube: a relational aggregation operator generalizing group-by, cross-tabs and subtotals", Journal of Data Mining and Knowledge Discovery, Springer Verlag, Vol.1, No1, 1997, pp 29--53.

[4] M. Delgado, C. Molina, D. Sanchez, A. Vila, L. Rodriguez-Ariza, "A fuzzy multidimensional model for supporting imprecision in OLAP", In Proc. of FUZZ-IEEE, 2004, vol.3, pp 1331- 1336

[5] P. Chamoni, S. Stock, "Temporal Structures in Datawarehousing", In Proc. of the DaWaK'99, ACM Portal, 1999 pp. 353 - 358

[6] T.B. Pedersen, C.S. Jensen, C.E. Dyreson, "A foundation for capturing and querying complex multidimensional data", In Information Systems, ACM Portal, Vol. 26, No. 5, 2001, pp. ages: 383 - 423

[7] A.O. Mendelzon, A. Vaisman, "Temporal Queries in OLAP", In Proc. of the 26th VLDB'00, 2000. ACM Portal, pp. 242 -253

[8] E. Rogova, P. Chountas, K. Atanassov, "Flexible hierarchies and fuzzy knowledge-based OLAP". – FSKD 2007, IEEE Computer Society Press Vol.2, pp. 7-11

[9] P. Chountas, "On Intuitionistic fuzzy sets over universes with hierarchical structures" Notes on Intuitionistic Fuzzy Sets, Vol. 13, No 1, pp 52-56

[10] K. Attanasov, "Intuitionistic Fuzzy Sets: Theory and Applications", Springer-Verlag, 1999

[11] K. Atanassov , "Remarks on the Intuitionistic fuzzy sets" – Fuzzy Sets and Systems, Vol. 51, 1992, No 1, 117-118

[12] B. Kolev, P. Chountas, E. Rogova, K. Atanassov "Representation of Value Im-perfection with the Aid of Background Knowledge: H-IFS" Studies in Computational Intelligence , Springer Verlag, Vol.109, pp 473-494.

[13] P. Chountas, K Atanassov, E. Rogova, "H-IFS: Modelling & Querying over Hierarchical Universes", International Conference On Information Processing And Management of Uncertainty-IPMU'2008, ISBN 978-84-612-3061-7, 2008, pp 1628-1634.

[14] E. Rogova, P. Chountas, "On Imprecision Intuitionistic Fuzzy Sets & OLAP – The Case for KNOLAP", IFSA'07, Springer-Verlag GmBh , Theoretical advances and application of fuzzy logic and soft computing, 2007, pp. 11-20

[15] R. Kimball, "The Data Warehouse Toolkit". New York: John Wiley & Sons, 1996.

[16] S. Chaudhuri, U. Dayal, V.Ganti, "Database Technology for Decision Support Systems". In: Computer, Vol. 34, p. 48-55, 2001

[17] M. Jarke et al. "Fundamentals of data warehouses" Springer, London, 2002

[18] C. Dyreson, "Information retrieval from an incomplete data cube", VLDB, Morgan Kaufman Publishers, 1996, pp. 532-543.

[19] H. Thomas & A. Datta, "A Conceptual Model and Algebra for On-Line Analytical Processing in Decision Support Databases", Information Systems Research 12: 2001, pp.83-102,

[20] E. Szmidt, J. Kacprzyk "Distances between Intuitionistic fuzzy sets, Fuzzy Sets and Systems", Vol. 114, No.3, 2000, pp505–518.

[21] E. Szmidt, J. Kacprzyk "A new concept of similarity measure for Intuitionistic fuzzy sets and its use in group decision making" LNAI 3558, 2005, pp272–282

[22] H. Prade, D. Dubois. "Handling bipolar queries in fuzzy information processing". In: Fuzzy information processing in databases. José Galindo (Eds.), Vol. 1, Information Science Reference, Hershey, PA. , 2008, pp. 97-114.

[23] S. Zadrozny "Bipolar queries revisited" In: V. Torra, et al. (Eds.): Modeling Decisions for Artificial Intelligence, LNCS, Springer Verlag Vol. 3558, 2005 pp. 387-398.

[24] S. Zadrozny J. Kacprzyk "Bipolar queries and queries with preference" Proc. 17th International Conference DEXA'06, 2006 pp. 415-419.

ISBN: 978-989-95079-6-8

Modeling Investor Optimism with Fuzzy Connectives

Milan Lovric, Rui J. Almeida, Uzay Kaymak[1] Jaap Spronk[2]

[1] Erasmus School of Economics, Erasmus University Rotterdam
Rotterdam, The Netherlands
[2] Rotterdam School of Management, Erasmus University Rotterdam
Rotterdam, The Netherlands
Email: {lovric,almeida,kaymak}@ese.eur.nl, jspronk@rsm.nl

Abstract— *Optimism or pessimism of investors is one of the important characteristics that determine the investment behavior in financial markets. In this paper, we propose a model of investor optimism based on a fuzzy connective. The advantage of the proposed approach is that the influence of different levels of optimism can be studied by varying a single parameter. We implement our model in an artificial financial market based on the LLS model. We find that more optimistic investors create more pronounced booms and crashes in the market, when compared to the unbiased efficient market believers of the original model. In the case of extreme optimism, the optimistic investors end up dominating the market, while in the case of extreme pessimism, the market reduces to the benchmark model of rational informed investors.*

Keywords— artificial financial market, agent-based modeling, behavioral finance, fuzzy aggregation, investor sentiment, optimism index.

1 Introduction

Artificial financial markets are models for studying the link between individual investor behavior and financial market dynamics. They are often computational models of financial markets, and are usually comprised of a number of heterogeneous and boundedly rational agents, which interact through some trading mechanism, while possibly learning and evolving. These models are built for the purpose of studying agents' behavior, price discovery mechanisms, the influence of market microstructure or the reproduction of the stylized facts of real-world financial time-series (e.g. fat tails of return distributions and volatility clustering).

A similar bottom-up approach has been utilized in agent-based computational economics (ACE) - the computational study of economies modeled as evolving systems of autonomous interacting agents [22]. A methodology analogous to agent-based modeling also comes from the physical sciences as the Macroscopic Simulation - a tool for studying complex systems by simulating many interacting microscopic elements [16]. A number of reviews of studies with artificial financial markets are available, e.g. [15] and [10].

Since agent-based models can easily accommodate complex learning behavior, asymmetric information, heterogeneous preferences, and ad hoc heuristics [4], such simulations are particularly suitable to test and generate various behavioral hypothesis. The idea of individual investors who are prone to biases in judgment, who are frame-dependent and use various heuristics, which might lead to anomalies on the market level, has been explored within the field of behavioral finance. Behavioral finance is the branch of finance which applies knowledge from psychology and sociology to discover and explain phenomena inconsistent with the paradigm of the expected utility of wealth and narrowly defined rational behavior [7]. A number of surveys and books on behavioral finance and behavioral economics topics can be found, for example [20], [8], and [2]. This complementarity of behavioral finance research and the agent-based methodology has been recognized in the literature (e.g. [15], [5]). Some of the early studies that pursue the idea of explicit accounting for behavioral theories in agent-based financial market simulations are [21] and [9]. In [21] the focus is on overconfidence and loss aversion, while [9] study the social interaction between investors.

One of the key characteristics that govern investor behavior is the optimism or pessimism of the investors. The link between asset valuation and investor sentiment has been the subject of considerable debate in the finance, and has been studied in the context of mispricing (departures from the fundamentals) [3], the limits of arbitrage [17], as well as the underreaction and overreaction of stock prices [1]. Two methodological approaches can be found in the finance literature. One is concerned with finding adequate proxies for the aggregate investor sentiment, and using them in statistical analysis to explain the variation of stock prices and the occurrences of mispricing, such as bubbles and crashes. The other one is a bottom-up approach that aims at modeling individual investor optimism and pessimism by using the insights from psychological theories. For these theories, it is important to have a flexible framework that can be adapted to capture the complexity of human decision making behavior.

In fuzzy decision theory, a wide range of connectives (aggregation operators) has been proposed and studied in order to model the flexibility of human decision making. In this sense, the use of fuzzy connectives for modeling elements of behavioral finance is promising, since the wide range of behaviors documented in the behavioral finance literature necessitates the use of a flexible framework for aggregating information. In this paper, we make a step in this direction by proposing a model of investor optimism based on fuzzy aggregation.

In probabilistic decision theory, such as the Prospect Theory [12, 23] and Rank-Dependent Utility Theory [19], optimism and pessimism are modeled using the probability weighting function. If, for example, the decision under risk is considered, a decision problem is presented using risky prospects, i.e. a set of possible outcomes and their probabilities. Because of the probability weighting, the decision weights asso-

ISBN: 978-989-95079-6-8

ciated with the outcomes are not equal to their probabilities (as would be in the case of Expected Value Theory or Expected Utility Theory). To model optimism we would need to specify and parameterize such a probability weighting function that gives more decision weight to good outcomes and less decision weight to bad outcomes. However, an empirically observed probability weighting function is usually S-shaped, which means that when dealing with such prospects, people are at the same time optimistic about the best outcomes, pessimistic about the worst outcomes, and insensitive to middle outcomes [25].

A decision maker's optimism or pessimism has also been studied within a fuzzy decision making setting. Various fuzzy connectives studied in this context have parameters that denote explicitly the optimism or pessimism degree of a decision maker. Apart from the well-known Hurwicz operator [11], the grade of compensation in Zimmermann–Zysno operator [26] can also be interpreted as an index of optimism. All these operators view the decision as a mixture of conjunctive and disjunctive behavior, and the degree of optimism determines which aggregation type dominates and to which degree. Another optimism–pessimism index was proposed in [24], where the parameter of the generalized averaging operator [6] is interpreted as the decision maker's characteristic degree of optimism. This is an intuitive way of modeling degree of optimism, since optimism is now modeled as the disposition of the decision maker to believe or give importance to positive events compared to his/her disposition to consider negative events [14]. An application of this operator in the risk management of power networks has been considered in [13].

In this paper, we propose a model of investor optimism based on the generalized averaging operator. The advantage of the proposed approach is that the influence of different levels of optimism can be studied by varying a single parameter. We study the effects of investor optimism in an artificial financial market based on the Levy, Levy, Solomon (LLS) model [16]. In previous publications, investor psychological biases such as overconfidence have also been studied by using this model [18].

The outline of the paper is as follows. Section 2 explains the basics of the LLS model in which we study the investor optimism and pessimism. Section 3 describes the setup of the experiments we have conducted. Section 4 presents the results of the simulations. Section 5 concludes the paper and discusses possible extensions for the future research.

2 Model Description

The proposed model of investor optimism is based on the LLS microscopic simulation model [16] with a small homogeneous subpopulation of efficient market believers (EMBs) as described in [16]. LLS model is a well-known and early econophysics model, rooted in a utility maximization framework. Variants of the model have been published in a number of articles and a book, and the model has also been critically evaluated in [27].

2.1 Asset Classes

As in the original LLS model, there are two investments alternatives: a risky stock (or market index) and a risk-free asset (bond). This is in line with many of the agent-based artificial financial markets, which typically do not deal with portfolio selection in multi-asset environments. The risky asset pays at the beginning of each period a dividend which follows a multiplicative random walk according to

$$\tilde{D}_{t+1} = D_t(1+\tilde{z}), \tag{1}$$

where $\tilde{z}$ is a random variable distributed uniformly in the interval $[z_1, z_2]$. The bond pays interest with a rate of r_f.

2.2 Agent Behavior

Many early agent-based artificial financial markets were based on a small number of relatively simple strategies. Such markets have been labeled as *few-type models* [15]. Typically, strategies (or agents who employ them) could be divided into two groups: *fundamental* (based on a perceived fundamental value) and *technical* (based on the past prices, e.g. some form of trend extrapolation). *Zero-intelligence* framework in which agents trade randomly, might be useful for to studying the influence of market microstructure, and sometimes a small number of such agents is included into a few type model to provide liquidity for other agents.

LLS model follows a standard framework where preferences (and risk attitude) are captured by an agent's utility function, and the objective is the maximization of expected utility. But even in such a framework there are many possibilities for the functional form of the utility, which differ in descriptive validity and analytical tractability. When empirical support is taken into account, most evidence suggests DARA (Decreasing Absolute Risk Aversion) and CRRA (Constant Relative Risk Aversion), which motivates the choice of power (myopic) utility function in [16]

$$U(W) = \frac{W^{1-\alpha}}{1-\alpha}. \tag{2}$$

LLS model contains two types of investors: (1) Rational Informed Investors (RII) and (2) Efficient Market Believers (EMB).

2.2.1 RII investors

RII investors know the dividend process, and therefore can estimate fundamental value as the discounted stream of future dividend, according to the Gordon model

$$P^f_{t+1} = \frac{D_t(1+\tilde{z})(1+g)}{k-g}, \tag{3}$$

where k is the discount factor of the expected rate of return demanded by the market for the stock, and g is the expected growth rate of the dividend. RII investors assume that the price will converge to the fundamental value in the next period. In each period RII investor i chooses the proportion of wealth to invest in stocks and bonds so that he or she maximizes the expected utility of wealth in the next period, given by the following equation from [16]:

$$\begin{aligned} EU(\tilde{W}^i_{t+1}) &= \frac{(W^i_h)^{1-\alpha}}{(1-\alpha)(2-\alpha)} \frac{1}{(z_2-z_1)} \left(\frac{k-g}{k+1}\right) \frac{P_h}{xD_t} \\ &\times \left\{ \left[(1-x)(1+r_f) + \frac{x}{P_h}\left(\frac{k+1}{k-g}\right) D_t(1+z_2)\right]^{(2-\alpha)} \right. \\ &\left. - \left[(1-x)(1+r_f) + \frac{x}{P_h}\left(\frac{k+1}{k-g}\right) D_t(1+z_1)\right]^{(2-\alpha)} \right\}. \end{aligned} \tag{4}$$

ISBN: 978-989-95079-6-8

Based on the optimal proportion, they determine the number of stocks demanded by multiplying this optimal proportion with their wealth. Since all RII investors are assumed to have the same degree or risk aversion (parameter α), they will all have the same optimal proportion x. The actual number of demanded shares might differ only if investors differ in their wealth. However, as in the experiments of [16] we assume that they all start with the same initial wealth.

2.2.2 EMB investors

EMB investors believe that the price accurately reflects the fundamental value. However, since they do not know the dividend process, they use *ex post* distribution of stock returns to estimate the *ex ante* distribution. EMB investor i uses a rolling window of size m^i, and is in the original model [16] said to be *unbiased* if, in absence of additional information, he or she assigns the same probability to each of the past m^i return observations [16]. Hence, the original, unbiased EMBs assume that returns come from a discrete uniform distribution

$$\Pr^i(\tilde{R}_{t+1} = R_{t-j}) = \frac{1}{m^i}, \text{ for } j = 1, ..., m^i. \tag{5}$$

The expected utility of EMB investor i is given by [16]

$$EU(\tilde{W}^i_{t+1}) = \frac{(W^i_h)^{1-\alpha}}{(1-\alpha)} \sum_{j=1}^{m^i} \Pr^i(\tilde{R}_{t+1} = R_{t-j}) \times \left[(1-x)(1+r_f) + xR_{t-j}\right]^{(1-\alpha)}. \tag{6}$$

In accordance with the LLS model [16], for all EMB investors an investor specific noise is added to the optimal investment proportion x^* (that maximizes the expected utility) in order to account for various departures from rational optimal behavior ($\tilde{\varepsilon}^i$ is truncated so that $0 \leq x^i \leq 1$, imposing the constraint of no borrowing and no short-selling), i.e.

$$x^i = x^{*i} + \tilde{\varepsilon}^i. \tag{7}$$

2.2.3 Sentiment EMBs

In this paper we create a new EMB type, called the *sentiment EMBs* by using a fuzzy set connective. *Sentiment EMBs* use generalized aggregation operator to estimate future returns, using the rolling window of size m^i. The prediction of the next period return for each investor i is given by

$$\tilde{R}_{t+1} = \left(\frac{1}{m^i} \sum_{j=1}^{m^i} (R_{t-j})^s \right)^{1/s}. \tag{8}$$

The higher the parameter s, the higher the estimate of the return (more closer to the maximum value from the sample), and vice versa. In such a way, we use parameter s to capture the phenomena of investor optimism and pessimism.

In our experiments we consider a several special cases of the generalized mean:

- $s \to -\infty$, the minimum of the sample;
- $s = -1$, the harmonic mean;
- $s \to 0$, the geometric mean;
- $s = 1$, the arithmetic mean;
- $s = 2$, the quadratic mean;
- $s \to \infty$, the maximum of the sample.

Since there is only one value for the expected return, instead of a probability distribution, the expected utility of sentiment EMB investor i is given by

$$EU(\tilde{W}^i_{t+1}) = \frac{(W^i_h)^{1-\alpha}}{(1-\alpha)} \left[(1-x)(1+r_f) + x\tilde{R}_{t+1}\right]^{(1-\alpha)}. \tag{9}$$

The investors will maximize this expected utility if in each period they invest all their wealth either in the stock or in the bond, depending on the actual comparison between the expected return on the stock $\tilde{R}_{t+1}$ and the return on the riskless bond $(1 + r_f)$.

2.3 Market Mechanism

LeBaron in [15] describes four types of market mechanisms used in agent-based artificial financial markets. In this paper, as in the original LLS model, we use clearing by *temporary market equilibrium*. RII and EMB investors determine optimal proportion in the stock so as to maximize the expected utility of their wealth in the next period. However, expected utility is the function of the future price, which is in the current period unknown. Investors therefore need to determine optimal proportions, and respective demands for shares, for various hypothetical prices. The equilibrium price P_t is set to that hypothetical price for which the total demand of all investors in the market equals the total number of outstanding shares, according to

$$\sum_i N^i_h(P_t) = \sum_i \frac{x^i_h(P_t) W^i_h(P_t)}{P_t} = N. \tag{10}$$

Table 1: Parametrization

Symbol	Value	Explanation
M	950	Number of RII investors
M_2	50	Number of EMB investors
m	10	Memory length of EMB investors
α	1.5	Risk aversion parameter
N	10000	Number of shares
r_f	0.01	Riskless interest rate
k	0.04	Required rate of return on stock
z_1	-0.07	Maximal one-period dividend decrease
z_2	0.10	Maximal one-period dividend growth
g	0.015	Average dividend growth rate

3 Experiments with investor optimism

In the benchmark model where only RII investors are present in the market, there is no trade, the log prices follow random walk, and there is no excess volatility of the market price [16]. In the experiment with a small fraction of homogeneous (with respect to memory length) and unbiased EMB investors (of the original model), the market dynamics show semi-predictable (unrealistic) booms and crashes, with substantial trading in the

ISBN: 978-989-95079-6-8

market and excess volatility [16]. This experimental setup of [16] is also the basis for the experiments in this paper.

In our new model we conduct six experiments for six different levels of optimism of EMB investors, that correspond to the special cases of the parameter s. In each experiment the market consists of 95% RII investors and 5% EMB investors, with the parametrization given in Table 1. We run 100 independent 1000-period-long simulations, with different initial seeds of the random number generators. The results in the Table 2 are averaged over these 100 simulations.

Table 2: Results

	$s=-\infty$	$s=-1$	$s=0$
$\sigma(P)$	6.0249	12.8370	17.8668
$\sigma(P^f)$	5.7159	5.7159	5.7159
excess volatility %	5.41	124.59	212.58
mean volume p.p. %	0.48	9.04	6.40
	$s=1$	$s=2$	$s=\infty$
$\sigma(P)$	27.4739	28.8751	25.0327
$\sigma(P^f)$	5.7159	5.7159	5.7159
excess volatility %	380.66	405.18	337.95
mean volume p.p. %	2.82	1.18	0.12

4 Results

Fig. 1 shows a typical price dynamics from the first experiment with pessimistic EMB investors. The market price closely follows the fundamental price which is driven by the random dividend process. Hence, this experiment resembles the benchmark model in which there are only RII investors in the market. Pessimistic investors predict next period return with the minimum return in the sample of past returns. The minimum return is almost always bellow the risk-less return, so the optimal investment for pessimistic EMB investors is to invest everything in bond. The actual investment proportion will slightly vary due to the error term in (7). Only in rare occasions when there is a series of returns higher than the risk-less return, the EMB investors will invest in the risky asset. The results in Table 2 show that for this experiment the volatility of the market price is similar to the volatility of the fundamental price, which means that there is a low excess volatility. The relative mean volume per period shows that there is very little trading in the market, i.e. from period to period the investors do not change much their portfolio holdings.

Fig. 2 shows the price development for the second experiment with slightly more optimistic investors that predict future return using the harmonic mean. The results of this experiment qualitatively and quantitatively resemble the results of the original model with a small fraction of unbiased EMB investors (which predict future returns using a uniform discrete distribution over the observed returns). The market exhibits cyclical booms and crashes to the fundamental value. According to Table 2, the market is more volatile, and there is also more trading. This exchange of risky assets between RII and EMB investors occurs mostly when the booms begin and when they crash.

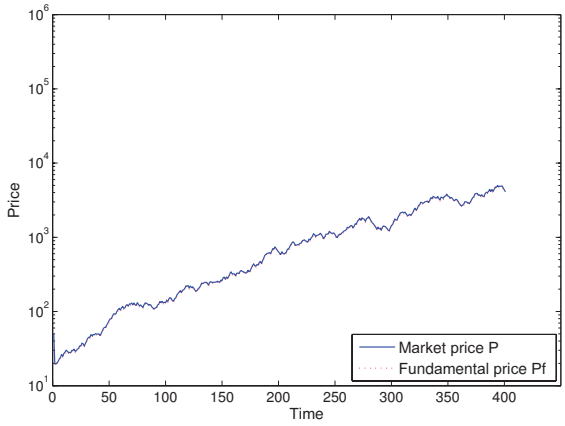


Figure 1: Price dynamics with 95% RII and 5% minimum sentiment EMB.

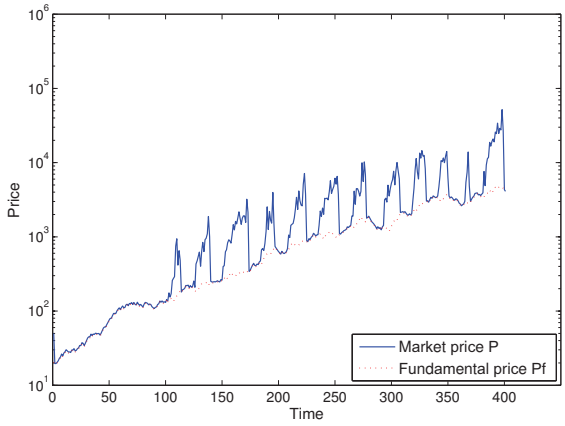


Figure 2: Price dynamics with 95% RII and 5% harmonic sentiment EMB.

Fig. 3, Fig. 4, and Fig. 5 depict the market dynamics when EMB investors are even more optimistic. As the index of optimism increases the market shows more extreme (longer lasting) booms, followed by very sharp crashes. During these bubbles the EMB investors aggressively invest in the risky asset, while the RII investor divest expecting that the overvalued asset would fall to its fundamental value. The crash occurs when there is a series of low returns, due to low dividend realizations, so the EMB investors suddenly shift toward a risk-less asset. However, as soon as a better return is realized, EMB investors invest in the risky asset and a new boom starts. From Table 2 it is also evident that the more optimistic EMB investors are, the more volatile market price is. However, the trading is reduced because the booms are longer lasting, i.e. the cycles of booms and crashes appear less frequently.

In the case of full optimism, there is an ongoing market bubble, as shown in Fig. 6. The market does not crash because the maximum return in the rolling window of past returns is always above the risk-less return, so the EMB investors are always highly invested in the risky asset. The trading in this experiment is even more reduced, but the volatility of the market price is also somewhat reduced. The reason for the latter is that the crash does not occur within the experiments.

ISBN: 978-989-95079-6-8

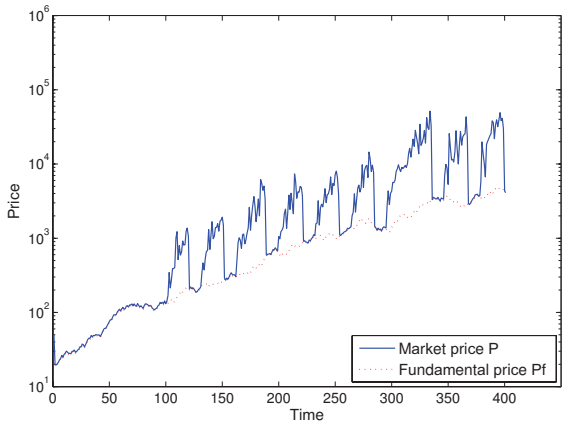


Figure 3: Price dynamics with 95% RII and 5% geometric sentiment EMB.

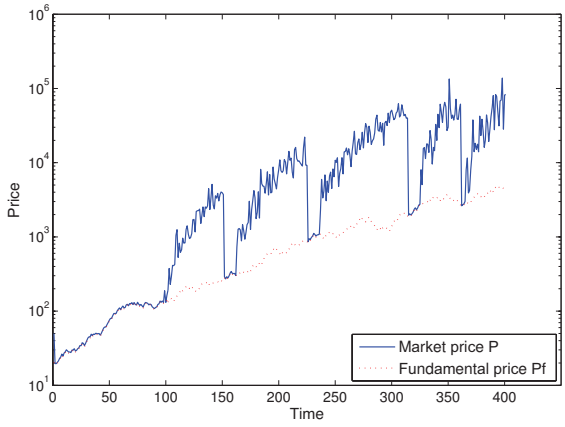


Figure 4: Price dynamics with 95% RII and 5% arithmetic sentiment EMB.

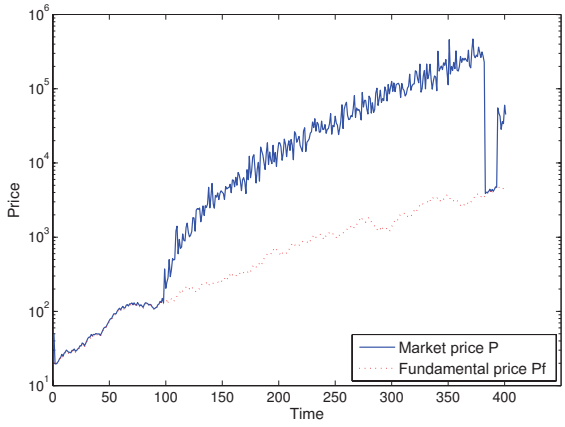


Figure 5: Price dynamics with 95% RII and 5% quadratic sentiment EMB.

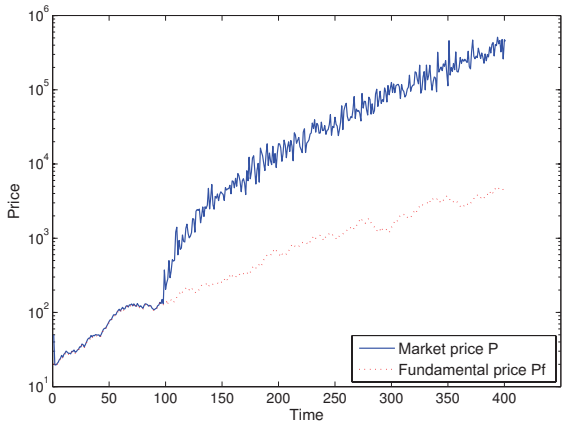


Figure 6: Price dynamics with 95% RII and 5% maximum sentiment EMB.

Fig. 7 shows the development of the relative wealth of RII investors over time. At the beginning, RII investors posses 95% of all the wealth in the market. In the case of extremely pessimistic EMB investors, RII investors end up asymptotically dominating the market. This is because the LLS market is a growing market, and only RII investors are investing in the risky asset and exploiting that growth. Conversely, in the case of extreme optimism, EMB investors are highly invested in the stock, and eventually dominate the market. In non-extreme cases of optimism, both types of investor coexist in the market.

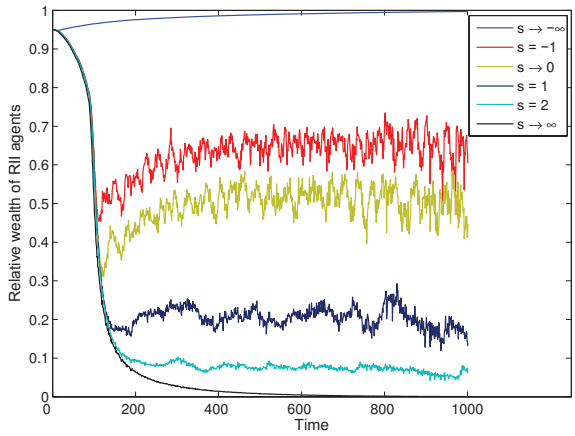


Figure 7: Relative wealth dynamics of RII against sentiment EMBs with various levels of optimism.

5 Discussion and conclusions

In this paper we have used a fuzzy connective to study investor optimism in the modified LLS model of the stock market [16]. We show how changes in the formation of expectations by EMB investors can have a marked impact on the price dynamics. The levels of investor optimism are related to the occurrences of market booms and crashes, as well as measures of excess volatility and trading volume.

ISBN: 978-989-95079-6-8

Since our current experiments focus only on the case of homogeneous EMB investors, we would like to conduct our next experiments in the case of heterogeneous EMB investors with various memory lengths. We expect that the choice of memory length has a great impact on the occurrence of booms and crashes, particularly in extreme cases of optimism, because the larger that window of past returns is, the less likely is that all the returns are bellow or above the risk-less return. In future research, we will extend our analysis to the interplay of investors with various degrees of optimism within the same market. Furthermore, we would like to implement an updating mechanism by which the level of investor optimism changes based on the past performance.

As we have used the same model to study investor overconfidence [18], a distinct although related behavioral phenomenon, it would also be interesting to study both phenomena at the same time. The overconfidence in the model [18] refers to the peakedness of the return distribution around the mean of return observations, while optimism in this model determines how that mean is chosen (ranging from the minimum observation to the maximum observation in the sample of past returns).

This paper demonstrates the advantage of using a fuzzy connective for modeling investor optimism, as we were able to control investor optimism by varying only a single parameter. The results of our experiments show that this parameter was a valid choice for an index of optimism in the context of financial markets. In future research other fuzzy set connectives could be investigated for agent decision making.

References

[1] N. Barberis, A. Shleifer, and R. Vishny. A model of investor sentiment. *Journal of Financial Economics*, 49(3):307–343, 1998.

[2] N. C. Barberis and R. H. Thaler. A survey of behavioral finance. In G. M. Constantinides, M. Harris, and R. M. Stulz, editors, *Handbook of the Economics of Finance*, volume 1, pages 1053–1128. Elsevier, 1 edition, 2003.

[3] G. W. Brown and M. T. Cliff. Investor sentiment and asset valuation. *Journal of Business*, 78(2):405–440, 2005.

[4] N. T. Chan, B. LeBaron, and T. Poggio. Agent-based models of financial markets: A comparison with experimental markets. MIT Artificial Markets Project, 1999.

[5] S.-H. Chen and C.-C. Liao. Behavioral finance and agent-based computational finance: toward an integrating framework. In *Proceedings of 7th Joint Conference on Information Sciences*, pages 1235–1238, North Carolina, USA, September 2003.

[6] H. Dyckhoff and W. Pedrycz. Generalized means as model of compensative connectives. *Fuzzy Sets and Systems*, 14(2):143–154, 1984.

[7] G. M. Frankfurter and E. G. McGoun. Market efficiency or behavioral finance: The nature of the debate. *The Journal of Psychology and Financial Markets*, 1(4):200–210, 2000.

[8] J. Goldberg and R. von Nitzsch. *Behavioral Finance*. John Wiley and Sons Ltd, 2001.

[9] A. O. I. Hoffmann, W. Jager, and E. J. H. V. Eije. Social simulation of stock markets: Taking it to the next level. *Journal of Artificial Societies and Social Simulation*, 10(2), 2007.

[10] C. H. Hommes. Heterogeneous agent models in economics and finance. In L. Tesfatsion and K. L. Judd, editors, *Handbook of Computational Economics*, volume 2, chapter 23, pages 1109–1186. Elsevier, 1 edition, 2006.

[11] L. Hurwicz. A class of criteria for decision-making under ignorance. Cowles Commission Discussion Paper Statistics No. 356, Cowles Foundation for Research in Economics, Yale University, Feb. 1951.

[12] D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979.

[13] U. Kaymak, R. Babuska, H. R. van Nauta Lemke, and G. Honderd. A fuzzy logic system for steady-state security analysis of power networks. *Journal of Intelligent and Fuzzy Systems*, 6(3):345–359, 1998.

[14] U. Kaymak and H. R. van Nauta Lemke. A sensitivity analysis approach to introducting weight factors into fuzzy multicriteria decision making. *Fuzzy Sets and Systems*, 97(2):169–182, 1998.

[15] B. LeBaron. Agent-based computational finance. In L. Tesfatsion and K. L. Judd, editors, *Handbook of Computational Economics*, volume 2, chapter 24, pages 1187–1233. Elsevier, 1 edition, 2006.

[16] M. Levy, H. Levy, and S. Solomon. *Microscopic Simulation of Financial Markets: From Investor Behavior to Market Phenomena*. Academic Press, 2000.

[17] J. B. D. Long, A. Shleifer, L. H. Summers, and R. J. Waldmann. Noise trader risk in financial markets. *Journal of Political Economy*, 98(4):703–738, 1990.

[18] M. Lovric, U. Kaymak, and J. Spronk. Overconfident investors in the LLS agent-based artificial financial market. In *Proceedings of the IEEE Symposium on Computational Intelligence for Financial Engineering (CIFEr)*, pages 58–65, Nashville, USA, March–April 2009.

[19] J. Quiggin. A theory of anticipated utility. *Journal of Economic Behavior and Organization*, 3(4):323–343, 1982.

[20] H. Shefrin. *Beyond Greed and Fear - Understanding Behavioral Finance and the Psychology of Investing*. Harvard Business School Press, 2000.

[21] H. Takahashi and T. Terano. Agent-based approach to investor's behavior and asset price fluctuation in financial markets. *Journal of Artificial Societies and Social Simulation*, 6(3), 2003.

[22] L. Tesfatsion and K. Judd. Agent-based computational economics: A constructive approach to economic theory. In L. Tesfatsion and K. L. Judd, editors, *Handbook of Computational Economics*, volume 2, chapter 16, pages 831–880. Elsevier, 1 edition, 2006.

[23] A. Tversky and D. Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323, 1992.

[24] H. R. van Nauta Lemke, J. G. Dijkman, H. van Haeringen, and M. Pleeging. A characteristic optimism factor in fuzzy decision-making. In *Proceedings IFAC Symposium on Fuzzy Information, Knowledge Representation and Decision Analysis*, pages 283–288, Marseille, France, July 1983.

[25] H. U. Weber. From subjective probabilities to decision weights: The effect of asymmetric loss functions on the evaluation of uncertain outcomes and events. *Psychological Bulletin*, 115(2):228–242, 1994.

[26] H.-J. Zimmermann and P. Zysno. Latent connectives in human decision making. *Fuzzy Sets and Systems*, 4:37–51, 1980.

[27] E. Zschischang and T. Lux. Some new results on the Levy, Levy and Solomon microscopic stock market model. *Physica A*, 291:563–573, 2001.

ISBN: 978-989-95079-6-8

Tail Point Density Estimation Using Probabilistic Fuzzy Systems

R. J. Almeida U. Kaymak

Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
Email: {rjalmeida,kaymak}@ese.eur.nl

Abstract— *Value at Risk (VaR) is a popular measure for quantifying the market risk that a financial institution faces into a single number. Due to the complexity of financial markets, the risks associated with a portfolio may vary over time. For accurate VaR estimation, it is necessary to have flexible methods that adapt to the underlying data distribution. In this paper, we consider VaR estimation by using probabilistic fuzzy systems (PFS). Contrary to previous publications, our focus is on the modeling of the tail points of the distribution of returns. We study two approaches to designing probabilistic fuzzy VaR models that take into account the extreme values of the data and compare their performance with the performance of a GARCH model. It is found that the VaR estimation process is simplified and improved by our proposed method.*

Keywords— Density estimation, extreme values, fuzzy histograms, probabilistic fuzzy systems, value-at-risk.

1 Introduction

Due to the volatile nature of the financial markets risk management is an important activity for financial institutions that operate in these markets. As a result of risk management, activities are undertaken to reduce the possibility of failure to an acceptable range. These activities may include portfolio adjustment, hedging or insurance [1, 2]. Nowadays, the financial sector operates under strict guidelines, which have been imposed through international agreements, partly due to various financial failures that have happened in 1990's. For example, due to the Basel Agreement, the financial institutes must have well documented procedures to manage the market risk, the credit risk and the operational risk that they are exposed to [3].

Managing risk is strongly dependent on the information available. When the amount of information grows beyond a specific level, there is a need for a concise representation of the risk a company or institution is facing. Due to the complex nature of financial markets in which many parties exchange information and interact through trading, the overall risk for a company is influenced by many internal and external factors. It is customary for management to classify different types of risk and develop models for dealing with each type of risk in order to keep the risk management problem tractable. One of the different types of risk that a financial institution has to deal with is the market risk, which is the exposure to the uncertain market value of a portfolio [4]. Value-at-risk (VaR) is a way to quantify the market risk. It is a single number for the senior management to express and summarise the total market risk of a portfolio with financial assets. Value at Risk measures the worst expected loss over a given horizon under normal market conditions at a given confidence level. Due to regulations, large banks must nowadays base their market risk capital requirements on the VaR estimate [3]. This drives the continued research into newer and better VaR models.

Different types of approaches have been proposed for VaR estimation, such as simulation and parametric approaches. In the parametric approaches, the risk is quantified in terms of volatility, which is expressed as the standard deviation σ of the portfolio. The simplest models of volatility assume that it does not vary over time, while more advanced models acknowledge that volatility varies dynamically over time. The dynamic aspect of volatility could be modelled in various ways. A popular model where volatility changes dynamically in time is the GARCH (Generalised Auto Regressive Heteroscedasticity) model [5]. For the GARCH $(1, 1)$ model, which is used quite often in practice, the variance is estimated using a first-order autoregressive model of the squared returns.

The disadvantage of the parametric approach is that, due to the complexity of financial markets, the data usually do not follow the parametric distributions that are assumed to underly the data generating process. For example, the returns are typically non-Gaussian, they have fat tails and volatility clustering is often observed in financial markets see *e.g.* [6, 7]. Therefore, flexible modelling approaches such as non-parametric modelling or semi-parametric modelling are needed in which the models can adapt themselves into the underlying actual data distribution. In this context, neural network models for VaR estimation have been studied by various researchers [8, 9, 10] as well as fuzzy set models [11, 12]. Probabilistic fuzzy systems [13] were also used to model VaR by estimating the whole density function, by following the distribution of the data and distributing more membership functions around the origin.

In this paper we focus on tail point density estimation using probabilistic fuzzy systems, to correctly estimate VaR. For this purpose we consider two approaches for estimating the conditional parameters of the PFS model, by giving more relevance to the extreme values, and comparing their performance in obtaining value-at-risk models. A Mamdami-type probabilistic fuzzy system [14] is used for this purpose. The location of the antecedent membership functions is determined by using fuzzy clustering and maximum likelihood parameter estimation is used for determining the probability parameters of the PFS. The validity of the VaR models is evaluated by using a statistical back-testing method based on failure rates.

The outline of the paper is as follows. In Section 2 we discuss the basics of probabilistic fuzzy systems and the concept of fuzzy histograms. In Section 3 we give a brief introduction to VaR modelling and VaR models. In Section 4 we present two methods for estimating the PFS parameters that give more relevance to the tail points of the distribution, where the VaR

ISBN: 978-989-95079-6-8

is estimated. The experimental setup for the empirical study using six different assets are given in Section 5, while the results are reported in Section 6. Finally, conclusions and future work are given in Section 7.

2 Probabilistic Fuzzy Systems

Probabilistic fuzzy systems (PFS) are based on the concept of the probability of a fuzzy event, as defined by Zadeh [15]. It is assumed that the input space is a subset of $\mathbb{R}^n$ and that the rule consequents are defined on a finite domain $Y \in \mathbb{R}$. The PFS consists of a set of rules whose antecedents are fuzzy conditions and whose consequents are probability distributions. In this study, we consider Mamdani PFS in which the rules have the following form [14].

$$\begin{aligned} \text{Rule } R_q\text{: If } \mathbf{x} \text{ is } A_q \text{ then} \\ y \text{ is } C_{q1} \text{ with } \Pr(C_{q1}|A_q) \text{ and} \\ y \text{ is } C_{q2} \text{ with } \Pr(C_{q2}|A_q) \text{ and } \ldots \text{ and} \\ y \text{ is } C_{qN} \text{ with } \Pr(C_{qN}|A_q). \end{aligned} \quad (1)$$

Hence, a Mamdani PFS is a generalisation of a Mamdani fuzzy system in which the deterministic fuzzy rules are replaced with probabilistic fuzzy rules. These rules specify a probability distribution over a collection of fuzzy sets that partition the output domain. The rules of a PFS also express linguistic information and they can be used to explain the model behavior by a set of linguistic rules. This way, the system deals both with linguistic uncertainty as well as probabilistic uncertainty. The interpretation of the probabilistic fuzzy rules is as follows. Given the occurrence of a (multidimensional) antecedent fuzzy event A_q, which is a conjunction of the fuzzy conditions defined on input variables, each of the consequent fuzzy events C_j is likely to occur. The selection of which of the consequent fuzzy events occurs, is done proportionally to the conditional probabilities $\Pr(C_j|A_q)$, $(j = 1, 2, \ldots, N)$. This applies for all the rules R_q, $q = 1, 2, \ldots, Q$. Note that two conditional probabilities $\Pr(C_j|A_q)$ and $\Pr(C_j|A_{q'})$ will be different, in general.

Let

$$\beta_q(\mathbf{x}) = \frac{\mu_{A_q}(\mathbf{x})}{\sum_{q'=1}^{Q} \mu_{A_{q'}}(\mathbf{x})} \quad (2)$$

be the normalised degree of fulfillment of rule R_q, where μ_{A_q} is the degree of fulfillment of rule R_q. When $\mathbf{x}$ is n-dimensional, μ_{A_q} is determined as a conjunction of the individual memberships in the antecedents computed by a suitable t-norm, *i.e.* ,

$$\mu_{A_q}(\mathbf{x}) = \mu_{A_{q_1}}(x_1) \circ \cdots \circ \mu_{A_{q_n}}(x_n), \quad (3)$$

where x_n is the n-th components of $\mathbf{x}$ and $\circ$ denotes a t-norm. Then, it can be shown that the output of the above Mamdani PFS is a conditional probability density function if an additive reasoning scheme is used with multiplicative aggregation of the rule antecedents [16]. The conditional probability of the output given an input vector $\mathbf{x}$ can be computed as

$$f(y|\mathbf{x}) = \sum_{j=1}^{N} \frac{\sum_{q=1}^{Q} \beta_q(\mathbf{x}) \Pr(C_j|A_q) \mu_{C_j}(y)}{\int_{-\infty}^{\infty} \mu_{C_j}(y) dy}, \quad (4)$$

assuming that the output space is well-formed, *i.e.* the output membership values satisfy

$$\sum_{j=1}^{N} \mu_{C_j}(y) = 1, \qquad \forall y \in Y. \quad (5)$$

It is also possible to compute the crisp output of the probabilistic fuzzy system by taking the conditional expectation of the output according to

$$E(y|\mathbf{x}) = \int_{-\infty}^{\infty} y f(y|\mathbf{x}) dy. \quad (6)$$

However, we are not interested in the expected output of the system in this paper. We are primarily interested in the output of the PFS as a fuzzy histogram, by using (4).

Assuming Y is fuzzily partitioned in a set of N fuzzy classes C_j described by membership functions $\mu_{C_j}(y)$, then the (fuzzy) column $f_j(y)$ for fuzzy class C_j can be estimated according to

$$f_j(y) = \frac{\Pr(C_j)\mu_{C_j}(y)}{\int_{-\infty}^{\infty} \mu_{C_j}(y) dy}, \quad (7)$$

The numerator in (7) describes a probability weighted with membership function $\mu_{C_j}(y)$. The denominator of (7) is a scaling factor representing the fuzzified size of class C_j (which in the one-dimensional continuous case, equals the *fuzzy length* of the interval C_q). The complete pdf $f(y)$ is again approximated by a summation of the functions $f_j(y)$:

$$f(y) \approx f_{app}(y) = \sum_{j=1}^{N} f_j(y) = \sum_{j=1}^{N} \frac{\Pr(C_j)\mu_{C_j}(x)}{\int_{-\infty}^{\infty} \mu_{C_j}(y) dy}. \quad (8)$$

Due to the overlap of the fuzzy sets, fuzzy histograms approximate probability distributions better, in practice. In Fig. 1 a representation of this phenomenon is shown, where a normal probability density function is approximated using both a crisp and a fuzzy histogram. In both cases, seven classes have been used.

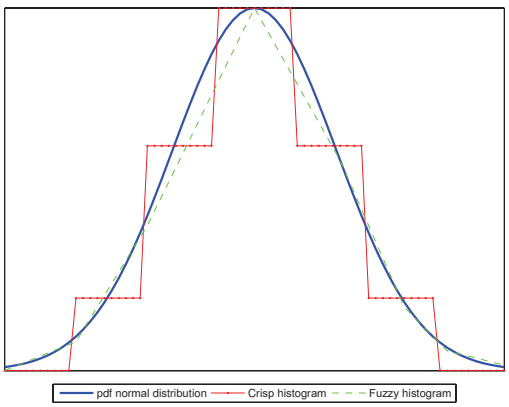


Figure 1: Fuzzy Histogram

3 VaR Estimation

In this section, we discuss value-at-risk estimation as an application example for probabilistic fuzzy systems. Value-at-risk is a single number for the senior management to express and

ISBN: 978-989-95079-6-8

summarise the market risk of a portfolio of financial assets. The VaR value of a portfolio is always calculated over a time horizon h at a significance level c. It indicates the maximum loss that a portfolio of assets will suffer over a horizon of h (days) with a confidence of c. An overview of the mainstream value at risk estimation methods can be found in [17]. Several methods are also discussed in [18]. Various building blocks of VaR measurement, methods for model validation as well as the differences between the parametric and nonparametric estimation approaches are discussed in [3].

3.1 Value at risk

Assume that a portfolio has value W_t at time t. Let r denote the one period percentage return of the portfolio. If $f(r)$ is the probability density function of the returns, define r_v such that

$$1 - c = \int_{-\infty}^{r_v} f(r)dr. \tag{9}$$

The value at risk V_t of the portfolio at time t is then defined as $V_t = -r_v W_t$. Assuming that the returns are distributed normally, the key step in the value at risk estimation can be formulated as determining the standard deviation σ of the returns distribution. This is also called *volatility estimation*.

Probabilistic fuzzy systems have been successfully applied in estimating the whole density distribution. In [13], good results were obtained by scaling the consequent membership functions, which were distributed with more membership functions are around the origin. The distribution of consequent functions alters the approximation of the density functions. Since VaR is calculated at the negative endpoint of the distribution function, it is logical to try to model the tail points more precisely. In Section 4 two approaches are presented that focus on the modeling of the tail points.

3.2 Model validation

Model validation is the process of checking whether a model performs adequately, and can be done in various ways. In this paper, we consider *exception based back testing*. Kupiec has developed a statistical test for assessing the validity of a VaR model [19]. Kupiec confidence regions are defined by the tail point of the log-likelihood ratio LR_{uc}

$$\begin{aligned} LR_{uc} &= -2\ln\left[c^{T-N}(1-c)^N\right] \\ &\quad +2\ln\left\{\left[1-\left(\frac{N}{T}\right)\right]^{T-N}\left(\frac{N}{T}\right)^N\right\}. \end{aligned} \tag{10}$$

In (10), N is the number of exceptions and T is the total number of observations. This ratio is shown to be asymptotically χ^2-distributed, with 1 degree of freedom, under the null hypothesis that the VaR model is valid [19]. Note that the Kupiec test statistic is two sided. Hence, the model is rejected both when there are too few exceptions, (the model is too conservative), as well as when there are too many exceptions, (the model underestimates the volatility). In this paper, we apply the Kupiec test with 95% confidence to assess the validity of the VaR models.

4 Parameter Learning for Tail Points

The parameters of the probabilistic fuzzy systems consist of the number of rules in the system, the parameters of the membership functions (*i.e.* number, type, location, etc.) and the probability parameters $\Pr(C_j|A_q)$ of the stochastic mapping between the antecedent and the consequents. The identification of all the parameters of the PFS simultaneously can be very time consuming and it suffers from the problem of multiple local minima. Thus, we use process knowledge to establish values of a subset of parameters. The other parameters are then optimised given the values of this subset of parameters.

In this work we determine the parameters of the antecedent membership functions by using a fuzzy clustering heuristic, that uses the fuzzy c-means algorithm [20] on the product space of the antecedent variables, and the spreads of the membership functions are derived from the distribution of the data, using a fuzzy covariance matrix, as proposed in [13]. Since the output membership functions must satisfy (5), it is convenient to use triangular membership functions for the output partition. These membership functions are combined with shouldered membership functions at the edges of the domain, as depicted in Fig. 2(a), to ensure that the domain is always covered by the fuzzy partition, no matter how extreme the values may be. The distribution of the membership functions can be uniform over the universe of discourse, or it can be varying. Two possible partitions for the output space are shown in Fig. 2. More membership functions placed around the edges of the universe of discourse, allows to better capture the variability in this region and to better model the tail points of the distribution. As one moves towards the edges of the universe of discourse, the membership functions have smaller supports and the separation between them decreases. Note that the triangular membership functions in this partition are not symmetric.

Assuming that the membership functions in the rule antecedents and the rule consequents have been defined, the optimal probability parameters $\Pr(C_j|A_q)$ can now be determined by using maximum likelihood parameter estimation, in which the log-likelihood function

$$J = \sum_{k=1}^{K} \ln\left(\Pr(y_k|\mathbf{x}_k)\right) \tag{11}$$

is maximised where K is the number of samples in the data set [21]. In (11), it is assumed that the samples in the data set are independent of one another. A suitable initialisation for iterative optimisation for maximum likelihood estimation is given by direct estimation from the data by using

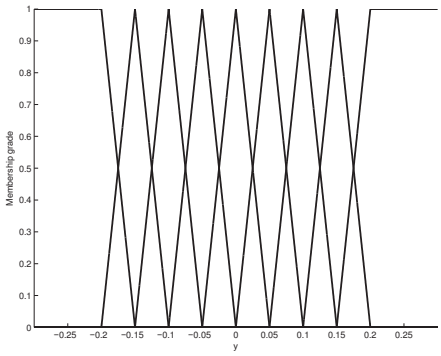
$$\Pr(C_j|A_q) = \frac{\sum_{k=1}^{K}\mu_{C_j}(y_k)\mu_{A_q}(\mathbf{x}_k)}{\sum_{k=1}^{K}\mu_{A_q}(\mathbf{x}_k)}. \tag{12}$$

During the parameter estimation of the optimal probability parameters $\Pr(C_j|A_q)$, more relevance can be given to the extremes of the universe of discourse. This can be accomplished by using an appropriate weight function $w(y_k)$ on the maximum likelihood parameter estimation, dependent on the distribution of the variable, that will give more weight to the tail points, in the form:

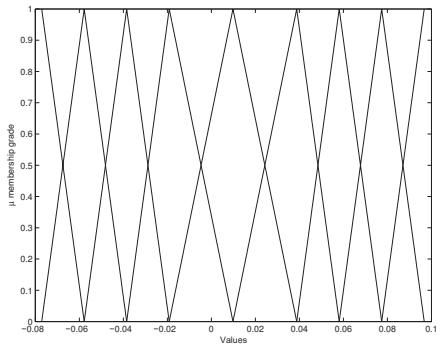
$$J = \sum_{k=1}^{K} \ln\left(w(y_k)\Pr(y_k|\mathbf{x}_k)\right). \tag{13}$$

The most simple kind of functions that can be used is an

ISBN: 978-989-95079-6-8



(a) Equally spaced.



(b) More sets towards the edges.

Figure 2: Two possible partitions for the output space.

affine function, such as

$$w(y_k) = \begin{cases} (\min_j(y_j)/2)^{-1} y_k & \text{if } y_k \leq 0, \\ (\max_j(y_j))^{-1} y_k & \text{if } y_k > 0. \end{cases} \quad (14)$$

Note that in this function more weight is given to the negative side, to better model the negative tail points and possibly obtain better VaR estimation.

5 Experimental Setup

We have studied the performance of the histogram-based probabilistic fuzzy systems to estimate VaR for six different stocks: KPN, ABN AMRO, JiaLing, BaoShan, COSCO and Merchant Bank. The performance of PFS models has been compared with the performance of the GARCH models.

In this paper, we consider value at risk models for one period ahead. In other words, the horizon over which the value at risk is computed is one day. Extensive models could be identified for the multiple day case (i.e. h-day horizon), but usually one suffices by using the simple $\sqrt{(h)}$ scaling, where the h-day value at risk is taken as $\sqrt{(h)}$ times the one-day value at risk.

The probabilistic fuzzy models that we consider use the returns r_t at period t to predict the distribution of the returns at period $t+1$. In our models, we have used nine antecedent membership functions and nine consequent membership functions. Hence, the fuzzy system had nine rules. The input space was partitioned using the FCM algorithm with nine clusters. In such a system, there are 81 probability parameters $\Pr(C_j|A_q)$ (nine for each rule). Since FCM has the tendency to place more clusters in regions covered with more data, there are more antecedent membership functions in the centre, where more samples are available.

The output space was partitioned with triangular membership functions. These triangular membership functions are combined with shouldered membership functions at the edges of the domain, to ensure that the domain is always covered by the fuzzy partitions, no matter how extreme the returns may be on a particular day.

Given the fuzzy membership functions whose parameters are determined as above, the conditional probability parameters for the PFS are determined by using maximum likelihood estimation. We used uniformly distributed triangular membership functions over the universe of discourse, combined with (13), and we name this model PFS$_{UD}$. For the case where we used the maximum likelihood parameter estimation as stated in (11), we used varying size of triangular membership functions over the universe of discourse, and we name this model PFS$_V$. Both approaches could be applied together. Nonetheless, in this work, we are interested in studying different ways of modelling the tail points of the probability density functions obtained with PFS, and thus we study them separately. Given the conditional probability distribution of one period returns, the value at risk of the portfolio is obtained by using (9).

The steps necessary for computing the one-period value-at-risk of a portfolio can now be summarised as follows for PFS$_1$ models.

1. Collect the price series regarding the portfolio and compute the one-period returns. Create training and validation data sets.
2. Determine antecedent membership functions by applying fuzzy c-means clustering to compute the locations of the membership functions and use cluster covariance to obtain the spreads of the membership functions.
3. Select the number of consequent membership functions and form a triangular partition, equally spaced in the case of PFS$_{UD}$ and with more sets towards the edges on PFS$_V$.
4. Given the definitions of the antecedent and the consequent membership functions, determine the optimal probability parameters of the PFS by maximising (13) for the PFS$_{UD}$ and (11) for the PFS$_V$.
5. Using the test set, compute the estimated conditional probability distribution function for the one-period returns for each observation in the test set.
6. Given the conditional probability distribution functions, compute the VaR by using (9).
7. Validate the model by using exception based back-testing as explained in Section 3.2.

6 Results

This section reports the application results for the proposed approaches to the stocks in study, for the validation data sets. Table 1 shows the obtained results of the exception-based back testing for the best GARCH and probabilistic fuzzy models.

ISBN: 978-989-95079-6-8

This table shows the number of exceptions that have occurred in the validation data for different levels of the confidence parameter c. The bold face numbers indicate that the model is not rejected according to the test statistic. The non-rejection region for the Kupiec test statistic is also shown.

Table 1: Failure rates for back testing

Asset	c	PFS_{UD}	PFS_V	GARCH	Non-Rejection Region
	95%	**26**	**30**	**19**	$16 < N < 36$
ABN	97.5%	**15**	**13**	**13**	$6 < N < 20$
	99%	**5**	**6**	**9**	$1 < N < 10$
	95%	**29**	**22**	11	$16 < N < 36$
KPN	97.5%	**14**	**15**	**8**	$6 < N < 20$
	99%	**10**	**9**	**4**	$1 < N < 10$
	95%	**30**	**35**	**22**	$16 < N < 36$
JiaLing	97.5%	**11**	**14**	**14**	$6 < N < 20$
	99%	**6**	**6**	6	$1 < N < 10$
	95%	**24**	**20**	12	$16 < N < 36$
BaoShan	97.5%	**14**	**13**	**8**	$6 < N < 20$
	99%	**9**	**4**	**6**	$1 < N < 10$
	95%	**26**	**25**	14	$16 < N < 36$
COSCO	97.5%	**9**	**9**	**11**	$6 < N < 20$
	99%	**5**	**6**	**5**	$1 < N < 10$
	95%	**27**	**27**	10	$16 < N < 36$
Merchant	97.5%	**15**	**14**	5	$6 < N < 20$
	99%	**6**	**3**	**4**	$1 < N < 10$

As can be seen in Table 1 the GARCH models are rejected for some data sets, while the PFS models are accepted for all data sets. This implies the presence of extreme values in the returns series that are not captured by the GARCH model [7]. Note that for KPN, the number of exceptions obtained for $c = 99\%$ is on the limit indicated by the Kupiec test.

Table 2 shows the initial probability parameters obtained with (12). It can be seen that all probability variables are positive according to this estimation. Table 3 and Table 4 shows the optimal probability parameters obtained after maximum likelihood estimation for PFS_{UD} and PFS_V, respectively.

Table 2: Initial probability parameters for ABN AMRO model.

	Consequent								
Rule	1	2	3	4	5	6	7	8	9
1	0.1003	0.1333	0.2066	0.0567	0.0441	0.1215	0.1012	0.1218	0.1143
2	0.0565	0.1351	0.1659	0.0792	0.0744	0.0972	0.1617	0.1489	0.0811
3	0.0459	0.1679	0.1495	0.1300	0.1024	0.0773	0.1463	0.1359	0.0448
4	0.0544	0.1683	0.1700	0.1105	0.0647	0.0802	0.1454	0.1579	0.0468
5	0.0516	0.1578	0.1800	0.1119	0.0692	0.0770	0.1547	0.1472	0.0506
6	0.0563	0.1648	0.1760	0.0956	0.0655	0.1206	0.1340	0.1227	0.0646
7	0.0700	0.1877	0.1659	0.0748	0.0405	0.0901	0.1147	0.2002	0.0562
8	0.0529	0.1625	0.1626	0.1122	0.0999	0.0930	0.1313	0.1286	0.0570
9	0.0539	0.1729	0.1624	0.1132	0.0746	0.0772	0.1476	0.1505	0.0476

Note that some of the probability parameters are now zero. As tables Table 3 and Table 4 show, more probability mass is centered around the middle consequents. A higher probability mass around the center may indicate that the extreme returns are much less frequent, but this is not always the case. The returns seem to revert to the average after a positive or negative extreme value, and the volatility seems to be rather different for the average returns and the extremes [16]. Hence these results capture the tail behavior in the returns distribution. When we compare these results with the ones obtained when distributing more membership functions are around the origin [13], we can see that the tails of the return distribution are more pronounced in the presented work. We believe that this is due to the fact the in the presented work, both proposed approaches better model and capture the extreme values of the data.

Table 3: Probability parameters for PFS_{UD} ABN AMRO model after optimisation.

	Consequent								
Rule	1	2	3	4	5	6	7	8	9
1	0	0	0.0397	0.2675	0.6928	0	0	0	0
2	0	0	0.1049	0.3454	0.5497	0	0	0	0
3	0	0.0090	0	0.3466	0.6225	0.0219	0	0	0
4	0.0562	0	0.0700	0.5111	0.1847	0.1224	0	0.0556	0
5	0	0.0121	0.0074	0.4865	0.4797	0	0.0142	0	0
6	0	0.0357	0	0.4496	0.4550	0.0085	0.0513	0	0
7	0.0145	0.0472	0	0.4219	0.4774	0.0226	0.0163	0	0
8	0	0	0	0.3355	0.5736	0	0.0908	0	0
9	0	0.0441	0	0.1853	0.6198	0.1286	0	0	0.0222

Table 4: Probability parameters for PFS_V ABN AMRO model after optimisation.

	Consequent								
Rule	1	2	3	4	5	6	7	8	9
1	0	0.0123	0.0326	0.2897	0.6654	0	0	0	0
2	0	0.0310	0	0.4509	0.4618	0	0.0563	0	0
3	0	0.0435	0	0.2054	0.6082	0.1208	0	0	0.0222
4	0.0596	0	0.0719	0.5362	0.1819	0.0909	0	0.0596	0
5	0	0.0122	0	0.3315	0.6317	0.0245	0	0	0
6	0	0	0.1068	0.4104	0.4828	0	0	0	0
7	0	0.0070	0.0244	0.4559	0.4989	0	0.0138	0	0
8	0.0229	0	0.0096	0.4270	0.5231	0.0071	0.0104	0	0
9	0	0	0	0.3320	0.5878	0	0.0801	0	0

The obtained probability density functions for ABN-AMRO using the PFS_V are shown in Fig. 3. As can be seen, for some cases the distribution does not resembles a normal distribution. Specifically, the cases were the distribution exhibits fat tails, correspond to high volatility data.

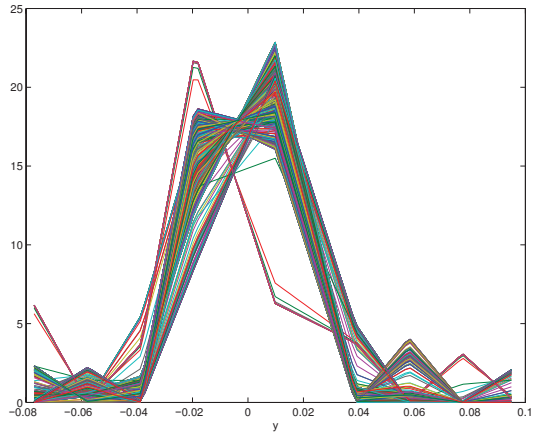


Figure 3: Obtained pdf ABN-AMRO using PFS_V

It is also interesting to consider how the VaR values estimated by the PFS compare to the values estimated by the GARCH models. Table 5 shows the sum of the differences between the VaR estimated and the actual losses in the periods where the VaR estimation is smaller than the actual losses, i.e., when exceptions occur. As can be seen in Table 5, the ex-

ISBN: 978-989-95079-6-8

Table 5: VaR Exceptions

Asset	c	PFS_{UD}	PFS_V	GARCH
ABN	95%	0.0339	0.0333	0.0446
	97.5%	0.0430	0.0452	0.0510
	99%	0.0586	0.0558	0.0587
KPN	95%	0.0591	0.0596	0.0887
	97.5%	0.0744	0.0724	0.1015
	99%	0.0902	0.0881	0.1166
JiaLing	95%	0.0301	0.0288	0.0405
	97.5%	0.0371	0.0348	0.0463
	99%	0.0469	0.0434	0.0533
BaoShan	95%	0.0439	0.0244	0.0350
	97.5%	0.0478	0.0314	0.0400
	99%	0.0531	0.0463	0.0460
COSCO	95%	0.0327	0.0330	0.0499
	97.5%	0.0428	0.0429	0.0571
	99%	0.0561	0.0570	0.0656
Merchant	95%	0.0264	0.0266	0.0362
	97.5%	0.0312	0.0318	0.0413
	99%	0.0371	0.0388	0.0475

pected losses are in most cases smaller in the PFS models, with the exception of BAOSHAN. In other cases where the GARCH model has smaller total expected losses than the PFS model, the GARCH model leads to a smaller number of exceptions, which may indicate a conservative model.

7 Conclusions

We have proposed two approaches for determining the model parameters of probabilistic fuzzy systems for value at risk modelling, where both approaches give more relevance to the tail points of the distribution of values. In one of the approaches we distributed more membership functions around the edges of the universe of discourse and used maximum likelihood to estimate the conditional parameters of the model. In the other approach we distributed the membership functions uniformly over the universe of discourse and used a weighted maximum likelihood to estimate the conditional parameters of the model.

The performance of the proposed models has been compared to the VaR estimation by using the popular GARCH (1,1) volatility estimation. It is found that PFS models are not rejected by back testing, while GARCH models are sometimes rejected. Furthermore, we show how VaR estimation is improved by using these approaches. This shows the added flexibility that comes through the use of the probabilistic fuzzy models, enabling them to adapt to the properties of the data and of the problem at study.

References

[1] R. A. Brealey and S. C. Myers. *Principles of Corporate Finance*. McGraw Hill, New York, 7th edition edition, 2001.

[2] J. Hull. *Options, Futures and Other Derivatives*. Prentice Hall, Upper Saddle River, 4 edition, 2000.

[3] Philippe Jorion. *Value at Risk : the new benchmark for managing financial risk*. McGraw–Hill, New York, 2 edition, 2001.

[4] Glyn A. Holton. *Value at Risk: thoery and practice*. Academic Press, Amsterdam, The Netherlands, 2003.

[5] T. Bollerslev. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31:307–327, 1986.

[6] Rama Cont. Volatility clustering in financial markets: empirical facts and agent-based models. In G. Teyssiére and A. P. Kirman, editors, *Long Memory in Economics*, pages 289–310. Springer, Berlin, 2007.

[7] N.S. Balke and T.B. Fomby. Large shocks, small shocks, and economic fluctuations: outliers in macroeconomic time series. *Journal of Applied Econometrics*, 9:181–200, 1994.

[8] J. W. Taylor. A quantile regression neural network approach to estimating the conditional density of multiperiod returns. *Journal of Forecasting*, 19(4):299–311, July 2000.

[9] N. Chapados and Y. Bengio. Cost functions and model combination for var-based asset allocation using neural networks. *IEEE Transactions on Neural Networks*, 12(4):890–906, July 2001.

[10] M. Jiang, Z. Q. Meng, and Q. Y. Hu. A neural network model on solving multiobjective conditional value-at-risk. In F. Yin, J. Wang, and C. Guo, editors, *Advances in Neural Networks*, number 3174 in Lecture Notes in Computer Science, pages 1000–1006. Springer, Berlin, 2004.

[11] Zdeněk Zmeškal. Value at risk methodology under soft conditions approach (fuzzy-stochastic approach). *European Journal of Operational Research*, 161:337–347, 2005.

[12] Umberto Cherubini and Giovanni Della Lunga. Fuzzy value-at-risk: accounting for market liquidity. *Economic Notes by Banca Monte dei Paschi di Siena SpA*, 30(2):293–312, 2001.

[13] Du Xu and Uzay Kaymak. Value-at-risk estimation by using probabilistic fuzzy systems. In *Proceedings of the 2008 World Congress on Computational Intelligence*, pages 2109–2116, Hong-Kong, June 2008.

[14] Uzay Kaymak, Willem-Max van den Bergh, and Jan van den Berg. A fuzzy additive reasoning scheme for probabilistic Mamdani fuzzy systems. In *Proceedings of the 2003 IEEE International Conference on Fuzzy Systems*, volume 1, pages 331–336, St. Louis, USA, May 2003.

[15] L. A. Zadeh. Probability measures of fuzzy events. *Journal of Mathematical Analysis and Applications*, 23(2):421–427, 1968.

[16] Jan van den Berg, Uzay Kaymak, and Willem-Max van den Bergh. Financial markets analysis by using a probabilistic fuzzy modelling approach. *International Journal of Approximate Reasoning*, 35:291–305, 2004.

[17] Darrell Duffie and Jun Pan. An overview of value at risk. *Journal of Derivatives*, 4(3):7–49, Spring 1997.

[18] Z. Wiener. Introduction to VaR (value-at-risk). In D. Galai, D. Ruthenberg, M. Sarnat, and B. Schreiber, editors, *Risk Management and Regulation in Banking*, pages 47–63. Kluwer Academic Publishers, Boston, 1999.

[19] Paul H. Kupiec. Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 2:73–84, 1995.

[20] J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York, 1981.

[21] L. Waltman, U. Kaymak, and J. van den Berg. Maximum likelihood parameter estimation in probabilistic fuzzy classifiers. In *Proceedings of the 14th Annual IEEE International Conference on Fuzzy Systems*, pages 1098–1103, Reno, Nevada, USA, May 2005.

ISBN: 978-989-95079-6-8

Optimization of an Oil Production System using Neural Networks and Genetic Algorithms

Guillermo Jimenez de la C[1,2] Jose A. Ruz-Hernandez[1] Evgen Shelomov[1]

Ruben Salazar M.[1,2]

1 Universidad Autonoma del Carmen, Av. 56 # 4 esq.
Av. Concordia, Col. Aviacion, C.P. 24180, Cd. del Carmen, Campeche, Mexico.
2 Instituto Mexicano del Petroleo, Av. Periferica Norte # 75 esq. Calle 35B, Col. San Agustin del Palmar,
C.P. 24118, Cd. del Carmen, Campeche, Mexico.
Email: gjc@mexico.com, jruz@pampano.unacar.mx, eshelomov@pampano.unacar.mx, rsalazar@imp.mx

Abstract— *This paper proposes an optimization strategy which is based on neural networks and genetic algorithms to calculate the optimal values of gas injection rate and oil rate for oil production system. Two cases are analyzed: a) A single well production system and b) A production system composed by two gaslifted wells. For both cases an objective function is maximized to reduce production cost. The proposed strategy shows the ability of the neural networks to approximate the behavior of an oil production system and the genetic algorithms to solve optimization problems when a mathematical model is not available.*

Keywords— Genetic algorithms, injection gaslift, neural network, optimization, oil production system, perceptron multilayer.

1 Introduction

The daily operation of an oil and gas production system, have many decisions, which affect the volumes produced and the cost of production oil. These decisions are taken at different levels in the organization, but eventually they will reach the physical production system [1]. Fig. 1 gives an overview of a physical gas lift production system. For such oil production systems, the decisions are related to find the lift gas rate for each well giving the maximum total oil production rate at very instance of time.

An objective function is a single-valued and well-defined mathematical function mapping the values of the decision variables into a performance measure. Examples of such performance measures are the total oil production rate, net present value (profit), or the recovery of the reservoir. To improve the performance of the production system, a question to be answered is: what decisions are better to maximize or minimize the objective function?. In the process of making good decisions, information about the production system is used. This information may include the physical properties such as pipe diameters and lengths, or it may include measurements from the production system.

To support making good decisions, well models may be used to develop the production plans. Typically, well test are performed to determine the gas to oil ratio, water cut, and production rates of each individual well. Well test are performed by routing a well to a dedicated separator. This separator will separate three phases, and a rate transmitter is connected to the outlet for each phase. The well model is update using the measurements taken during a test. Fluid sampling may be used to obtain the fluid composition including the water cut.

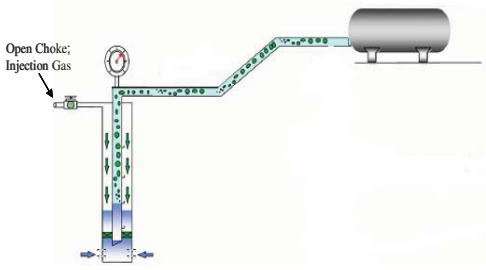


Figure 1: Gas lift production system of a single well.

The objective of gas lift is to increase oil production or allow nonrateing wells to rate by reducing the hydrostatic head of the fluid column in the well [2]. By injecting gas into the tubing, the density of the wellbore fluid decrease; thus, the pressure-drop component resulting from gravity is reduced. However, the gas lift also gives a larger pressure-drop component resulting from friction, giving some optimal gas lift rate for the well. Usually, the available lift-gas for a group of wells is less than the sum of the individual optimum lift-gas rates for each well. The gas-lift optimization problem is established to find the lift gas rates for each well giving the maximum total oil production rate subject to a gas lift processing capacity constraint and possibly other operational and processing constraints.

In this paper, a model-based optimization via neural networks and genetic algorithms is developed and used to calculate the optimal values of gas injection rate and oil rate of a gas lift production system. Two cases were analyzed: a) A single well production system and b) A production system composed by two gas lifted wells. For both cases maximize the objective function to reduce production cost. The proposed strategy shows the ability of the neural networks to approximate the behavior of an oil production system and to solve optimization problems when a mathematical model is not available.

This paper presents a methodology of hybrid computational intelligent using neural networks and genetic algorithm. Others related references for example; genetic algorithms for neural network training on transputers, neural network weight selection using genetic algorithms, studies on the speed of convergence of neural network training using genetic algorithm, automatic generation of neural networks

ISBN: 978-989-95079-6-8

with parameter setting based on genetic algorithms, evolutionary algorithms for neural network design and training, others.

2 Optimization strategy based on a neural network and genetic algorithms

In order to solve the gas-lift optimization problem, an optimization procedure based on a neural network and genetic algorithms was developed. The strategy selected is based on three components as illustrated in Fig. 2. The first one is a neural network which is used to approximate the gas lift performance curve, the second one uses an objective function to satisfy a performance index and the third one is used to solve the optimization problem via genetic algorithms.

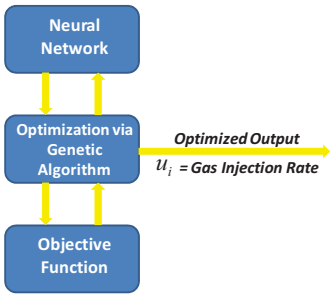


Figure 2: Strategy of optimization based on a neural network and genetic algorithms.

2.1 Neural Network

A Multi-Layer Perceptron (MLP) is selected. The use of this kind of networks to approximate functions and carry out identification process goes back to more than one decade [3]. The following equation is used to determine the structure of a MLP with a single hidden or intermediate layer, a neuron in the output layer with function of linear activation and M hidden neurons is used;

$$F(x_1,...,x_P)=\sum_{i=1}^{M}\alpha_i g(\sum_{j=1}^{p} w_{ij}x_j-\theta_i) \qquad (1)$$

The expression to define the neural network used in our strategy is given by

$$Q_i = g[u_i(i-1)] \qquad (2)$$

where:

Q_i : is the estimated produced oil rate (STB/day).

u_i : is the Gas lift rate injected into the well (MMscf/day)

g: function activation

In this application, the approximation procedure is done using a neural network multilayer perceptron with three layers. The hidden layer has neurons using sigmoid activation function and the output layer has a unique neuron with linear activation function. Different MLPs are trained by means of Levenberg-Marquardt Algorithm, which uses the criterion of middle square error to update the neural network weights. The corresponding MLP is displayed in Fig. 3.

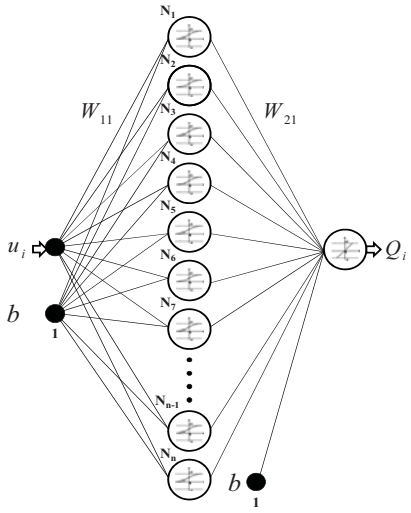


Figure 3: Neural network MLP.

2.2 Objective function

The expression to define the objective function used in our strategy of optimization is given by:

$$J(u)=\sum_{i=1}^{N}\alpha_i Q_i(u_i)-\beta_i u_i \qquad (3)$$

where:

α_i : Production costs by 1 STB/day produced oil.

β_i : Supply costs by 1 MMscf/day injected gas.

N : amount wells.

α_i and β_i are required to balance units

[STB/day]/[MMscf/day]= STBD/MMscfd

The objective function given in (3) can give a net gain, relating adequately processes numerical comparison of the output of the desired product and quantity of flow of injection gas.

2.3 Optimization via genetic algorithms

In order to maximize the objective function, given in (3), the following simple genetic algorithm is applied [4].

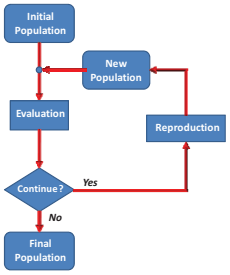


Figure 4: Simple genetic algorithm pseudocode diagram.

ISBN: 978-989-95079-6-8

The properties of genetic algorithms which are applied are the following ones

- Type of genetic algorithm: Simple genetic algorithm [4].
- Amount of unit in population: 100 chromosomes
- Amount of units that contain "N" genes, to find the optimal value: 200 generations.
- Probability of mutation= 0.001.
- In order to select the units the match method is used.

In this application we take into account that the values of injection gas must satisfy the following conditions.

$$u_i = \{0.00, 0.01, 0.002, \ldots, 5.99, 6\} \quad (4)$$

The previous condition represents the precision considered when the injection gas is sampled.

Other conditions are:

- The rate of gas injection will be a value in the following interval (MMscf/day):

$$0.00 \le u_i \le 6.00 \quad (5)$$

- There is not mathematical model of the process; however, experimental data can be obtained from a well simulator.
- A neural network model of the process is available, which can be used to approximate its behavior and to construct the objective function.

Two cases are simulated; the first one corresponds to a single well production system, then the equation (3) can be rewritten like:

$$Max[J(u)] = \alpha_1 Q_1 - \beta_1 u_1 \quad (6)$$

and the second one considers a production system composed by two wells, then the equation (3) can be rewritten like:

$$Max[J(u)] = \alpha_1 Q_1 + \alpha_2 Q_2 - \beta_1 u_1 - \beta_2 u_2 \quad (7)$$

3 Results and Discussions

3.1 First case: Produced oil by a single well

The corresponding oil production system is illustrated in Fig. 1. For this case, two single wells are considered. The first one has a pressure in the well head (P_{wh}) equal to 14 kg/cm^2 and the second one has 12 Kg /cm^2. A simulation program is used to collect data and to train neural networks using neural networks toolbox of Matlab.

The best trained neural networks are described in Table 1 and Table 2. Other parameters like the static pressure in the reservoir, lengths and diameters of the tubing and lines, chokes and others pipe components are also considered. Figure 5 illustrates the simulation for a well with P_{wh}= 14 kg/cm^2.

Table 1: Neural network architecture and obtained errors with P_{wh} = 14 Kg/cm^2.

Number of neurons		error	error (%)
Intermedia layer	Output layer		
30	1	0.1365	0.002

Table 2: Neural network architecture and obtained errors with P_{wh} = 12 Kg/cm^2.

Number of neurons		error	error (%)
Intermedia layer	Output layer		
30	1	0.0119	0.0004

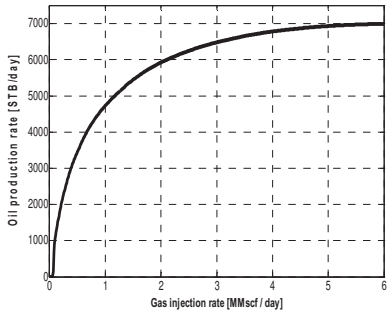


Figure 5: Estimated production curve oil production rate vs gas injection rate, P_{wh} = 14 Kg/cm2.

Figure 6 displays the obtained results when the gain of production for this well with P_{wh}=14 Kg/cm^2 is maximized. Furthermore,α_1=*28*USD/STBD and β_1=*5250*USD/MMscfd are considered in equation (6); α_1 *and* β_1 estimates cost for this year.

The obtained results in this case are contained in Table 3.

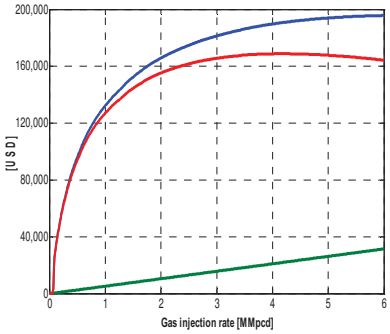


Figure 6: Optimization results; income of production (blue), costs by supply of injection gas (green) and objective function (Optimal gain-red).

ISBN: 978-989-95079-6-8

Table 3: Gain of production results

Optimal gain production (USD)	Income optimal production (USD)	Costs by supply of injection gas (USD)	Obtained optimal gas injection rate (MMscf / day)
168,840	190,680	21,840	4.16

Figure 7 illustrates the simulation results for a well when P_{wh}=12 kg/cm^2 is used.

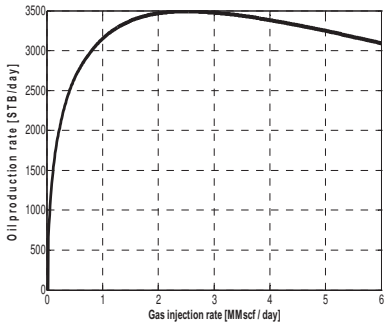


Figure 7: Estimated production curve oil production rate vs gas injection rate, P_{wh} = 12 Kg/cm2.

Figure 8 displays the obtained results via optimization procedure to maximize the gain of production for this well when P_{wh}=12 Kg/cm2 is used. As considered above, α_1=*28*USD/STBD and β_1=*5250*USD/MMscfd are used in equation (6).

The obtained results in this case are contained in Table 4.

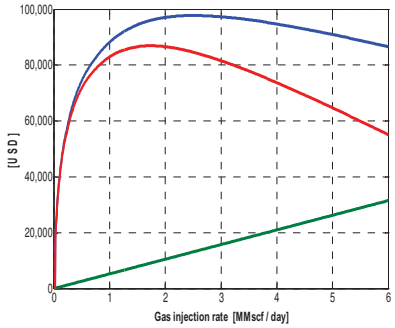


Figure 8: Optimization results; income of production (blue), costs by supply of injection gas (green) and objective function (Optimal gain-red).

Table 4: Gain of production results

Optimal gain production (USD)	Income optimal production (USD)	Costs by supply of injection gas (USD)	Obtained optimal gas injection rate (MMscf / day)
86,921	96,004	9,082	1.73

3.2 Second case: Produced oil rate by a production system based on two wells.

In Fig. 9 is shown an oil production system of two wells. Production system made up of wells of 12 kg/cm^2 and 14 Kg /cm^2. A simulation program is used to collect data and to train neural networks using neural networks toolbox of Matlab. The best trained neural networks is described in Table 5.

Table 5: Neural network architecture and obtained errors, two wells.

Number of neurons		error	error (%)
Intermedia layer	Output layer		
30	1	-0.3702	-0.0037

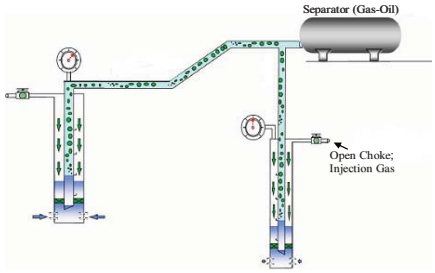


Figure 9: Gas lift production system composed by two wells.

3.2.1 Sub case 1. $u_i = u_1 = u_2$

The Fig.10 illustrates the simulation for a system composed by two wells (Black).

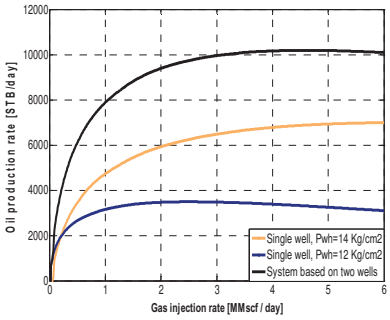


Figure 10: Estimated production curve oil production rate vs gas injection rate. System composed by two wells.

Figure 11 displays the simulation results of maximizing of the gain of production for two wells. Assume that α_1= α_2=*28*USD/STBD and β_1=β*2=5250*USD/MMscfd are considered in equation (7). Furthermore, objective function is subject to the next constrains:

$$0.00 \le u_i \le 6.00 \text{ and } u_2 = u_1 .$$

The obtained results in this subcase are contained in Table 6.

ISBN: 978-989-95079-6-8

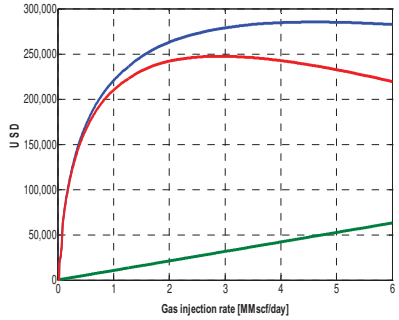


Figure 11: Optimization results; income of production (blue), costs by supply of injection gas (green) and objective function (Optimal gain-red).

Table 6: Gain of production results

Optimal gain production (USD)	Income optimal production (USD)	Costs by supply of injection gas (USD)	Obtained optimal gas injection rate (MMscf / day)	
247,260	277,550	30,240	2.88	2.88

3.2.2 Sub case 2. $u_1+u_2=K$

The Fig. 12 illustrates the simulation for a system composed by two wells (Black).

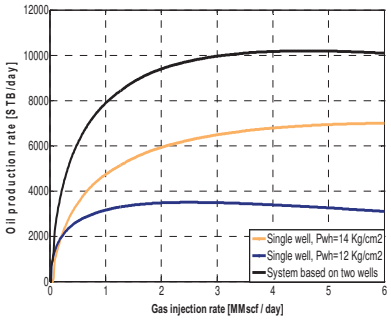


Figure 12: Estimated production curve oil production rate vs gas injection rate. System composed by two wells.

Figure13 presents the results of maximizing of the gain of production for two wells. α_1 *and* β_1 considered in objective function subject to constrains :

$$0.00 \le u_i \le 6.00$$

and

$$u_2 + u_1 = K$$

where $K = 4$. Furthermore, $\alpha_1=\alpha_2=28$USD/STBD and $\beta_1=\beta_2=5250$USD/MMscfd are considered in equation (7).

The obtained results in this subcase are contained in table 7.

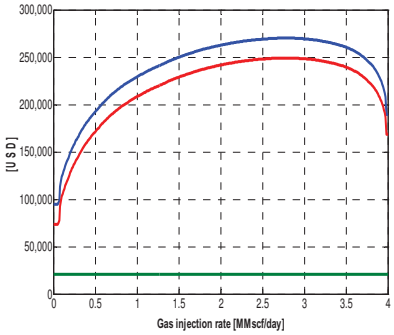


Figure 13: Optimization results; income of production (blue), costs by supply of injection gas (green) and objective function (Optimal gain-red).

Table 7: Gain of production results

Optimal gain production (USD)	Income optimal production (USD)	Costs by supply of injection gas (USD)	Obtained optimal gas injection rate (MMscf / day)	
249,360	270,360	21,000	2.79	1.21

4 Conclusions

The obtained results shows that neural networks and genetic algorithms are useful to optimize the costs and gains in an oil production systems. The implementation of these methodologies to be applied in petroleum industry allows to increase the gains, to reduce the costs, save natural elements and to extend the life of wells.

Acknowledgements

The authors thank support of Universidad Autonoma del Carmen, Mexico on project PR/131/2008.

References

[1] H.P. Bieker, O. Slupphaug, and T. A. Johansen, "Real-Time Production Optimization of Oil and Gas Production Systems: A Technology Survey," *SPE Production & Operations Journal*, vol. 22, no. 4, pp. 382-391, November 2007.

[2] A. Vasper, "Auto, Natural, or In-Situ Gas-Lift Systems Explained," *SPE Production & Operations Journal*, vol. 23, no. 1, pp. 75-80, February 2008.

[3] J. Nocedal, and S.J. Wright, *Numerical Optimization*. New York: Springer, 1999. K Narendra, and K. Parthasarathy, "Identification and control of dynamical systems using neural networks". IEEE Transactions on Neural Networks, vol.1, no 1. pp. 4-27, 1990.

[4] D. Goldberg, (1989). Genetic Algorithms in Search, Optimization, and Machine Learning (Addison-Wesley, Readings, MA).

[5] W.Y. Fang, and K.K. Lo, "A Generalized Well-Management Scheme for Reservoir Simulation," *SPE Reservoir Engineering Journal*, vol. 11, no. 2, pp. 116-120, 1996.

[6] E.P. Kanu, J. Mach, and K.E. Brown. "Economic Approach to Oil Production and Gas Allocation in Continuous Gas Lift," *Journal of Petroleum Technology*, vol. 33, no. 10, pp. 1887-1892, 1981.

ISBN: 978-989-95079-6-8

[7] K Hornik, M. Stinchcombe, and White H. “Multilayer Feedforward Networks are Universal Aproximators”, Neural Networks, vol. 2, no. 5, 1989.

[8] M. Norgaard, O. Ravn; Poulsen, N. K., y L. K., Hansen, “Neural Networks for Modelling and Control of Dynamic Systems”, Springer-Verlag, London, 2000.

[9] Prosper. Petroleum Experts Ltd. User’s Guide IPM V5.3Build #173 Jun 6, 2006.

[10] J. A. Ruz-Hernandez., D. A. Suarez, E. Shelomov. and A. Villavicencio, “Predictive Control Based on an Auto-Regressive Neuro-Fuzzy Model Applied to the Steam Generator Startup Process at a Fossil Power Plant”, Revista de Computación y Sistemas, vol. 6, no. 3, CIC-IPN, ISSN 1405 5546, Mexico, pp.204-212, 2003.

[11] E. N. Sánchez, G A. Alanís, “Redes Neuronales, Conceptos Fundamentales y Aplicaciones al Control Automático”, Prentice Hall, 2006.

[12] J. M. Corchado, F. Díaz. “Redes Neuronales Artificiales, un enfoque Practico”. Universidad de Vigo. ISBN. 848158145-3.

[13] G. Cybenko, “Aproximation by superpositions of sigmoidal function”, Math. Control, Sygnals and System, vol. 2, nº 4, 1989.

ISBN: 978-989-95079-6-8

Rule Based Fuzzy Cognitive Maps in Socio-Economic Systems

Joao Paulo Carvalho[1,2] Jose A. B. Tome [1,2]

1 INESC-ID - Instituto de Engenharia de Sistemas e Computadores

2 IST – Instituto Superior Tecnico, TULisbon
Lisboa, Portugal
Email: {joao.carvalho, jose.tome}@inesc-id.pt

Abstract *— This paper shows how Rule Based Fuzzy Cognitive Maps can be useful to model qualitative socio-economic systems, by discussing and presenting a macroeconomic model that, although developed eight years ago and based on simple qualitative rules, described and predicted the current economic crisis when the policy behavior of the banking system regarding tax-rates was assumed similar to the one that ended up occurring.*

Keywords — Qualitative Modelling of Dynamic Systems, Rule Based Fuzzy Cognitive Maps, Socio-Economic Systems.

1 Introduction♠

Economic models have traditionally been based on mathematics. Econometry, the quantitative science of modelling the economy, focus on creating models to help explain and predict variables of interest in economics. However the most common econometric models are usually very imprecise and are not usually valid but on a very short term. This can easily be seen on the regular predictions made to most macroeconomic indicators: most yearly predictions made by governments, economic entities or independent experts, must usually get corrected every trimester due to inaccuracies in the models used to predict their values. This is essentially due to the fact that most econometric models tend to ignore the existence of the feedback loops that make any alteration in any component of the model to potentially be propagated until that component is affected by its own previous change on a relatively short term.

The more precise models that try to address this issue are usually based on differential equations [13, 20]. However, due to the dimension of these systems (very high number of variables involved), these models demand a strong knowledge in mathematics, and a huge amount of time to be developed. However, even when these kinds of models are possible, they usually tend to ignore that economy is a social science, and therefore is subject to qualitative uncertainties due to human and social factors that are not easily captured using strict quantitative mathematic models.

♠ This work was supported in part by the FCT - Portuguese Foundation for Science and Technology under project PTDC/MAR/66231/2006

The use of approaches that include both the existence of feedback cycles and the intrinsic qualitative social nature associated with economy, should lead to the implementation of more accurate models.

Dynamic Cognitive Maps (DCM), where Rule Based Fuzzy Cognitive Maps (RB-FCM) are included, are a qualitative approach to modeling and simulating the Dynamics of Qualitative Systems (like, for instance, Social, Economical or Political Systems) [2,3,4,5,6,7,8,9,10,11]. RB-FCM were developed as a tool that can be used by non-engineers and/or non-mathematicians and eliminates the need for complex mathematical knowledge when modeling qualitative dynamic systems. In this paper one shows that RB-FCM can be used to model socio-economical systems using as an example a model that was originally designed eight years ago, but that can explain and simulate the current world economic situation.

2 Dynamic Cognitive Maps

The term Dynamic Cognitive Maps has been recently used to describe techniques that allow simulating the evolution of cognitive maps through time. Axelrod [1] work on cognitive maps (CM) introduced a way to represent real-world qualitative systems that could be analyzed using several methods and tools. However, those tools only provided a way to identify the most important structural elements of the CM. Complete, efficient and practical mechanisms to analyze and predict the evolution of data in CM were not available for years due to several reasons. System Dynamics tools like those developed by J.W.Forrester [12] could have provided the solution, but since in CM numerical data may be uncertain or hard to come by, and the formulation of a mathematical model may be difficult, costly or even impossible, then efforts to introduce knowledge on these systems should rely on natural language arguments in the absence of formal models. Fuzzy Cognitive Maps (FCM), as introduced by Kosko [15, 16, 17], were developed as a qualitative alternative approach to system dynamics. However, although very efficient and simple to use, FCM are causal maps (a subset of cognitive maps that only allow basic symmetric and monotonic causal relations between concepts)[9], and, in most applications, a

ISBN: 978-989-95079-6-8

FCM is indeed a man-trained Neural Network that is not fuzzy in a traditional sense and does not exploit usual fuzzy capabilities. They do not share the properties of other fuzzy systems and the causal maps end up being quantitative matrixes without any qualitative knowledge.

Several evolutions and extensions have been added to the original FCM model, but none addresses so many FCM issues as RB-FCM. RB-FCM were developed as a tool that models and simulates real world qualitative system dynamics while trying to avoid the limitations of those approaches. The following sub sections resume some features of RB-FCM that are useful to the comprehension of this paper.

2.1 Rule Based Fuzzy Cognitive Maps

RB-FCM allow a representation of the dynamics of complex real-world qualitative systems with feedback, and the simulation of events and their influence in the system. They can be represented as fuzzy directed graphs with feedback, and are composed of fuzzy nodes (Concepts), and fuzzy links (Relations). RB-FCM are true cognitive maps since are not limited to the representation of causal relations. Concepts are fuzzy variables described by linguistic terms, and Relations are defined with fuzzy rule bases.

RB-FCM are essentially iterative fuzzy rule based systems where we added fuzzy mechanisms to deal with feedback, introduced timing mechanisms [8] and new ways to deal with uncertainty propagation, and were we defined several kinds of Concept relations (Causal, Inference, Alternatives, Probabilistic, Opposition, Conjunction, etc. [2,5]) to cope with the complexity and diversity of the dynamic qualitative systems we are trying to model. Among new contributions brought by RB-FCM, there is a new fuzzy operation – the Fuzzy Carry Accumulation - [4,7], which is essential to model the mechanisms of qualitative causal relations (FCR – Fuzzy Causal Relations) while maintaining the simplicity and versatility of FCM.

There are 2 main classes of Concepts: Levels , that represent the absolute values of system entities (e.g., LInflation is Good); and Variations, that represent the change in value of a system entity in a given amount of time (e.g., VInflation increased very much). By allowing the definition of both the absolute value of a concept and its change through time, RB-FCM have the means to properly model the dynamics of a system.

2.1.1 Expressing Time in Dynamic Cognitive Maps

Time is probably the most essential factor when modeling a dynamic system. However, most DCM approaches seem to ignore this fact. In order to maintain consistency in the process of modeling the dynamics of a qualitative system, it is necessary to develop and introduce timing control mechanisms. To allow the representation of time flow, delays, and the inhibition of certain relations when they have no influence on a given instant, changes were made to the engine of RB-FCM. More details regarding RB-FCM time mechanisms can be found in [8,11]

3 A Qualitative Macro Economic Model as an Example of DCM Modeling in Socio-Economic Systems

In this section one presents a model that was developed eight years ago. The primary goal when this problem was approached was to show the capabilities and ease of use of RB-FCM to model the dynamics of qualitative real world systems. Even if the final model is rather complex and does not contains apparent flaws, it is not, and was never intended to be, a complete model, since it wasn't developed by economic experts (even though some were consulted). However, as it can be seen in the obtained results, the model exhibits a behaviour that is able to describe the current economic crisis and the reasons that lead to it.

Classic cognitive mapping techniques [1] were used as the first step to obtain the model: the concepts and relations were extracted from a short column published in Portuguese newspaper Público in the year 2001 consisting on an economic expert analysis regarding "Short-term Tax Rate evolution in Europe" [18]. Throughout the text, the author introduced several concepts, supporting its theories while explaining the relations between concepts using qualitative knowledge. The "classic" CM obtained was much simpler than the one presented here, which was expected, since the analysis of the dynamics of a much more complex model – like the one we ended up obtaining – would require several months of work using traditional quantitative econometric approaches. The first model used only the most important concepts (the ones really necessary to a short term analysis): Tax Rates, Inflation, Consumption, Oil Price, and Food Cost. Even with such a few concepts, a realistic model becomes rather difficult to analyse due to the complexity of the relations that affect the involved concepts. However, since the goal was to show RB-FCM potential to deal with larger systems and long-term simulations, the model was evolved to be more realistic therefore including more concepts and much more relations. On this step, 13 concepts were added to the original 5 (Figure 7).

At the end of this phase of the modelling process one obtained a classic Cognitive Map – basically a graph where the nodes were the Concepts and each edge represented an existing unknown relation between a pair of Concepts.

3.1 Concept Modelling

The next step was refining the concepts to obtain a linguistic fuzzy representation for each one. This step consisted in defining the class(es) (Variation, Level) and the linguistic terms and membership functions for each concept. In dynamic systems, variations are much more important than absolute values, therefore, most concepts are Variations, some are Levels, and a few key concepts like Inflation, Tax Rate, etc., are both Variations and Levels (the Level value of these concepts is actualized according to its Variation using a special LV relation [5]).

ISBN: 978-989-95079-6-8

The linguistic terms of Levels must have a direct correspondence with the real world values. Therefore we allied common sense and expert consulting (using straight questions like "what do you consider a high value for Inflation?", and receiving answers like "around 4%") to define their membership functions. In the particular case of Levels that depend on LV relations, it was also necessary to define the real-world meaning of a certain amount of variation (e.g., a "Small" increase on inflation is around 0.3%). Figure 5 shows the linguistic terms of the Level concept LInflation.

Variation linguistic terms usually represent qualitative terms without a direct correspondence to absolute values. E.g., VInflation has 11 linguistic values ranging from "Huge Decrease" to "Huge Increase" (Figure 6). Linguistic terms of Variations can usually be represented by standard sets, which simplify and accelerate the modelling process [10].

3.2 Qualitative Modelling of a Qualitative Dynamic System

The huge advantages of using Fuzzy Rule Bases (FRB) to define qualitative relations between Concepts has been largely discussed and proved [2,3,5,9,11]. The major drawback of rule-based fuzzy inference, the combinatorial explosion of the number of rules, is avoided in RB-FCM by the use of Fuzzy Causal Relations and the Fuzzy Carry Operation [4]. Another important feature of RB-FCM is the simplicity of the process of insertion and removal of Concepts and/or Relations, which also reduces the modelling complexity of FRB [4]. Therefore one has in RB-FCM an adequate tool to model qualitative relations. However, the single fact of using linguistic rule bases to model relations does not guarantee the qualitative nature of the model. Let us see the example of Inflation modelling:

A pseudo-qualitative approach using FRB would try to closely map the widespread quantitative approaches: Inflation value is predicted by a weight averaged sum of several factors (Estimated Oil inflation, estimated Food price inflation, etc.). This method is highly dependent on the precision and validity of each factor real-world absolute value. In the proposed model, a novel approach where rules are independent from the real world absolute values was used. The model is based on a qualitative definition of inflation: Economics theory states that economic growth depends on inflation – without inflation there is no growth; In fact, the worst economic crisis (30´s for instance) are associated with deflation; Therefore, it is desirable and expected that all factors that affect inflation have a certain cost increase – If all factors suffer a normal increase, then the inflation will maintain its normal and desired value. Therefore, one can state the following qualitative relation for each of those *n* factors:

"If $factor_n$ has a normal increase, then Inflation will maintain"

This statement is part of the fuzzy rule base of a causal relation. Since fuzzy causal effects are accumulative and their effect is a variation in the value of the consequent, then if all factors that cause inflation have the normally expected increase, Inflation will not vary. If some factors increase more than expected and the others maintain their value then inflation will somehow increase. If a factor increase less than normal, or even decreases, then its effect is a decrease in inflation (note that the final variation of Inflation is given by the accumulation of all causal variation effects – e.g., if some pull it down a bit, and one pulls it up a lot, in the end inflation still can maintain its normal value).

It is possible to build a completely qualitative and sound causal FRB to model each factor influence of Inflation, without ever referring to absolute values. If one intends to model inflation in South America, one can maintain the rule base. All that needs to be changed are the linguistic terms of the Level Concept associated to Inflation (for instance, normal inflation would become around 8 %, and so on...). Obviously some factors are more important than others (a large increase in food might cause a large increase in Inflation, but what is considered a large increase in Oil might only cause a small increase in Inflation – average Oil price varied over 100% in the last 2 years, but other factors had a slightly above average increase, and therefore inflation had mild increase instead of a sever one…). This "relative" importance is easily modelled as a causal effect in a FRB. Table I represents an example of a causal FRB. One can also mention the fact that oil price variation has a delayed effect in inflation. RB-FCM provide mechanisms to model these kinds of timing issues [8,11]

TABLE I: FCR7 +sl Food Cost, Inflation

If Food Cost Decreases VeryMuch,	Inflation has a Large Decrease
If Food Cost Decreases Much,	Inflation has a Large Decrease
If Food Cost Decreases,	Inflation has a Large Decrease
If Food Cost Decreases Few,	Inflation Decreases
If Food Cost Decreases VFew,	Inflation Decreases
If Food Cost Maintains,	Inflation Decreases
If Food Cost Increases VFew,	Inflation has a Small Decrease
If Food Cost Increases Few,	Inflation has a Very Small Decrease
If Food Cost Increases Normally,	Inflation Maintains
If Food Cost Increases M,	Inflation has a Small Increase
If Food Cost Increases VM,	Inflation Increases

This kind of qualitative approach was used throughout the model when causal relations were involved.

As it was mentioned above, Variations usually have a standard set of linguistic terms. These allow the predefinition of certain common fuzzy causal relations (FCR). These FCR are called macros and were used to reduce the modelling effort.

The model includes other than causal relations. For instance: Oil price variation was modelled using a classic fuzzy inference rule base (FIRB) based on oil Offer/Demand (where Oil offer was decided in simulated periodic OPEP meetings); the Tax Rates were modelled considering that Banks were managed as a common business with profit in

ISBN: 978-989-95079-6-8

mind – for example, an increase in money demand would increase Tax Rates (this would be changed later (see 3.4)).

Regarding timing considerations, the system was modelled considering a one month period between iterations.

It is obviously impossible to detail every aspect of the system modelling in this paper. provides a graphic representation of the final RB-FCM model. The system consists of 18 concepts and around 400 fuzzy rules to express relations (most were automatically generated using macros). The system was described using RB-FCMsyntax (a dedicated language) – a complete description is available in [2]. Here are some guidelines regarding the description of relations in Figure 7: "FCR+" stands for a standard positive causal relation (an increase in the antecedent will cause an increase in consequent), and "FCR-" a standard negative relation (increase causes a decrease). Several "+" or "–" represent stronger effects. A "/" represents an attenuated effect. "sl" and "sr" represent biased effects (non symmetric causal relations. A "?" represents a relation which cannot be symbolically described (one must consult the FCR). A "d" represents a delay in the effect. FIR stands for Fuzzy Inference Relation. The number after FCR or FIR is the label for the complete description of the rule base.

3.3 Simulation results[1,2]

The simulation of the original system provided rather interesting results. The evolution of the system through time was rather independent from the initial values and the external effects. After a certain period of time, that could vary from a few months to several years (depending on a conjugation of external factors like a war, or a severe cut in oil production), economy would end up collapsing: deflation, negative growth, 0% tax-rates. Figure 1 represents one of those cases.

Initially one could think that there was a major flaw in the model (or in the RB-FCM mechanisms), but after a discussion and analysis of the results with an economics expert, the culprit was found: the model approached the economic situation before the creation of entities that control Interest Rates (like the U.S. Federal Reserve, the European Central Bank). The lack of these entities was the main cause to economic instability until 1930's. In fact, Economics was known in the 18th and 19th century as the "Dark Science", because all theories indicated that economy was not sustainable. According to the simulation results, depression always comes after a growth period and due to an exaggerate increase in tax rates (the banks try to maximize their profit in a short period, and their greed cause an apparently avoidable crisis). Notice the similarity with the present economic crisis - one will return to this point later on. Therefore, to support the referred theory, a simple model of the European Central Bank behaviour regarding interest rates was added to the model.

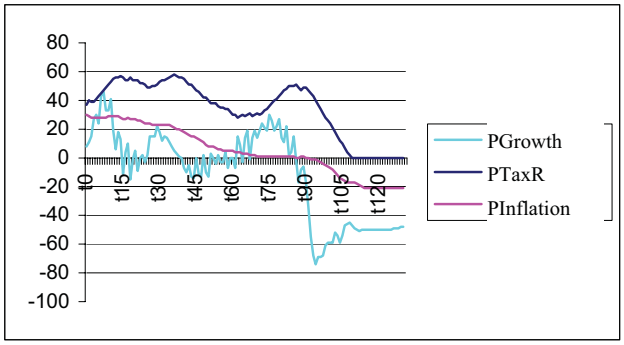


Figure 1 – Serious economic crisis: Negative growth and deflation (Predicted Growth, Predicted TaxRate, Predicted Inflation,)

3.4 Modelling European Central Bank Influence

To simulate ECB influence, a Fuzzy Inference Subsystem (FISS) – a RB-FCM block used to model the process of decision making of system entities (FISS timing mechanisms are independent of the RB-FCM) – was added to the model (Figure 2). This FISS ended up as a simple FRB with 48 rules (each with 2 antecedents)[2]. These rules were designed to inhibit the greedy bank behaviour that was identified as the cause to the unavoidable crisis.

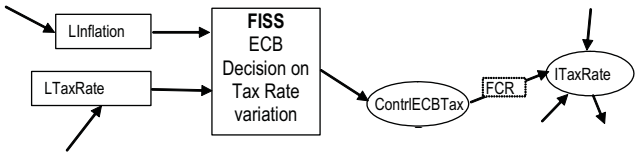


Figure 2 – FISS: ECB decision on Interest Rate variation

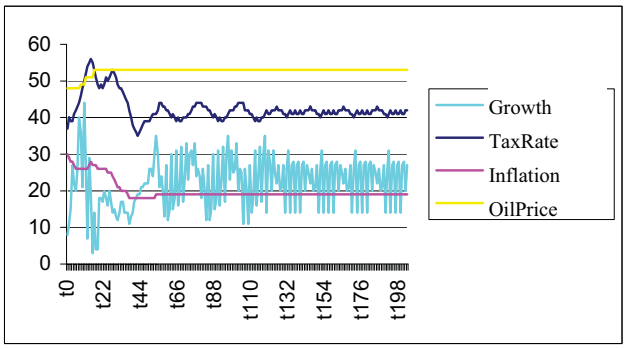


Figure 3 – Avoiding economic crisis trough ECB Interest Rate control

3.5 Complete Model Simulation

With the introduction of the ECB-FISS, system behaviour changed completely and serious crisis were avoided under normal circumstances (Figure 3). One of the most interesting results was the fact that, under normal circumstances, the economic model stabilizes around the real-world BCE predicted ideal target value for inflation (slightly below 2%)

[1] Note: This section refers to an initial version of the system that did not include concept 17

[2] All variables are in % and must be divided by 10 (e.g., 15 represents 1.5%)

ISBN: 978-989-95079-6-8

and growth (averaging slightly above 2%). Note that these values are not imposed anywhere in the model, they result from the system itself.

However, the ECB and private bank behaviour in the last two years was incredibly similar to the greedy behaviour exhibited by the model without ECB. Tax rates – see Euribor historical data [14], Figure 4 – were severely increased between 2006-08 under the pretext of controlling inflation, but, as it was found later, mostly because private banks were needing to increase their tax rates to protect themselves against prior mistakes. The variation of the Euribor tax rate + 0.8% spread rate in the last 10 years (120 month), is very much similar to what was predicted in the "greedy" model that was presented 8 years ago. As a result of that, and, as the original model predicted, we are entering a severe economic crisis as a result of that policy. Given that this is a very long term 10 year simulation (done 8 years ago), the results are incredibly more accurate than current models, that don't usually attempt to predict for longer than 2 years and usually with very inaccurate results. As a proof of this, less than 1 year ago all major economic actors were still insisting on increasing tax rates having inflation control in mind, and no major economic actor was even suspecting that deflation would be the real concern in less than 8 months.

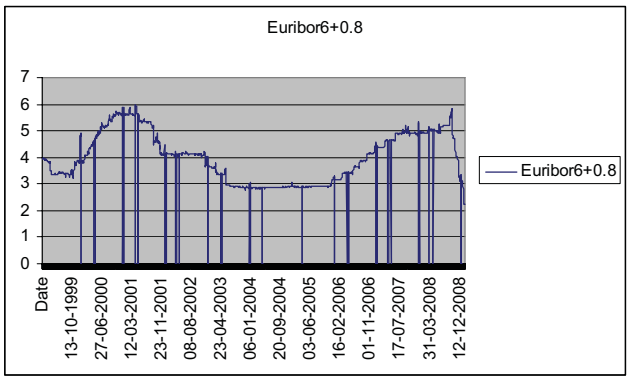


Figure 4 – Euribor6+0.8 for the last 120 months

4 Conclusions, Applications, and Future Developments

In this paper we exemplified how one can use DCM to model complex qualitative socio-economic systems, avoiding the need to use extensive and time consuming differential equation models, while obtaining very interesting and encouraging results.

By using true qualitative modelling techniques, one obtained results that look more realistic (plausible) than those obtained using quantitative approaches – where results almost never show the short term uncertainties that are so characteristic of qualitative real-world dynamic systems. In the end, the results of the presented model, that was developed eight years ago, are surprisingly realistic and could have been used to predict and avoid the current world economic crisis, even if one considers its necessary incompleteness.

5 References

[1] Axelrod, R., "The Structure of Decision: Cognitive Maps of Political Elites", Princeton University Press, 1976

[2] Carvalho, J.P., "Mapas Cognitivos Baseados em Regras Difusas: Modelação e Simulação da Dinâmica de Sistemas Qualitativos", PhD thesis, Instituto Superior Técnico, Universidade Técnica de Lisboa, Portugal, 2002

[3] Carvalho, J.P., Tomé, J.A., "Qualitative Optimization of Fuzzy Causal Rule Bases using Fuzzy Boolean Nets", Fuzzy Sets and Systems, 158(17), pp. 1931-1946, Elsevier, 2007

[4] Carvalho, J.P., Tomé,J.A., "Views on Fuzzy Sets and Systems from Different Perspectives. Philosophy and Logic, Criticisms and Applications. (Studies in Fuzziness and Soft Computing), Chapter 19 - Fuzzy Mechanisms for Qualitative Causal Relations", Edited by. Rudolf Seising, Springer, 2009

[5] Carvalho, J.P., Tomé, J.A., "Rule Based Fuzzy Cognitive Maps – Qualitative Systems Dynamics", Proceedings of the 19th International Conference of the North American Fuzzy Information Processing Society, NAFIPS2000, Atlanta, 2000

[6] Carvalho, J.P., Tomé, J.A.,"Issues on the Stability of Fuzzy Cognitive Maps and Rule-Based Fuzzy Cognitive Maps", Proceedings of the 21st International Conference of the North American Fuzzy Information Processing Society, NAFIPS2002, New Orleans, 2002

[7] Carvalho, J.P., Tomé,J.A., "Fuzzy Mechanisms For Causal Relations", Proceedings of the Eighth International Fuzzy Systems Association World Congress, IFSA'99, Taiwan, 1999

[8] Carvalho, J.P., Tomé,J.A.,"Rule Based Fuzzy Cognitive Maps - Expressing Time in Qualitative System Dynamics", Proceedings of the 2001 FUZZ-IEEE, Melbourne, Australia, 2001

[9] Carvalho,J.P., Tomé, J.A., "Rule Based Fuzzy Cognitive Maps and Fuzzy Cognitive Maps - A Comparative Study", Proceedings of the 18th International Conference of the North American Fuzzy Information Processing Society, NAFIPS99, New York, 1999

[10] Carvalho,J.P., Tomé, J.A., "Using Interpolated Linguistic Term to Express Uncertainty in Rule Based Fuzzy Cognitive Maps", Proceedings of the 22nd International Conference of the North American Fuzzy Information Processing Society, NAFIPS2003, Chicago, 2003

[11] Carvalho, J.P., Wise,L., Murta, A., Mesquita, M., "Issues on Dynamic Cognitive Map Modelling of Purse-seine Fishing Skippers Behavior", WCCI2008 - 2008 IEEE World Congress on Computational Intelligence, Hong-Kong , pp. 1503-1510 , 2008

[12] Forrester, J.W., several papers available online at http://sysdyn.mit.edu/sd-intro/home.html

[13] Halanay, A., Samuel, J., "Differential Equations, Discrete Systems and Control: Economic Models (Mathematical Modelling: Theory and Applications)", Kluwer Academic, 1997

[14] http://www.euribor.org/html/content/euribor_data.html

[15] Kosko, B., "Fuzzy Cognitive Maps", International Journal of Man-Machine Studies, 1986

[16] Kosko, B., "Fuzzy Engineering", Prentice-Hall International Editions, 1997

[17] Kosko, B., "Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence", Prentice-Hall International Editions, 1992

[18] Neves, A., "Taxas de Juro podem descer já este mês", Público 2001/08/10

ISBN: 978-989-95079-6-8

[19] Zadeh, L., "Fuzzy Sets and Applications: Selected Papers, Wiley-Interscience", 1987

[20] Zhang, W.-B., "Differential Equations, Bifurcations and Chaos in Economics", World Scientific, 2005

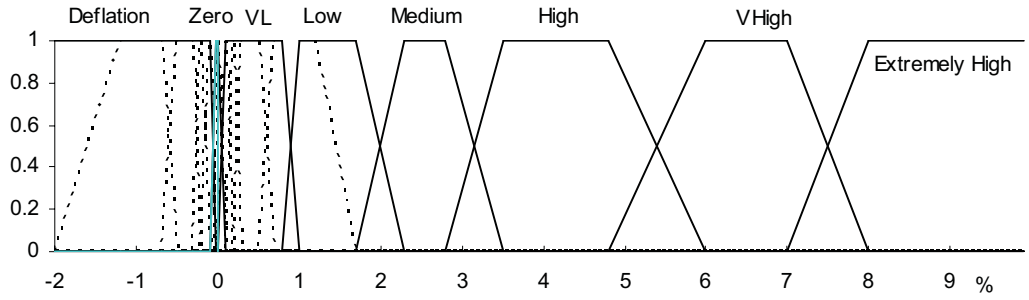


Figure 5 – Level "Inflation" linguistic terms. Dotted linguistic terms represent the variation degrees of Inflation {Huge_Decrease, Large_Decrease, Decrease, Small_Decrease, VS_Decrease, Maintain, VS_Increase,…., Huge_Increase}

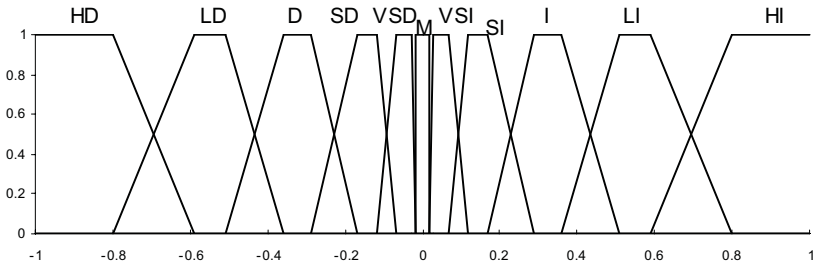


Figure 6 – Variation "Inflation" linguistic terms: {Huge Decrease, Large Decrease,…,Huge Increase}. x-scale values are normalized values. There is no relation no real world values

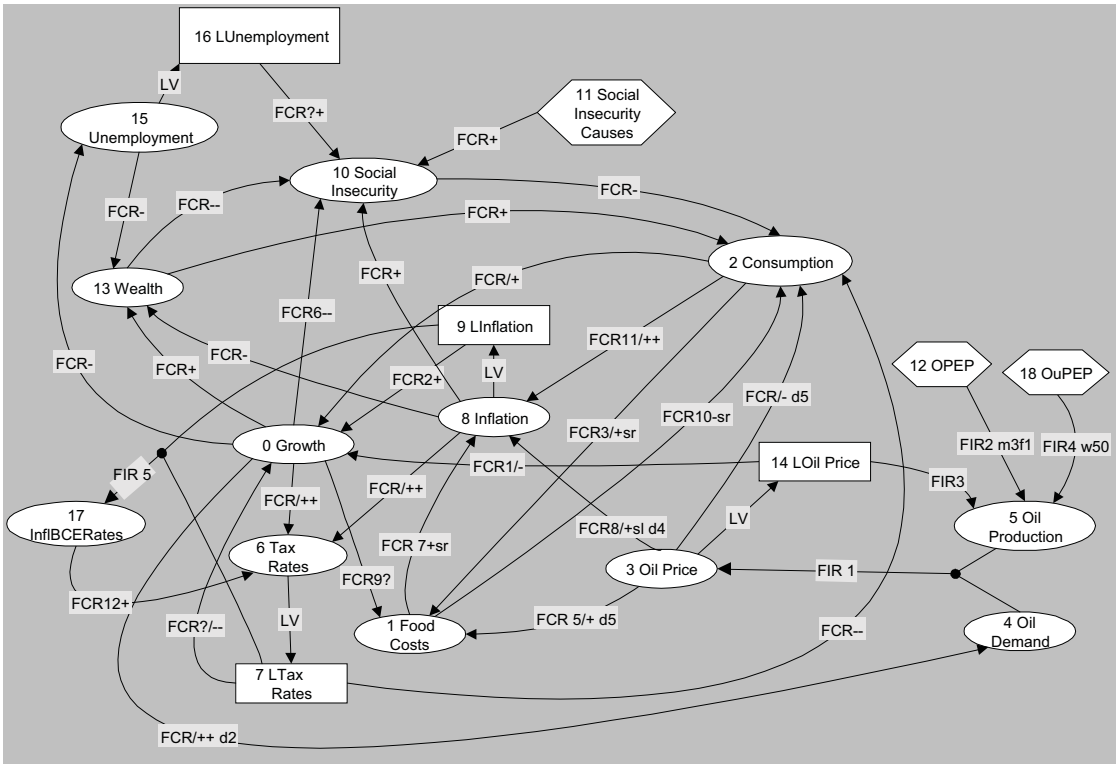


Figure 7 – RB-FCM: A qualitative model of economy. Concept 17, representing a simple FISS was added later

ISBN: 978-989-95079-6-8

Operational Semantics for a Fuzzy Logic Programming System with Defaults and Constructive Answers

Hannes Strass Susana Munoz-Hernandez Victor Pablos Ceruelo

Technical University of Madrid (Spain) *
E-mail: hannes.strass@alumnos.upm.es, {susana, vpablos}@fi.upm.es

***Abstract*— *In this paper we present the operational semantics of RFuzzy, a fuzzy Logic Programming framework that represents thruth values using real numbers from the unit interval. RFuzzy provides some useful extensions: default values to represent missing information, and typed terms to intuitively restrict predicate domains. Together, they allow the system to give constructive answers in addition to truth values. RFuzzy does not confine to a particular Fuzzy Logic, but aims at being as general as possible by using the notion of aggregation operators.*

Keywords*— *Operational Semantics, Logic Programming Application

1 Introduction

For many real-world problems, crisp knowledge representation is not perfectly adequate. Information that we handle might be imprecise, uncertain, or even both. Classical two-valued logic cannot easily represent these qualitative aspects of information. To address this issue, multiple frameworks for incorporating uncertainty in logic have been developed over the years: fuzzy set theory, probability theory, multi-valued logic, or possibilistic logic; to mention only some.

From the point of view of practical tools to support this reasoning the field is not so rich. Logic programming is however a perfect candidate for the implementation of these tools because it is traditionally used for problem solving and knowledge representation.

1.1 Fuzzy Logic approaches

The result of introducing Fuzzy Logic into Logic Programming has been the development of several fuzzy systems over Prolog. These systems replace the inference mechanism, SLD-resolution, of Prolog with a fuzzy variant that is able to handle partial truth [1]. Most of these systems implement the fuzzy resolution introduced by Lee in [2], examples being the Prolog-Elf system [3], the Fril Prolog system [4] and the F-Prolog language [5]. However, there is no common method for fuzzifying Prolog as has been noted in [6].

1.2 Fuzzy Prolog

One of the most promising fuzzy tools for Prolog was the "Fuzzy Prolog" system [7, 8]. This approach is more general than others in some respects:

1. A truth value is a finite union of sub-intervals on $[0, 1]$.
2. A truth value is propagated through the rules by means of an *aggregation operator*. The definition of *aggregation operator* is general.
3. Crisp and fuzzy reasoning are consistently combined in a Prolog compiler [9].

Fuzzy Prolog adds fuzziness to a Prolog compiler using CLP($\mathcal{R}$) instead of implementing a new fuzzy resolution as other former fuzzy Prologs do. So, it uses Prolog's built-in inference mechanism, and the constraints and their operations provided by CLP($\mathcal{R}$) to handle the concept of partial truth. It represents intervals as constraints over real numbers and *aggregation operators* as operations with these constraints.

There are other proposals, e.g. in [10], that provide an interpretation of truth values as intervals, but Fuzzy Prolog proposed to generalise this concept to unions of intervals for the first time.

1.3 Multi-adjoint logic

Over the last few years several papers have been published by Medina et al. [11, 12, 13] about multi-adjoint programming. The theoretical model described in these works led to the development of FLOPER [14], another Fuzzy Logic Programming system. It has a Logic-Programming-inspired syntax and provides free choice of aggregation operators and credibility of rules just as RFuzzy does. There are however some things that FLOPER cannot do: (1) deal with missing information (which RFuzzy does by default truth value declarations), (2) type atoms and predicates to give constructive answers, and (3) provide syntactic sugar to express truth value functions.

1.4 Motivation

The generality of the approach pursued in [7, 8] turned out to make many users feel uncomfortable: Fuzzy Prolog is rather expressive, so it is not always clear how knowledge should be represented. Furthermore, interpreting the output that comes as a sequence of constraints maybe possible for a human but is very hard to do for a computer program – especially basing a decision upon it may not be straightforward.

To address these issues, we propose the RFuzzy framework. It is considerably simpler to use than the above-mentioned Fuzzy Prolog, but still contains many of its nice features. In RFuzzy, truth values will be represented by real numbers from the unit interval $[0, 1]$. This simplifies making modelling decisions and interpreting the output of the system. We still use the general concept of aggregation operators to be able to model different Fuzzy Logics. The rules of RFuzzy programs will have attached a *credibility* value to them: it allows the author

*This work is partially supported by the project DESAFIOS - TIN 2006-15660-C02-02 from the Spanish Ministry of Education and Science, by the Spanish Ministry of Science and Innovation Research Staff Training Program - BES-2008-008320 and by the project PROMESAS - S-0505/TIC/0407 from the Madrid Regional Government.

ISBN: 978-989-95079-6-8

of the rule to express how much they confide in the relation expressed by the rule. In addition, RFuzzy offers features that are very useful for knowledge representation, namely default values and types.

The rest of the paper is organised as follows: Section 2 introduces the abstract syntax of RFuzzy. In Section 3 we present an operational semantics for RFuzzy and illustrate it with an example. The last but one section shortly sketches the key points of the implementation and Section 5 concludes.

2 RFuzzy Syntax

We will use a signature Σ of function symbols and a set of variables V to "build" the *term universe* $\mathrm{TU}_{\Sigma,V}$ (whose elements are the *terms*). It is the minimal set such that each variable is a term and terms are closed under Σ-operations. In particular, constant symbols are terms.

Similarly, with use a signature Π of predicate symbols to define the *term base* $\mathrm{TB}_{\Pi,\Sigma,V}$ (whose elements are called *atoms*). Atoms are predicates whose arguments are elements of $\mathrm{TU}_{\Sigma,V}$. Atoms and terms are called *ground* if they do not contain variables. As usual, the *Herbrand universe* H is the set of all ground terms, and the *Herbrand base* B is the set of all atoms with arguments from the Herbrand universe.

To combine truth values in the set of real truth values $[0,1]$, we will make use of *aggregation operators*. A function $\hat{F} : [0,1]^n \to [0,1]$ is called an aggregation operator if it verifies $\hat{F}(0,\ldots,0) = 0$ and $\hat{F}(1,\ldots,1) = 1$. We will use the signature Ω to denote the set of used operator symbols F and $\hat{\Omega}$ to denote the set of their associated aggregation operators $\hat{F}$. An n-ary aggregation operator is called *monotonic in the i-th argument*, if additionally $x \leq x'$ implies $\hat{F}(x_1,\ldots,x_{i-1},x,x_{i+1},\ldots,x_n) \leq \hat{F}(x_1,\ldots,x_{i-1},x',x_{i+1},\ldots,x_n)$. An aggregation operator is called *monotonic* if it is monotonic in all arguments.

Immediate examples for aggregation operators that come to mind are typical examples of t-norms and t-conorms: minimum $\min(a,b)$, maximum $\max(a,b)$, product $a \cdot b$, and probabilistic sum $a + b - a \cdot b$.

The above general definition of aggregation operators subsumes however all kinds of minimum, maximum or mean operators.

Definition. Let Ω be an aggregation operator signature, Π a predicate signature, Σ a term signature, and V a set of variables.

A *fuzzy clause* is written as

$$A \xleftarrow{c,F_c}_{F} B_1,\ldots,B_n$$

where $A \in \mathrm{TB}_{\Pi,\Sigma,V}$ is called the head, $B_1,\ldots,B_n \in \mathrm{TB}_{\Pi,\Sigma,V}$ is called the body, $c \in [0,1]$ is the credibility value, and $F_c \in \Omega^{(2)}$ and $F \in \Omega^{(n)}$ are aggregation operator symbols (for the credibility value and the body resp.)

A *fuzzy fact* is a special case of a clause where $n = 0$, $c = 1$, F_c is the usual multiplication of real numbers "$\cdot$" and $F = v \in [0,1]$. It is written as $A \leftarrow v$.

A *fuzzy query* is a pair $\langle A, v\rangle$, where $A \in \mathrm{TB}_{\Pi,\Sigma,V}$ and v is either a "new" variable that represents the initially unknown truth value of A or it is a concrete value $v \in [0,1]$ that is asked to be the truth value of A. ⌟

Intuitively, a clause can be read as a special case of an implication: we combine the truth values of the body atoms with the aggregation operator associated to the clause to yield the truth value for the head atom. For this truth value calculation we are completely free in the choice of an operator.

Example. Consider the following clause, that models to what extent cities can be deemed good travel destinations – the quality of the destination depends on the weather and the availability of sights:

$$\text{good-destination(X)} \xleftarrow{1.0,\cdot} \text{nice-weather(X), many-sights(X).}$$

The credibility value of the rule is 1.0, which means that we have no doubt about this relationship. The aggregation operator used here in both cases is the product "$\cdot$". We enrich the knowledge base with facts about some cities and their continents:

$$\begin{gathered}
\text{nice-weather(madrid)} \leftarrow 0.8,\\
\text{nice-weather(istanbul)} \leftarrow 0.7,\\
\text{nice-weather(moscow)} \leftarrow 0.2,\\
\text{many-sights(madrid)} \leftarrow 0.6,\\
\text{many-sights(istanbul)} \leftarrow 0.7,\\
\text{many-sights(sydney)} \leftarrow 0.6,\\
\text{city-continent(madrid, europe)} \leftarrow 1.0,\\
\text{city-continent(moscow, europe)} \leftarrow 1.0,\\
\text{city-continent(sydney, australia)} \leftarrow 1.0,\\
\text{city-continent(istanbul, europe)} \leftarrow 0.5,\\
\text{city-continent(istanbul, asia)} \leftarrow 0.5.
\end{gathered}$$

Some queries to this program could ask if Madrid is a good destination, $\langle \text{good-destination(madrid)}, \mathsf{v}\rangle$. Another query could ask if Istanbul is the perfect destination, $\langle \text{good-destination(istanbul)}, 1.0\rangle$.The result of the first query will be the real value 0.48 and the second one will fail. It can be seen that no information about the weather in Sydney or sights in Moscow is available although these cities are "mentioned". ◇

In the above example, the knowledge that we represented using fuzzy clauses and facts was not only vague but moreover incomplete. As this is rather the norm than the exception, we would like to have a mechanism that can handle non-present information.

In standard logic programming, the closed-world assumption is employed, i.e. the knowledge base is not only assumed to be sound but moreover to be complete. Everything that can not be derived from the knowledge is assumed to be false. This could be easily modelled in this framework by assuming the truth value 0 as "default" truth value, so to speak. Yet we want to pursue a slightly more general approach: arbitrary default truth values will be explicitly stated for each predicate. We even allow the definition of different default truth values for different arguments of a predicate. This is formalised as follows.

Definition. A *default value declaration* for a predicate $p \in \Pi^{(n)}$ is written as $\texttt{default}(p(X_1,\ldots,X_n)) =$

ISBN: 978-989-95079-6-8

$[\delta_1$ if $\varphi_1, \ldots, \delta_m$ if $\varphi_m]$ where $\delta_i \in [0,1]$ for all i. The φ_i are first-order formulas restricted to terms from $\mathrm{TU}_{\Sigma,\{X_1,\ldots,X_n\}}$, the predicates $=$ and $\neq$, the symbol true, and the junctors $\wedge$ and $\vee$ in their usual meaning. ⌟

Example (continued). Let us add the following default value declarations to the knowledge base and thus close the mentioned gaps.

$$\texttt{default}(\mathsf{nice\text{-}weather(X)}) = 0.5,$$
$$\texttt{default}(\mathsf{many\text{-}sights(X)}) = 0.2,$$
$$\texttt{default}(\mathsf{good\text{-}destination(X)}) = 0.3$$

They could be interpreted as: when visiting an arbitrary city of which nothing further is known, it is likely that you have nice weather but you will less likely find many sights. Irrespective of this, it will only to a small extent be a good travel destination.

To model the fact that a city is not on a continent unless stated otherwise, we add another default value declaration for city-continent: $\texttt{default}(\mathsf{city\text{-}continent(X, Y)}) = 0.0$. Notice that in this example $m = 1$ and $\varphi_1 = \mathsf{true}$ for all the default value declarations. ◇

The default values allow our knowledge base to answer arbitrary questions about predicates that occur in it. But will the answers always make sense? To stay in the above example, if we ask a question like "What is the truth value of nice-weather(australia)?" we will get the answer "0.5" which does not make too much sense since Australia is not a city, but a continent.

To address this issue, we introduce types into the language. Types can be viewed as inherent properties of terms – each term can have zero or more types. We use them to restrict the domains of predicates.

Definition. A *term type declaration* assigns a type $\tau \in \mathcal{T}$ to a term $t \in H$ and is written as $t : \tau$. A *predicate type declaration* assigns a type $(\tau_1, \ldots, \tau_n) \in \mathcal{T}^n$ to a predicate $p \in \Pi^n$ and is written as $p : (\tau_1, \ldots, \tau_n)$, where τ_i is the type of p's i-th argument. ⌟

Example (continued). Using the set of types $\mathcal{T} = \{\mathsf{City}, \mathsf{Continent}\}$, we add some term type declarations to our knowledge base:

$$\mathsf{madrid} : \mathsf{City}, \mathsf{istanbul} : \mathsf{City},$$
$$\mathsf{sydney} : \mathsf{City}, \mathsf{moscow} : \mathsf{City};$$
$$\mathsf{africa} : \mathsf{Continent}, \mathsf{america} : \mathsf{Continent},$$
$$\mathsf{antarctica} : \mathsf{Continent},$$
$$\mathsf{asia} : \mathsf{Continent}, \mathsf{europe} : \mathsf{Continent}.$$

We also type the predicates in the obvious way:

$$\mathsf{nice\text{-}weather} : (\mathsf{City}),$$
$$\mathsf{many\text{-}sights} : (\mathsf{City}),$$
$$\mathsf{good\text{-}destination} : (\mathsf{City}),$$
$$\mathsf{city\text{-}continent} : (\mathsf{City}, \mathsf{Continent}).$$

◇

For a ground atom $A = p(t_1, \ldots, t_n) \in B$ we say that it is *well-typed with respect to* T iff $p : (\tau_1, \ldots, \tau_n) \in T$ implies $\tau_i \in \mathfrak{t}_T(t_i)$ for all i.

For a ground clause $A \xleftarrow{c,F_c}_F B_1, \ldots, B_n$ we say that it is well-typed w.r.t. T iff all B_i are well-typed for $1 \leq i \leq n$ implies that A is well-typed (i.e. if the clause preserves well-typing). We say that a non-ground clause is well-typed iff all its ground instances are well-typed.

Example (continued). With respect to the given type declarations, city-continent(moscow, antarctica) is well-typed whileas city-continent(asia, europe) is not. ◇

A *fuzzy logic program* P is a triple $P = (R, D, T)$ where R is a set of fuzzy clauses, D is a set of default value declarations, and T is a set of type declarations.

From now on, when speaking about programs, we will implicitly assume the signature Σ to consist of all function symbols occurring in P, the signature Π to consist of all the predicate symbols occurring in the program, the set $\mathcal{T}$ to consist of all types occurring in type declarations in T, and the signature Ω of all the aggregation operator symbols. For Ω we will furthermore require that all operators from $\hat{\Omega}$ be monotonic.

Lastly, we introduce the important notion of a "well-defined" program.

Definition. A fuzzy logic program $P = (R, D, T)$ is called *well-defined* iff

- for each predicate symbol p/n occurring in R, there exist both a predicate type declaration and a default value declaration;
- all clauses in R are well-typed;
- for each default value declaration $\texttt{default}(p(X_1, \ldots, X_n)) = [\delta_1$ if $\varphi_1, \ldots, \delta_m$ if $\varphi_m]$, the formulas φ_i are pairwise contradictory and $\varphi_1 \vee \cdots \vee \varphi_m$ is a tautology, i.e. exactly one default truth value applies to each element of p/n's domain.

⌟

3 Operational Semantics

The possibility to define default truth values for predicates offers us a great deal of flexibility and expressivity. But it also has its drawbacks: reasoning with defaults is inherently non-monotonic – we might have to withdraw some conclusions that have been made in an earlier stage of execution. To capture this formally, we attach to each truth value an attribute that indicates how this value has been concluded. There are 3 different cases of how a truth value can be determined:

- exclusively by application of program facts and clauses, represented by the symbol ▼ denoting the attribute value *safe*,
- by indirect use of default values, represented by the symbol ♦ denoting the attribute value *unsafe (mixed)*, or
- directly via a default value declaration, represented by the symbol ▲ denoting the attribute value *unsafe (pure)*.

We need to be able to compare the attributes (in order to be able to prefer one conclusion over another) and to combine them to keep track of default value usage in the course of

ISBN: 978-989-95079-6-8

computation. This is formalised by setting the ordering $<_a$ on truth value attributes such that $\blacktriangle <_a \blacklozenge <_a \blacktriangledown$.

The operator $\circ : \{\blacktriangle, \blacklozenge, \blacktriangledown\} \times \{\blacktriangle, \blacklozenge, \blacktriangledown\} \to \{\blacktriangle, \blacklozenge, \blacktriangledown\}$ is then defined as:

$$x \circ y := \begin{cases} \blacktriangledown & \text{if } x = y = \blacktriangledown \\ \blacktriangle & \text{if } x = y = \blacktriangle \\ \blacklozenge & \text{otherwise} \end{cases}$$

The operator $\circ$ is designed to keep track of attributes during computation: only when two "safe" truth values are combined, the result is known to be "safe", in all other cases it is "unsafe". It should be noted that "$\circ$" is monotonic.

The truth values that we use in the description of the semantics will be real values $v \in [0, 1]$ with an attribute (i.e. a $z \in \{\blacktriangle, \blacklozenge, \blacktriangledown\}$) attached to it. We will write them as zv. The ordering $\preccurlyeq$ on the truth values will be the lexicographic product of $<_a$, the ordering on the attributes, and the standard ordering $<$ of the real numbers. The set of truth values is thus totally ordered as follows:

$$\bot \prec \blacktriangle 0 \prec \cdots \prec \blacktriangle 1 \prec \blacklozenge 0 \prec \cdots \prec \blacklozenge 1 \prec \blacktriangledown 0 \prec \cdots \prec \blacktriangledown 1.$$

A *valuation* $\sigma : V \to B$ is an assignment of ground terms to variables. Each valuation σ uniquely constitutes a mapping $\hat{\sigma} : \mathrm{TU}_{\Sigma,V} \to B$ that is defined in the obvious way.

A *fuzzy Herbrand interpretation* (or short, *interpretation*) of a fuzzy logic program is a mapping $I : B \to \mathbb{T}$ that assigns truth values to ground atoms.

The *domain* of an interpretation is the set of all atoms to which a "proper" truth value is assigned: $\mathsf{Dom}(I) := \{A \mid A \in B, I(A) \succ \bot\}$.

For two interpretations I and J, we say I *is less than or equal to* J, written $I \sqsubseteq J$, if $I \sqsubseteq J$ iff $I(A) \preccurlyeq J(A)$ for all $A \in B$.

Accordingly, the infimum (or intersection) and supremum (or union) of interpretations are, for all $A \in B$, defined as $(I \sqcap J)(A) := \min(I(A), J(A))$ and $(I \sqcup J)(A) := \max(I(A), J(A))$.

The pair $(\mathcal{I}_P, \sqsubseteq)$ of the set of all interpretations of a given program with the interpretation ordering forms a complete lattice. This follows readily from the fact that the underlying truth value set $\mathbb{T}$ forms a complete lattice with the truth value ordering $\preccurlyeq$.

Definition. [Model] Let $P = (R, D, T)$ be a fuzzy logic program.

For a clause $r \in R$ we say that I *is a model of the clause* r and write

$$I \Vdash A \xleftarrow{c,F_c}_F B_1, \ldots, B_n$$

iff for all valuations σ, we have: if $I(\sigma(B_i)) = z_i v_i \succ \bot$ for all i, then $I(\sigma(A)) \succcurlyeq z'v'$ where $z' = z_1 \circ \cdots \circ z_n$ and $v' = \hat{F}_c(c, \hat{F}(v_1, \ldots, v_n)))$.

For a default value declaration $d \in D$ we say that I *is a model of the default value declaration* d and write

$$I \Vdash \mathtt{default}(p(X_1, \ldots, X_n)) = [\delta_1 \text{ if } \varphi_1, \ldots, \delta_m \text{ if } \varphi_m]$$

iff for all valuations σ, we have: if $\sigma(p(X_1, \ldots, X_n))$ is well-typed (w.r.t. T), then there exists an $1 \le j \le m$ such that $\sigma(\varphi_j)$ holds and $I(\sigma(p(X_1, \ldots, X_n))) \succcurlyeq \blacktriangle \delta_j$.

We write $I \Vdash R$ if $I \Vdash r$ for all $r \in R$ and similarly $I \Vdash D$ if $I \Vdash d$ for all $d \in D$.

Finally, we say that I *is a model of the program* P and write $I \Vdash P$ iff $I \Vdash R$ and $I \Vdash D$. ⌟

The operational semantics will be formalized by a transition relation that operates on (possibly only partially instantiated) computation trees. Here, we will not need to keep track of default value attributes $\{\blacktriangle, \blacklozenge, \blacktriangledown\}$ explicitly, it will be encoded into the computations.

Definition. Let Ω be a signature of aggregation operator symbols and W a set of variables with $W \cap V = \varnothing$.

A *computation node* is a pair $\langle A, e \rangle$, where $A \in \mathrm{TB}_{\Pi,\Sigma,V}$ and e is a term over $[0, 1]$ and W with function symbols from Ω. We say that a computation node is *ground* if e does not contain variables. A computation node is called *final* if $e \in [0, 1]$. A final computation node will be indicated as $\underline{\langle A, e \rangle}$.

We distinguish two different types of computation nodes: C-nodes, that correspond to applications of program clauses, and D-nodes, that correspond to applications of default value declarations.

A *computation tree* is a directed acyclic graph whose nodes are computation nodes and where any pair of nodes has a unique (undirected) path connecting them. We call a computation tree ground or final if all its nodes are ground or final respectively.

For a given computation tree t we define the *tree attribute*

$$z_t = \begin{cases} \blacktriangledown & \text{if } t \text{ contains no D-node} \\ \blacklozenge & \text{if } t \text{ contains both C- and D-nodes} \\ \blacktriangle & \text{if } t \text{ contains only D-nodes} \end{cases}$$

⌟

Computation nodes are essentially generalizations of queries that keep track of aggregation operator usage.

Computation trees as defined here should not be confused with the usual notion of SLD-trees. While SLD-trees describe the whole search space for a given query and thus give rise to different derivations and different answers, computation trees describe just a state in a single computation.

The computation steps that we perform on computation trees will be modelled by a relation between computation trees.

Definition. [Transition relation] For a given fuzzy logic program $P = (R, D, T)$, the transition relation $\rightarrowtriangle$ is characterized by the following transition rules:

Clause: $t/\left[\boxed{\langle A', v \rangle}\right] \rightarrowtriangle$

$$t/\left[\boxed{\langle A', v \rangle} \quad / \quad \begin{array}{c} \mathsf{C}\langle A', F_c(c, F(v_1, \ldots, v_n)) \rangle \\ \diagup \quad \diagdown \\ \langle B_1, v_1 \rangle \ \cdots \ \langle B_n, v_n \rangle \end{array}\right] \mu$$

If there is a (variable disjoint instance of a) program clause $A \xleftarrow{c,F_c}_F B_1, \ldots, B_n \in R$ and $\mu = \mathrm{mgu}(A', A)$. (Take a non-final leaf node and add child nodes according to a program clause; apply the most general unifier of

ISBN: 978-989-95079-6-8

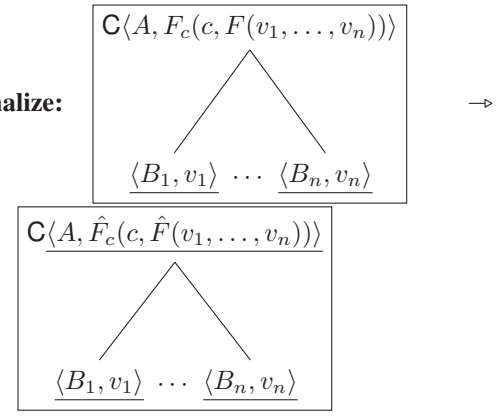
the node atom and the clause head to all the atoms in the tree.)

Note that we immediately finalize a node when applying this rule for a fuzzy fact.

Default: $\boxed{t\,[\langle A, x\rangle]} \quad \rightarrowtriangle \quad \boxed{t\left[\langle A, x\rangle / \mathsf{D}\underline{\langle A, \delta_j\rangle}\right]\mu}$

If A does not match with any program clause head, there is a default value declaration $\texttt{default}(p(X_1, \ldots, X_n)) = [\delta_1 \text{ if } \varphi_1, \ldots, \delta_m \text{ if } \varphi_m] \in D$, μ is a substitution such that $p(X_1, \ldots, X_n)\mu = A\mu$ is a well-typed ground atom, and there exists a $1 \leq j \leq m$ such that $\varphi_j\mu$ holds. (Apply a default value declaration to a non-final leaf node thus finalizing it.)

Finalize:



(Take a non-final node whose children are all final and replace its truth expression by the corresponding truth value.)

⌟

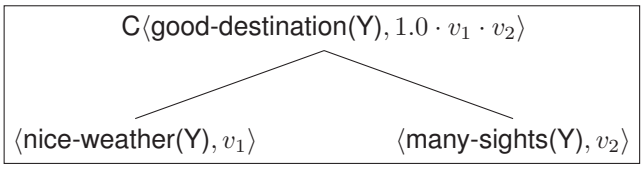
Here, the notation $t[A]$ means "the tree t that contains the node A somewhere". Likewise, $t[A/B]$ is to be read as "the tree t where the node A has been replaced by the node B".

Asking the query $\langle A, v\rangle$ corresponds to applying the transition rules to the initial computation tree $\langle A, v\rangle$. The computation ends *successfully* if a final computation tree is created, the truth value of the instantiated query can then be read off the root node. We will illustrate this with an example computation.

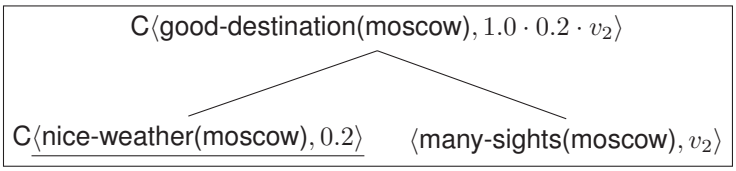
Example (continued). We start with the tree

$$\boxed{\langle \text{good-destination(Y)}, v\rangle}.$$

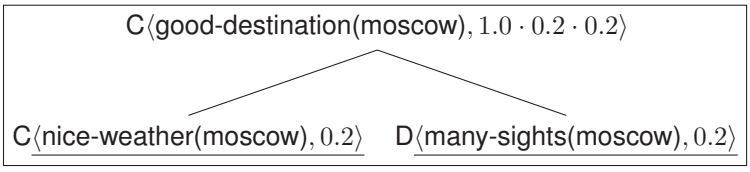
Applying the **Clause**-transition to the initial tree with the program clause good-destination(X) $\xleftarrow{1.0,\cdot}$ nice-weather(X), many-sights(X) yields



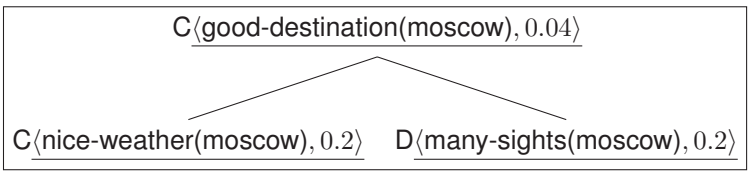
Now we apply **Clause** to the left child with nice-weather(moscow) ← 0.2:



Since there exists no clause whose head matches many-sights(moscow), we apply the **Default**-rule for many-sights to the right child.



In the last step, we finalise the root node.



The calculated truth value for good-destination(moscow) is thus 0.04. ◇

The actual operational semantics is now given by the truth values that can be derived in the defined transition system. This "canonical model" can be seen as a generalisation of the success set of a program.

Definition. Let P be a well-defined fuzzy logic program. The canonical model of P for $A \in B$ is defined as follows:

$$\mathsf{cm}(P) := \left\{ A \mapsto z_t v \;\middle|\; \begin{array}{l}\text{there exists a computation starting}\\ \text{with } \langle A, w\rangle \text{ and ending with a fi-}\\ \text{nal computation tree } t \text{ with root node}\\ \underline{\langle A, v\rangle}\end{array} \right\}$$

⌟

It can be verified that the canonical model $\mathsf{cm}(P)$ is indeed a model of P.

4 Implementation

RFuzzy is implemented as a package of the Ciao Prolog System [15]. It consists essentially of a set of rules that translate the RFuzzy Syntax to ANSI Prolog using the expansion of code of the packages in Ciao Prolog. The predicates of the program that have been declared as fuzzy get an additional argument that makes the truth value explicit. The resulting program can then be interpreted and executed "as usual".

Example (continued). We have no space to describe the implementation syntax used at RFuzzy, but it can be easily deduced from the implementation of our running example (not all the clauses are shown here).

```
nice_weather(madrid) value 0.8.
nice_weather(moscow) value 0.2.
```

ISBN: 978-989-95079-6-8

```
many_sights(madrid) value 0.6.
many_sights(sydney) value 0.6.

good_destination(X) cred (prod,1.0):~ prod
        nice_weather(X),
        many_sights(X).
```

The default value declarations are also very similar to the abstract syntax.

```
:- default(nice_weather/1, 0.5).
:- default(many_sights/1, 0.2).
:- default(good_destination/1, 0.3).
```

We use crisp predicates to represent types.

```
city(madrid).
city(moscow).
city(sydney).

:- set_prop(nice_weather/1) => city/1.
:- set_prop(many_sights/1) => city/1.
:- set_prop(good_destination/1) => city/1.
```

To ask queries to the system, we add a variable that is going to be instantiated with the truth value.

```
?- good_destination(moscow, V).

V = 0.04 ?
yes
```

But we cannot only ask for truth values of fully instantiated atoms. The real power of RFuzzy lies in the ability to provide constructive answers. For example, if we want to know "What is the best travel destination according to the knowledge base?" we just ask the following query. It looks for a destination `D` with a truth value `V` for which no destination with a higher truth value `V1` exists.

```
?- good_destination(D, V),
   \+ (good_destination(_, V1),
       V1 > V).

D = madrid,
V = 0.48 ?

yes
```

As we see, the system returns "Madrid" as best destination and thus answers a question with an object rather than a truth value. ◇

5 Conclusions and Future Work

We presented the operational semantics of the RFuzzy framework for Fuzzy Logic Programming and showed some features of the implementation [1] via an example. We finally remark that a least model semantics and a least fixpoint semantics for RFuzzy also have been defined and proven equivalent to the operational semantics shown here.

[1] A complete release of the implementation is available at `http://babel.ls.fi.upm.es/software/rfuzzy`

References

[1] P. Vojtas. Fuzzy logic programming. *Fuzzy Sets and Systems*, 124(1):361–370, 2001.

[2] R. C. T. Lee. Fuzzy Logic and the resolution principle. *Journal of the Association for Computing Machinery*, 19(1):119–129, 1972.

[3] M. Ishizuka and N. Kanai. Prolog-ELF incorporating fuzzy Logic. In *IJCAI*, pages 701–703, 1985.

[4] J. F. Baldwin, T. P. Martin, and B. W. Pilsworth. *Fril: Fuzzy and Evidential Reasoning in Artificial Intelligence*. John Wiley & Sons, 1995.

[5] D. Li and D. Liu. *A Fuzzy Prolog Database System*. John Wiley & Sons, New York, 1990.

[6] Z. Shen, L. Ding, and M. Mukaidono. Fuzzy resolution principle. In *Proc. of 18th International Symposium on Multiple-valued Logic*, volume 5, 1989.

[7] C. Vaucheret, S. Guadarrama, and S. Munoz-Hernandez. Fuzzy prolog: A simple general implementation using clp(r). In P.J. Stuckey, editor, *Int. Conf. in Logic Programming, ICLP 2002*, number 2401 in LNCS, page 469, Copenhagen, Denmark, July/August 2002. Springer-Verlag.

[8] S. Guadarrama, S. Munoz-Hernandez, and C. Vaucheret. Fuzzy Prolog: A new approach using soft constraints propagation. *Fuzzy Sets and Systems, FSS*, 144(1):127–150, 2004. ISSN 0165-0114.

[9] S. Munoz-Hernandez, C. Vaucheret, and S. Guadarrama. Combining crisp and fuzzy Logic in a prolog compiler. In J. J. Moreno-Navarro and J. Mariño, editors, *Joint Conf. on Declarative Programming: APPIA-GULP-PRODE 2002*, pages 23–38, Madrid, Spain, September 2002.

[10] H. T. Nguyen and E. A. Walker. *A first Course in Fuzzy Logic*. Chapman & Hall/Crc, 2000.

[11] J. Medina, M. Ojeda-Aciego, and P. Votjas. Multi-adjoint Logic Programming with continuous semantics. In *LPNMR*, volume 2173 of *LNCS*, pages 351–364, Boston, MA (USA), 2001. Springer-Verlag.

[12] J. Medina, M. Ojeda-Aciego, and P. Votjas. A procedural semantics for multi-adjoint Logic Programming. In *EPIA*, volume 2258 of *LNCS*, pages 290–297, Boston, MA (USA), 2001. Springer-Verlag.

[13] J. Medina, M. Ojeda-Aciego, and P. Votjas. A completeness theorem for multi-adjoint Logic Programming. In *International Fuzzy Systems Conference*, pages 1031–1034. IEEE, 2001.

[14] P.J. Morcillo and G. Moreno. Floper, a fuzzy logic programming environment for research. In *Proceedings of the Spanish Conference on Programming and Computer Languages, PROLE 2008*, Gijn, Spain, 2008.

[15] M. Hermenegildo, F. Bueno, D. Cabeza, M. Carro, M. García de la Banda, P. López-García, and G. Puebla. The CIAO Multi-Dialect Compiler and System: An Experimentation Workbench for Future (C)LP Systems. In *Parallelism and Implementation of Logic and Constraint Logic Programming*, pages 65–85. Nova Science, Commack, NY, USA, April 1999.

ISBN: 978-989-95079-6-8

Applying Bacterial Memetic Algorithm for Training Feedforward and Fuzzy Flip-Flop based Neural Networks

László Gál[1,2] János Botzheim[3,4] László T. Kóczy[1,4] Antonio E. Ruano[5]

1 Institute of Information Technology and Electrical Engineering,
Széchenyi István University, Győr, Hungary
2 Department of Technology, Informatics and Economy,
University of West Hungary Szombathely, Hungary
3 Department of Automation Széchenyi István University Győr, Hungary
4 Department of Telecommunication and Media Informatics
Budapest University of Technology and Economics, Budapest, Hungary
5 Centre for Intelligent Systems, FCT, University of Algarve, Portugal
Email: gallaci@ttmk.nyme.hu, {botzheim, koczy}@{sze, tmit.bme}.hu, aruano@ualg.pt

Abstract—*In our previous work we proposed some extensions of the Levenberg-Marquardt algorithm; the Bacterial Memetic Algorithm and the Bacterial Memetic Algorithm with Modified Operator Execution Order for fuzzy rule base extraction from input-output data. Furthermore, we have investigated fuzzy flip-flop based feedforward neural networks. In this paper we introduce the adaptation of the Bacterial Memetic Algorithm with Modified Operator Execution Order for training feedforward and fuzzy flip-flop based neural networks. We found that training these types of neural networks with the adaptation of the method we had used to train fuzzy rule bases had advantages over the conventional earlier methods.*

Keywords— Bacterial Memetic Algorithm, Fuzzy Flip-Flop, Levenberg-Marquardt method, Neural Network.

1 Introduction

Bacterial type evolutionary algorithms are inspired by the biological bacterial cell model [1,2]. The *Bacterial Memetic Algorithm* (BMA) is a recent method for fuzzy rule base extraction from input-output data for a certain system [7]. We have investigated its properties intensely and found some points where its performance in the fuzzy rule base identification could be improved. The recent bacterial type algorithms we proposed were named *Bacterial Memetic Algorithm with Modified Operator Execution Order* (BMAM), *Improved Bacterial Memetic Algorithm* (IBMA) and *Modified Bacterial Memetic Algorithm* (MBMA) [3,4]. They are both memetic algorithms and apply alternatively global and local search for identifying fuzzy rule bases from input-output data automatically when no human expert to define the rules is available.

Neural Networks belong to the Soft Computing area like Fuzzy Systems and Evolutionary Computing. They can be used for modeling a certain system where input-output data pairs exist. The neural networks are inspired by biological phenomena: the brain itself and other parts of the neural system.

Fuzzy Flip-Flops are extended forms of the binary flip-flops that are widely used in digital technics [5]. They use fuzzy logic operations instead of Boolean logic ones and require fuzzy inputs, furthermore they produce fuzzy outputs instead of digital values.

Our previous works were developing the *Bacterial Memetic Algorithm* applied for *fuzzy rule base identification* (FRBI) and investigating various types of *Fuzzy Flip-Flops* (F^3) used in *feedforward neural networks* (FFNN) as replacements of the neurons [6]. We trained the *Fuzzy Flip-Flop based Neural Networks* (FNN) with the *Levenberg-Marquardt* (LM) based training method as it is a widely used and accepted one. However, we faced the same problems with the LM based feedforward neural network training as in the fuzzy rule base identification. Therefore we have adopted the *Bacterial Memetic Algorithm with Modified Operator Execution Order* (BMAM) for training Neural Networks. Our goal was to improve the learning capabilities of feedforward neural networks with a bacterial type evolutionary approach.

In this paper we propose the adaptation of the BMAM for training feedforward neural networks, and we study and evaluate the respective results. From another aspect another paper was proposed here where we report on the findings of our investigations of the properties of different types of FNNs trained with BMAM [14].

2 Bacterial Memetic Algorithm with Modified Operator Execution Order (BMAM)

2.1 Bacterial Memetic Algorithm (BMA)

Bacterial Memetic Algorithm (BMA) is a very recent approach used for fuzzy rule base identification (FRBI) [7]. It combines global and local search. For the global search it uses bacterial type evolutionary approach and for the local search the *Levenberg-Marquardt* method is deployed. Previous work confirmed that the *Pseudo-Bacterial Genetic Algorithm* (PBGA) and the *Bacterial Evolutionary Algorithm* (BEA) were rather more successful in this area than the conventional genetic algorithms [1,2].

2.1.1 Bacterial mutation

PBGA is a special kind of *Genetic Algorithm* (GA) [8], it introduces a new "genetic" operation called *bacterial mutation*. For the algorithm, the first step is to determine how the problem can be encoded in a *bacterium* (*chromosome*). In case of modelling fuzzy systems the

ISBN: 978-989-95079-6-8

parameters of the model – all the breakpoints of the rule base – have to be encoded in the *chromosome*. The next step is to generate initial *bacteria* randomly. Then an optimization process is started utilizing mainly the *bacterial mutation*, in order to refine the model parameters.

The *bacterial mutation* operation tries to improve the parts of the *chromosomes*. Therefore each individual (*bacterium)* is selected one by one and a number of copies of the selected individual (*clones*) are created. Then the same part or parts are randomly chosen from all clones and it (they) is (are) mutated (except one single clone that remains unchanged). Mutation means to replace the part with a random value in a specified range.

After the mutations all the clones are evaluated (SSE, MSE, BIC) and the best clone is selected whose mutated part or parts are transferred to the other clones. Theoretically, this operation copies just a few parameters from one clone to the other clones (*gene transfer*), but in practice, the other clones will not differ from the best clone at the end at all. So, this operation can be done with discharging all the clones except the best one and then cloning further the best clone.

After the selection of the best clone and transferring its mutated part or parts to the other clones the above procedure is repeated until all the parts are mutated exactly once. The final best clone is remaining in the population and all the other clones are destroyed. The *bacterial mutation* cycle is done on the other individuals in the population e.g. in a parallel processing way.

At the end of the *complete bacterial mutation cycle* a new generation of bacteria is arisen.

The *Bacterial Evolutionary Algorithm* (BEA) is based on the PBGA supported by a new genetic operation called *gene transfer* operation. This operation can play an important role in the FRBI process as it establishes relationships among the individuals of the population (useful in somewhat changing environment) and is able to increase or decrease the number of the rules in a fuzzy rule base (useful in determining the appropriate size of the fuzzy rule base). Because this behaviour is not exploited in our investigations when training neural networks this operation is not described in detail.

2.1.2 The Levenberg-Marquardt method (LM)

The *Levenberg-Marquardt* (LM) method [9] is a gradient based iterative procedure. It is used for least squares curve fitting for a given set of empirical data, minimizing the *sum of squared error* function (SSE). It can be used for fuzzy rule extraction alone, but it generates only locally optimal rule base in the neighbourhood of the initial rules.

The *Error Back Propagation* algorithm (EBP or BP) was a great improvement in neural network research, but it has weak convergence rate. The LM algorithm is more complex and requires more computational effort than the BP, but it has much better convergence rate properties. Therefore the LM algorithm is one of the most popular training functions for feedforward back propagation networks.

2.1.3 Bacterial Memetic Algorithm (BMA)

Memetic algorithms combine evolutionary and local search methods [10]. The evolutionary part is able to find the global optimum region, but is not suitable to find the accurate minimum in practice. The gradient based part is able to reach the accurate optimum, but is very sensitive to the initial position in the search space and is unable to avoid the local optimum. Combining global and local search is expected to be beneficial.

Bacterial Memetic Algorithm (BMA) combines the *Bacterial Evolutionary Algorithm* (BEA) and the *Levenberg-Marquardt* (LM) method. In the past we used it for fuzzy rule base extraction, among others.

The BMA integrates its two components, the BEA and the LM method in the following way:

1. *Bacterial Mutation* operation for each individual,
2. *Levenberg-Marquardt* method for each individual,
3. *Gene Transfer* operation for a partial population.

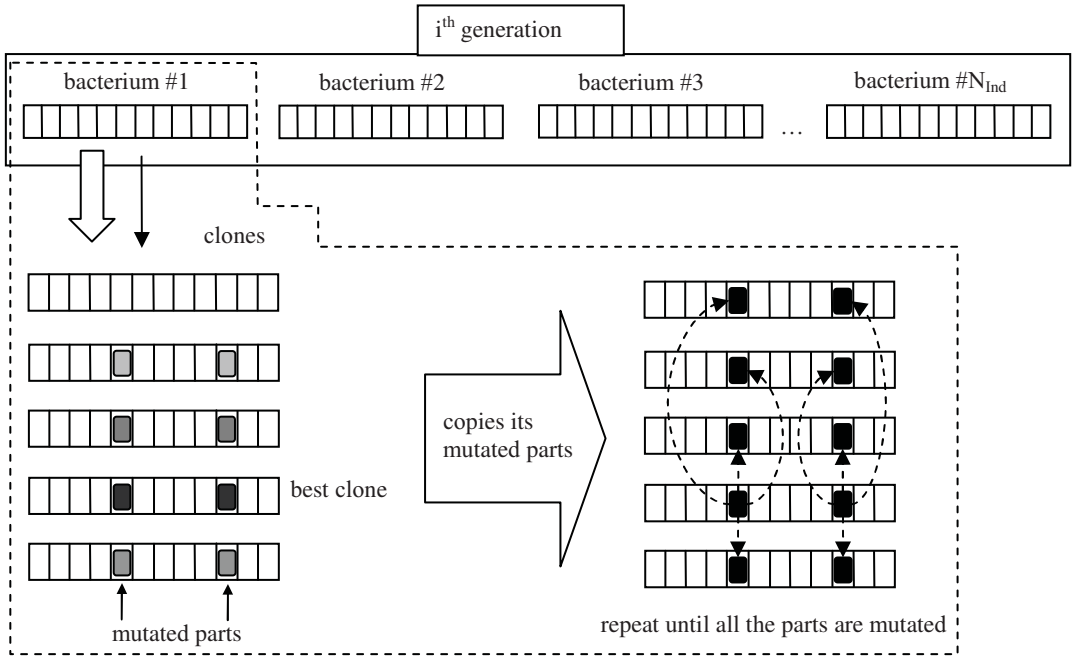


Figure 1: Bacterial mutation (one individual)

ISBN: 978-989-95079-6-8

This way the LM method is nested into the BEA, so that local search is done for every global search cycle.

2.2 Bacterial Memetic Algorithm with Modified Operator Execution Order (BMAM)

Although BMA provides a very good speed of convegrenve towards the optimal model parameters there are some points of the algorithm where the performance could be increased. We proposed new techniques to improve its performance. Some of them contain modifications that are not useful in training FFNNs (handling the *knot order violation* that can occur in applying LM for FRBI) (IBMA, MBMA) [3]. Another improvement to BMA is the *Bacterial Memetic Algorithm with Modified Operator Execution Order* [4] which exploits the *Levenberg-Marquardt* method more efficiently.

Instead of applying the LM cycle *after* the *bacterial mutation* as a separate step, the modified algorithm executes several LM cycles *during* the *bacterial mutation* after *each mutational* step.

The *bacterial mutation* operation changes one or more parameters of the modeled system randomly, and then it checks whether the model obtained by this way performs better than the previous models or the models that have been changed concurrently this way in the other clones. The mutation test cycle is repeated until all the parameters of the model have gone through the bacterial mutation.

In the mutational cycle it is possible to gain a temporary model that has an instantaneous fitness value that is worse than the one in the previous or the concurrent models. However, it is potentially better than those, because it is located in such a region of the search space that has a better local optimum than the other models do. In accordance to this, if some Levenberg-Marquardt iterations are executed after *each* bacterial mutational step, the test step is able to choose some potentially valued clones that could be lost otherwise.

In the *Bacterial Memetic Algorithm with Modified Operator Execution Order*, after *each mutational step* of *every single bacterial mutation iteration* several LM iterations are done. Several tests have shown it is enough to run just 3 to 5 of LM iterations per mutation to improve the performance of the whole algorithm. The usual test phase of the *bacterial mutation* operation follows after the LM iterations. After the complete *modified bacterial mutation* follows the LM method that is used in the original BMA, where more, e.g. 10 iterational steps, are done with all the individuals of the population towards reaching the local optimum. After all this the *gene transfer* operation is executed if needed.

3 Fuzzy flip-flops (F^3)

Fuzzy flip-flops are extended forms of binary flip-flops used in the conventional digital technics. We have dealt with the fuzzy extensions (complements) of the binary J-K and D flip-flops. Various types of fuzzy flip-flops are implemented and tested (set, reset type and the general type using the unified equation; J-K, D and Choi type D; based on minmax, algebraic, Yager, Dombi, Hamacher and Frank t-norms and co-norms, resp.) [11]. Because of an interesting property some fuzzy flip-flops can be used for implementing a sigmoid like transfer function and so constructing Multilayer Perceptron Neural Networks.

In our previous works we studied the behavior of various type fuzzy flip-flops, illustrating their characteristics by their respective graphs. We proposed also the concept of *fuzzy flip-flop based neural networks* and investigated their function approximation capabilities [6, 12].

4 Fuzzy flip-flop based feedforward neural networks (FNN)

In our team extensive research was done with the leadership of R. Lovassy in the field of fuzzy flip-flops. As we mentioned it before, various fuzzy norms can be used for building *fuzzy flip-flop based neural networks* (FNNs). The basic idea was to substitute the fuzzy flip-flops with sigmoidal transfer function instead of traditional neurons. The flip-flops are based on various norms, consequently, their transfer functions have different slopes. Fixing the value of the present state Q (in the characteristical equation), often we obtained "good" enough sigmoidal transfer function character [6]. First of all, to train this kind of neural network with a usual training method BP or LM the derivatives of these transfer functions have to be also calculated. Then the FNN can be used and trained in the usual way. We found that the FNNs we created had good approximation properties. [12].

5 Training feedforward neural networks

With an appropriate transfer function and its derivative the *Error Back Propagation* algorithm (BP) can be used for training feedforward neural networks (FFNN). However, it has weak convergence rates.

The LM algorithm is more complex and requires more computational effort than the BP one, but it has much better convergence rates. The LM algorithm is one of the most popular training methods for feedforward neural networks despite of its higher memory requirements and higher complexity.

The training of the FFNNs begins with the random generation of initial weights and biases. Then the training method selected is applied. An update vector is generated that has to be applied for the vector that contains the weights and biases.

When using BP or LM based training methods one faces the drawback of these local searchers described in the next section.

6 Using BMAM in training FFNNs

Although the LM method based training of the neural networks works much more efficiently than the BP based one it has all drawbacks of the local search methods. The training is very sensitive to the (parameter's) initial position of the search space. An inconveniently generated random parameter set with the initial weights and biases determines a hardly trainable neural network with a weak performance at the end of the LM method based training procedure. This is because the LM method is a local searcher and thus it is unable to avoid the local minima.

We decided to apply bacterial type evolutionary algorithms because they proved to be rather successful in our previous

ISBN: 978-989-95079-6-8

works, better than the other evolutionary approaches. We preferred the *Bacterial Memetic Algorithm with Modified Operator Execution Order* because it converged faster than the original BMA. (And contained not only FRBI related improvements, like IBMA and MBMA do.)

We did not implement here the *gene transfer* operation because it was not useful with the neural network training we have done (in not changing environments, there was no need to change the structure of the NN or the number of the neurons).

The detailed steps of the BMAM used for NN training are described below:

1. Create the initial population – neural networks with two hidden layers – and initialize the neural network's input, layer weights and biases randomly as before a usual LM training procedure. Each individual contains the weights and biases – the parameters of the model – encoded in the chromosome. In a 1-4-3-1 NN the number of the parameters to be encoded are 2*4+4*3+2*3+1 = 27 parameters per individual.
2. Apply the *modified bacterial mutation* for each individual.
    a. Each individual is selected one by one.
    b. N_{Clones} copies of the selected individual are created ("*clones*").
    c. Choose the same part or parts randomly from the clones and mutate it (except one single clone that remains unchanged during this mutation cycle).
    d. Run some conventional LM method based NN training iterations (3-5 epochs).
    e. Select the best clone (simulate and evaluate the NNs) and transfer *all* of its parts to the other clones.
    f. Repeat the part selection-mutation-*LM*-selection-transfer cycle until all the parts are mutated, improved and tested.
    g. The best individual is remaining in the population, all other clones are deleted.
    h. This process is repeated until all the individuals have gone through the *modified bacterial mutation*.
3. Apply the LM method based NN training to each individual (e.g. 10 epochs per individual per generation)
4. Repeat the procedure above from the *modified bacterial mutation* step until a certain termination criterion is satisfied (e.g. maximum number of generations = 20 generation).

The experimental setup was:

- General PC (2GHz), Windows XP, Matlab
- Test function: $f(x) = \frac{\sin(c_1 \cdot x) \cdot \sin(c_2 \cdot x)}{c_3} + c_4$
- $c_1 = 0.2, c_2 = 0.07, c_3 = 2, c_4 = 0.5$
- Number of individuals in the population: 5
- Number of clones: 5.

We tested the new training algorithm in two ways.

6.1 Test 1

In the first test group we applied the new BMAM based NN training. We created a 1-4-3-1 feedforward neural network with the usual sigmoid transfer function and with selected fuzzy flip-flop based neurons. Our goal was to investigate the improvement of BMAM based training over the conventional LM based one so four transfer functions were selected: sigmoid (tansig), Dombi Fuzzy D Flip-Flop (Dombi DF^3), Frank Fuzzy D Flip-Flop (Frank DF^3) and Frank Choi-type Fuzzy D Flip-Flop (Frank CDF^3).

Our goal was here to train a NN that is hard to be trained [12]. The number of neurons (4 and 3 in the hidden layers) was relatively low. It makes possible to recognise the performance improvement (MSE) better, to see that the model complexity may be reduced with the better training, and to avoid overfitting.

We ran 30-30 trainings for each case mentioned above. Then the maximum, minimum, median and mean MSE values calculated.

In our previous work we chose the median to characterize the trainability of the FFF based NNs because in case of training with the LM based way there were several unsuccessful trainings where the final model was unusable and produced too high MSE. That is why we had to analyse 30-100 runs. The best value (minimum MSE) was more or less randomly good so it could not be used as a reliable value for indicate the trainability. In case of the mean value a single one unsuccessful training deteriorates many very successful training results. With using the median this random extreme values could be avoided.

Table 1 and 2 contains the results of these tests. One can see that using BMAM result much more better quality models (lower MSE). Using BMAM results lover maximum MSE values than median or mean values of the LM based training respectively. Furthermore the median and mean MSE values of the BMAM based training are very close to the minimum values of the LM based training respectively.

Table 1: MSE values of LM based training

LM based	Max	Min	Median	Mean
Tansig	0.06452	1.2×10^{-5}	0.00712	0.01891
Dombi DF^3	0.11962	0.04263	0.05732	0.06045
Frank DF^3	0.06644	0.00459	0.04642	0.04159
Frank CDF^3	0.06645	0.00593	0.05486	0.04697

Table 2: MSE values of BMAM based training

BMAM	Max	Min	Median	Mean
Tansig	0.00180	3.14×10^{-7}	2.38×10^{-5}	0.00034
Dombi DF^3	0.03994	0.02342	0.03362	0.03294
Frank DF^3	0.01091	0.00187	0.00680	0.00716
Frank CDF^3	0.02519	0.00460	0.00833	0.00933

Figure 2 to 9 show MSE histograms of 30-30 runs with various transfer functions and training methods. One can see that if using BMAM there is no need to run 30 or 100 complete training cycles to gain an excellent quality model because every BMAM trained model have very low MSE value.

ISBN: 978-989-95079-6-8

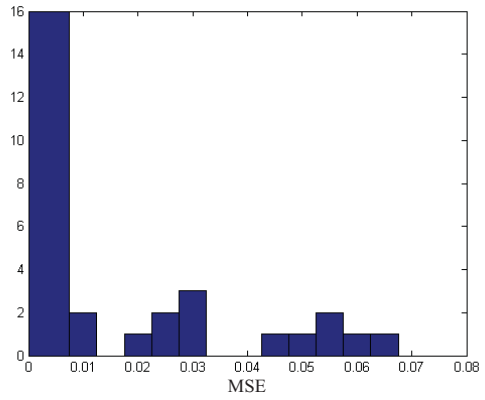


Figure 2: LM – Tansig NN histogram

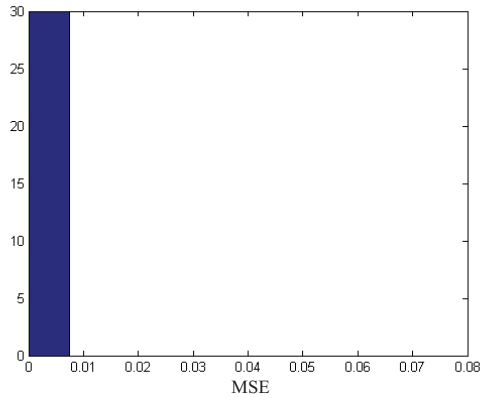


Figure 3: BMAM – Tansig NN histogram

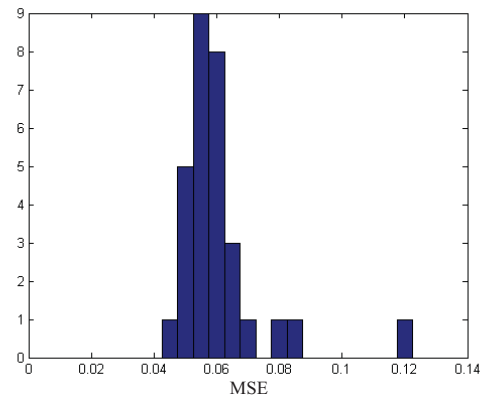


Figure 4: LM – Dombi DF^3 histogram

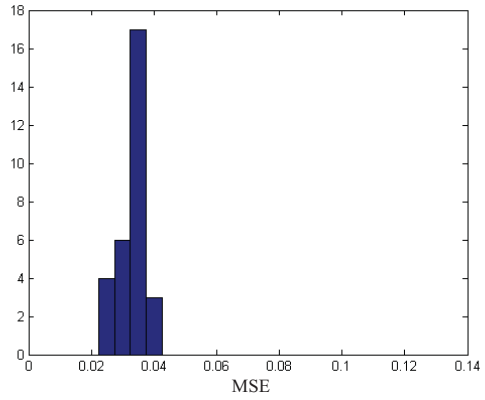


Figure 5: BMAM – Dombi DF^3 NN histogram

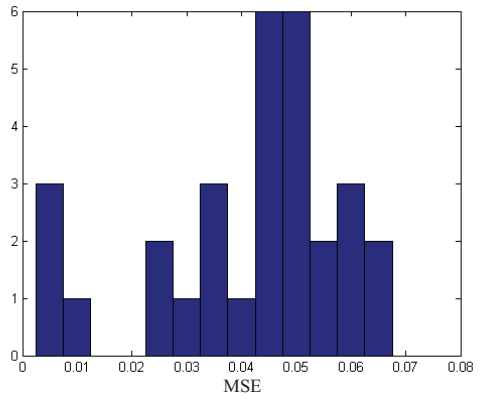


Figure 6: LM – Frank DF^3 NN histogram

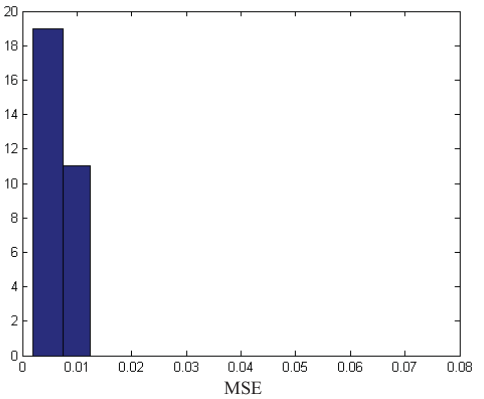


Figure 7: LM – Frank DF^3 histogram

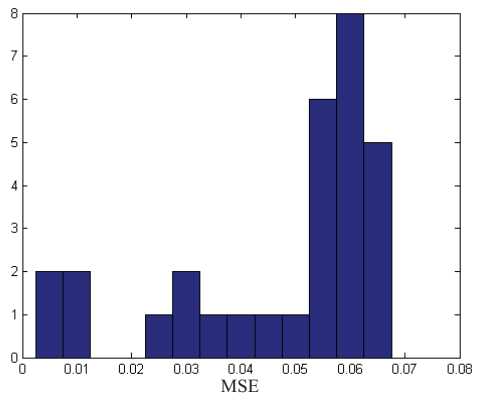


Figure 8: LM – Frank CDF^3 NN histogram

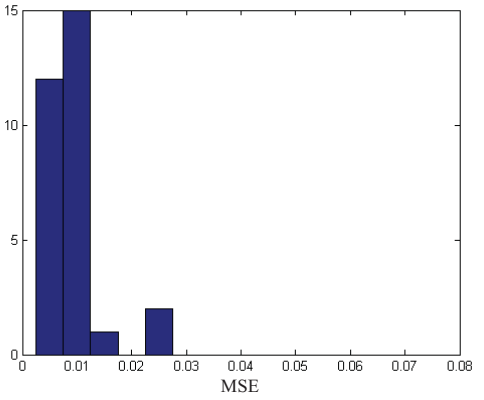


Figure 9: BMAM – Frank CDF^3 NN histogram

ISBN: 978-989-95079-6-8

6.2 Test 2

In the second group of tests we utilized the BMAM to identify quasi optimal parameter values of various types of fuzzy flip-flops used in FNNs. 1-8-8-1 neural networks were created because the good trainability was much more important than the lower complexity of the model here. This way the optimal parameter values are easier to identify.

Therefore we enhanced the capability of the BMAM based training method in a manner that the parameter (internally fixed Q values) of the fuzzy flip-flop used in Fuzzy Flip-Flop Neural Networks (FNN) can be encoded into the *chromosome*. This way it participates in the *bacterial mutation* cycle so the quasi optimal value of this parameter can be identified at the end of the BMAM based training.

Because this parameter is not affected by the LM training we applied two different versions of the *bacterial mutation* especially for this parameter. The first one is the original *bacterial mutation* (generate random values in the range of [0, 1]), while the second one increments or decrements the current fixed Q value with a very fine random step.

Table 2 shows the expected ranges and the quasi-optimal internally fixed Q values of several FNNs identified by the BMAM training method. The expected ranges were derived from our previous work [12].

Table 2: Expected ranges and fixed Q values by BMAM

Type of FNN	Expected range	Fixed Q value identified by BMAM
Algebraic JK FF	0 – 0.4	0.25
Algebraic D FF	~0.1,~0.5, ~0.9	0.91
Algebraic C D FF	<0.15, 0.4 – 0.6, >0.85	0.53
Dombi D FF	<0.1 or >0.9	0.924
Frank D FF	0.25 – 0.45	0.31

Further investigations will be focused on using BMAM based training method to identify of the other variable parameters of Yager, Dombi, Hamacher, Frank norms based FNNs.

7 Conclusions

In this paper we introduced the adaptation of *the Bacterial Memetic Algorithm with Modified Operator Execution Order* for *training feedforward neural networks*, especially *neural networks* built from *Fuzzy Flip-Fops* (F^3s).

We applied this new approach to training neural networks and *fuzzy flip-flop based neural networks*. Our goal was to get a quasi-optimal result with only a single one or a very low number of training sequences whose error does not exceed (or very rarely exceeds) an acceptable level. Despite the usual tradeoffs between the complexity and accuracy [13] this way there is no need to run a few hundred of training cycles to get an acceptable model.

Our tests have shown that BMAM used for training FFNNs and fuzzy flip-flop based FFNNs is a very successful tool. Although it requires more computational effort than the conventional training methods it produces a higher quality model (so the complexity of the model can be reduced) with only one training cycle.

Furthermore we enhanced the capability of the BMAM based training method in a manner that the parameter or parameters of the fuzzy flip-flop used in *Fuzzy Flip-Flop based Neural Networks* (FNN) can be encoded into the chromosome. This way it participates in the *bacterial mutation* cycle so the quasi optimal values of these parameters can be identified at the end of the BMAM based training.

Acknowledgment

This paper was supported by the Széchenyi University Main Research Direction Grant 2009, National Scientific Research Fund Grant OTKA T048832 and K75711, SEK Scientific Grant 2009, and the National Office for Research and Technology.

References

[1] Nawa, N. E., Hashiyama, T., Furuhashi, T. and Uchikawa, Y., *A study on fuzzy rules discovery using pseudo-bacterial genetic algorithm with adaptive operator*, Proceedings of IEEE Int. Conf. on Evolutionary Computation, ICEC'97, 1997.

[2] Nawa, N. E. and Furuhashi, T., *Fuzzy Systems Parameters Discovery by Bacterial Evolutionary Algorithms*, IEEE Transactions on Fuzzy Systems 7, 1999, pp. 608-616.

[3] Gál, L., Botzheim, J. and Kóczy, L. T., *Improvements to the Bacterial Memetic Algorithm used for Fuzzy Rule Base Extraction*, Computational Intelligence for Measurement Systems and Applications, CIMSA 2008, Istanbul, Turkey, 2008, pp. 38-43.

[4] Gál, L., Botzheim, J. and Kóczy, L. T., *Modified Bacterial Memetic Algorithm used for Fuzzy Rule Base Extraction*, 5th International Conference on Soft Computing as Transdisciplinary Science and Technology, CSTST 2008, Paris, France, 2008.

[5] K. Ozawa, K. Hirota and L. T. Kóczy, *Fuzzy flip-flop*, In: M. J. Patyra, D. M. Mlynek, eds., Fuzzy Logic. Implementation and Applications, Wiley, Chichester, 1996, pp. 197-236.

[6] R. Lovassy, L. T. Kóczy and L. Gál, *Multilayer Perceptron Implemented by Fuzzy Flip-Flops*, IEEE World Congress on Computational Intelligence, WCCI 2008, Hong Kong, pp. 1683-1688.

[7] Botzheim, J., Cabrita, C., Kóczy, L. T. and Ruano, A. E. , *Fuzzy rule extraction by bacterial memetic algorithm*, IFSA 2005, Beijing, China, 2005, pp.1563-1568.

[8] Holland, J. H., *Adaptation in Nature and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, The MIT Press, Cambridge, MA, 1992.

[9] Marquardt, D., *An Algorithm for Least-Squares Estimation of Nonlinear Parameters*, SIAM J. Appl. Math., 11, 1963, pp. 431-441.

[10] Moscato, P., *On evolution, search, optimization, genetic algorithms and martial arts: Towards memetic algorithms*, Technical Report Caltech Concurrent Computation Program, Report. 826, California Institute of Technology, Pasadena, California, USA,1989.

[11] R. Lovassy, L. T. Kóczy and L. Gál, *Analyzing Fuzzy Flip-Flops Based on Various Fuzzy Operations*, Acta Technica Jaurinensis Series Intelligentia Computatorica vol. 1, no. 3, 2008, pp. 447-465.

[12] R. Lovassy, L. T. Kóczy and L. Gál, *Function Approximation Capability of a Novel Fuzzy Flip-Flop Based Neural Network*, IJCNN 2009 Atlanta - accepted.

[13] L. T. Kóczy and A. Zorat, *Fuzzy systems and approximation*, Fuzzy Sets and Systems, Vol.85, pp. 203-222, 1995.

[14] R. Lovassy, L. T. Kóczy and L. Gál, *Optimizing Fuzzy Flip-Flop Based Neural Networks by Bacterial Memetic Algorithm*, IFSA/EUSFLAT 2009, Lisbon, Portugal, 2009 - accepted

ISBN: 978-989-95079-6-8

Interactive Fuzzy Modeling by Evolutionary Multiobjective Optimization with User Preference

Yusuke Nojima, Hisao Ishibuchi

Department of Computer Science and Intelligent Systems, Osaka Prefecture University
1-1 Gakuen-cho, Naka-ku, Sakai, Osaka 599-8531, Japan
Email: {nojima, hisaoi}@cs.osakafu-u.ac.jp

Abstract— *One of the new trends in genetic fuzzy systems (GFS) is the use of evolutionary multiobjective optimization (EMO) algorithms. This is because EMO algorithms can easily handle two conflicting objectives (i.e., accuracy maximization and complexity minimization) when we design accurate and compact fuzzy rule-based systems from numerical data. Since the main advantage of fuzzy rule-based systems compared with other non-linear ones is their linguistic interpretability, the design of fuzzy rule-based systems can be viewed as linguistic data mining from numerical data. From the data mining point of view, the required knowledge strongly depends on its user. That is, the interpretability of fuzzy rule-based systems should be evaluated by taking into account the user's preference. Although there exist a number of interpretability measures in the literature, users usually do not know which measure represents their preference beforehand. In this paper, we propose interactive fuzzy modeling by evolutionary multiobjective optimization with user's preference. User's preference is represented by several satisfaction level functions which can be interactively modified by the user. The user's preference is used as one of multiple objectives in an EMO algorithm. As a case study, we apply our approach to real world time-series data of land price movements in Japan and demonstrate a user interface of our approach.*

Keywords— evolutionary multiobjective optimization, fuzzy modelling, interactive evolutionary computation, user preference.

1 Introduction

There are two major goals in the design of fuzzy rule-based systems: accuracy maximization and complexity minimization. Since the mid-1990s, a large number of approaches have been proposed for improving the accuracy of fuzzy rule-based systems while maintaining their interpretability [1-22]. Genetic algorithms have been frequently used under the name of genetic fuzzy systems (GFS) [23-25]. We can easily handle two conflicting objectives: accuracy maximization and complexity minimization by the weighted sum of them or multiobjective formulations using evolutionary multiobjective optimization (EMO) algorithms [26-28]. One of the hottest issues in GFS is how to measure the interpretability of fuzzy rule-based systems [29-34]. A number of interpretability measures have been already proposed and implemented in GFS. Interpretability is, however, very subjective for users. Let us assume the following two rule sets.

[Fuzzy rule-based system A]
If x_1 is *big* then y is 10,
If x_3 is *small* then y is 5,
If x_2 is *very small* then y is 2,
If x_3 is *big* then y is 11.

[Fuzzy rule-based system B]
If x_1 is *big* and x_2 is *big* and x_3 is *small* then $y=x_1+5x_2+9$,
If x_1 is *very big* and x_2 is *very small* then $y = 2x_2+2$,
If x_1 is *very small* and x_2 is *small* and x_3 is *big* then y=12.

If we assume that a rule set with a small number of rules is interpretable, the fuzzy rule-based system B is more interpretable. However, the fuzzy rule-based system A seems to be more interpretable with respect to the rule length and the rule type. From this observation, we can say that the interpretability is totally dependent on user's preference (see [35] for more detailed discussions). There is another problem on interpretability. Users usually know which is interpretable for them only after comparing among some alternative fuzzy rule-based systems. That is, an interactive optimization process must be needed for the users.

In our former studies [36, 37], we have proposed an interactive genetic fuzzy rule selection method for pattern classification problems. A preference function is used as one objective function in an EMO algorithm. The preference function is composed of several satisfaction level functions. The reason why we used several satisfaction level functions is that users don't know appropriate criteria and their priorities among them beforehand. These satisfaction level functions are interactively modified during the evolution under the framework of interactive evolutionary computation [38-41]. In this paper, we apply this idea to a fuzzy modeling problem. We deal with time-series data of land price movements and demonstrate the prototype of a user interface.

This paper is organized as follows: Section 2 explains fuzzy modeling and its interactive optimization process. Section 3 explains the tackled problem and demonstrates the effectiveness of our method with a prototype of our user interface. Section 4 concludes this paper.

2 Interactive Fuzzy Modeling

2.1 Fuzzy Modeling

In this paper, for an n-input and single-output nonlinear function $y = y(\mathbf{x})$, we use the following fuzzy if-then rules:

$$\text{Rule } R_k : \text{If } x_1 \text{ is } A_{k1} \text{ and } ... \text{ and } x_n \text{ is } A_{kn} \text{ then } y \text{ is } B_k, \quad k = 1, 2, ..., N, \tag{1}$$

where x_i is the i-th input variable of an n-dimensional input vector $\mathbf{x} = (x_1, \ldots, x_n)$, y is an output variable, k is a rule index, A_{ki} is an antecedent linguistic label (e.g., *small* and *large*) for

ISBN: 978-989-95079-6-8

x_i, B_k is a consequent linguistic value, and N is the total number of fuzzy if-then rules.

The following fuzzy reasoning method has been frequently used in fuzzy rule-based systems since its first proposal in a neuro-fuzzy system [42]:

$$\hat{y}(\mathbf{x}) = \frac{\sum_{k=1}^{N} \mu_k(\mathbf{x}) \cdot b_k}{\sum_{k=1}^{N} \mu_k(\mathbf{x})}, \quad (2)$$

where $\mu_k(\mathbf{x})$ is the compatibility grade of the fuzzy if-then rule R_k with the input vector $\mathbf{x}$, and b_k is a representative real number of the consequent linguistic value B_k. The compatibility grade $\mu_k(\mathbf{x})$ is usually calculated by the product operation as

$$\mu_k(\mathbf{x}) = \mu_{k1}(x_1) \times \cdots \times \mu_{kn}(x_n), \quad (3)$$

where $\mu_{ki}(\cdot)$ is the membership function of the antecedent linguistic value A_{ki}. The representative real number b_k can be viewed as a result of the defuzzification of the consequent linguistic value B_k.

The fuzzy reasoning method in (2) can be viewed as a simplified version of the Takagi-Sugeno (TS) model where a linear function is used in the consequent part of each fuzzy if-then rule. The simplified fuzzy reasoning method in (2) has several advantages. For example, its reasoning mechanism is very simple, and it is suitable for gradient-based learning algorithms.

Since we use multiple granularities of fuzzy sets for an input vector, sometime the effect of specific rules becomes lower due to general rules. For giving specific rules more weight, we use an idea of inclusion-based fuzzy reasoning [43]. We extend the inclusion relation $R_A \subset R_B$ in [43]. When only the two rules R_k and R_q with the relation $|\mathbf{A}_k| > |\mathbf{A}_q|$, are compatible with the input vector $\mathbf{x}$, the specific rule R_q is mainly used in fuzzy reasoning. That is, the weight of the general rule R_k is discounted. Our idea is to determine the amount of the discount for R_k using the compatibility grade $\mu_q(\mathbf{x})$ of the specific rule R_q with the input vector $\mathbf{x}$. More specifically, the weight of R_k is defined as $(1-\mu_q(\mathbf{x}))$. When the specific rule R_q is fully compatible with the input vector $\mathbf{x}$, the weight of the general rule R_k is zero. This means that R_k has no effect on the calculation of the estimated output value $\hat{y}(\mathbf{x})$. On the other hand, when the compatibility grade of R_q with $\mathbf{x}$ is very small, the amount of the discount for R_k is also very small. In this case, R_k has almost the same weight as R_q. Since the general rule R_k may include multiple rules, its weight is defined as

$$w(R_k, \mathbf{x}) = \prod_{\substack{q=1 \\ |\mathbf{A}_q| < |\mathbf{A}_k|}}^{N} (1 - \mu_q(\mathbf{x})). \quad (4)$$

When there are no compatible fuzzy if-then rule smaller than R_k, $w(R_k, \mathbf{x})$ is specified as $w(R_k, \mathbf{x}) = 1$ because the weight of R_k should not be discounted in this case. It should be noted that the weight of each rule depends on the compatibility grades of other rules with the input vector $\mathbf{x}$. This means that the weight is context-dependent. Different weights are assigned to the same rule for different input vectors. Moreover, the same rule may have different weights for the same input vector in different rule bases because the weight of each rule depends on other rules.

Using the rule weight $w(R_k, \mathbf{x})$ of each fuzzy if-then rule R_k, our inclusion-based fuzzy reasoning method is written as

$$\hat{y}(\mathbf{x}) = \frac{\sum_{k=1}^{N} w(R_k, \mathbf{x}) \cdot \mu_k(\mathbf{x}) \cdot b_k}{\sum_{k=1}^{N} w(R_k, \mathbf{x}) \cdot \mu_k(\mathbf{x})}. \quad (5)$$

2.2 Multiobjective Fuzzy Rule Selection for Modeling

We use a simple two-stage method for designing rule sets. In the first phase, a large number of candidate rules are generated from the possible combinations of membership functions. The consequent real number is specified as the weighted average of output values of compatible input-output pairs as

$$b_k = \frac{\sum_{p=1}^{m} \mu_{\mathbf{A}_k}(\mathbf{x}_p) \cdot y_p}{\sum_{p=1}^{m} \mu_{\mathbf{A}_k}(\mathbf{x}_p)}, \quad (6)$$

where $\mu_{\mathbf{A}_k}(\mathbf{x}_p)$ is the compatibility grade of the input vector $\mathbf{x}_p$ with the antecedent part $\mathbf{A}_k$ of the linguistic rule R_k.

In the second phase of our rule selection, a number of fuzzy rule sets are selected by a multiobjective genetic algorithm. Any subset S of the N candidate rules can be represented by a binary string of length N as

$$S = s_1 s_2 \cdots s_N, \quad (7)$$

where s_k=1 and s_k=0 represent the inclusion of the k-th candidate rule R_k in S and the exclusion of R_k from S.

Each fuzzy rule set S is evaluated by the three objectives:

$f_1(S)$: the total square error by S,

$f_2(S)$: the number of selected fuzzy rules in S,

$f_3(S)$: the overall user preference for S.

The first and second objectives have been frequently used and correspond to accuracy maximization and complexity minimization, respectively. The first objective is calculated by,

$$f_1(S) = \sum_{p=1}^{m} (y_p - \hat{y}_p(\mathbf{x}_p))^2. \quad (8)$$

The third objective $f_3(S)$ is the newly proposed objective in this paper. We explain it in the next subsection. The problem formulation of multiobjective genetic fuzzy rule selection is written as

Minimize $f_1(S)$ and $f_2(S)$, and maximize $f_3(S)$. (9)

ISBN: 978-989-95079-6-8

We use NSGA-II of Deb et al. [27] to search for a number of non-dominated fuzzy rule-based systems with respect to these three objectives. In this paper, uniform crossover and biased bit-flip mutation are used in NSGA-II. The biased mutation is that a larger probability is assigned to the mutation from 1 to 0 than that from 0 to 1.

Figure 1 shows the whole procedure of the proposed method. We specify an interval (i.e., the number of generations) for internal evaluations. During this interval, the satisfaction level functions are not changed. After the interval, the user checks some of non-dominated rule sets and modifies the satisfaction level functions. Then another internal evaluation process starts. By repeating this interactive process, the user can specify the own preference and find the rule set with the high user preference value.

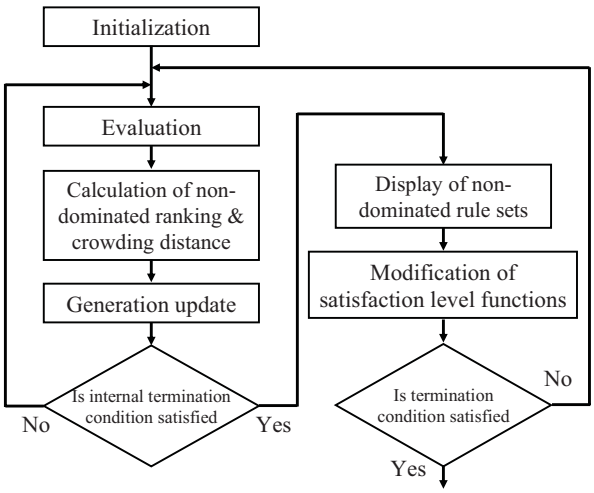


Figure 1: The whole procedure of the proposed method.

2.3 *Preference Function*

In our former study [36, 37], we have proposed a preference function composed of several satisfaction level functions. The inputs for the satisfaction level functions are criteria on the accuracy and interpretability of fuzzy rule sets. Each satisfaction level function is represented by a trapezoidal function in Fig. 2. $g^r(S)$ is the value of r-th criterion for the rule set S. $u^r(g^r(S))$ is the output of r-th satisfaction level function. Users specify the preference and priority for each criterion by moving the point B in Fig. 2. That is, the u_x and u_y of the point B mean the preference and priority for the criterion, respectively. The u_x can be also regarded as the maximum criterion level.

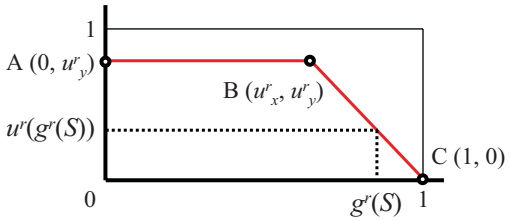


Figure 2: A trapezoidal function for representing satisfaction level functions of each criterion.

The satisfaction level function can be viewed as the requirement level. During evolution, users can modify the satisfaction level function according to the temporally obtained non-dominated rule sets.

The third objective function $f_3(S)$ for an overall user's preference is calculated by

$$f_3(S) = \sum_{r=1}^{N_C} u^r\left(g^r(S)\right), \tag{10}$$

where N_c is the number of criteria.

3 Case Study

3.1 *Problem Description*

We apply our proposed method to a simple fuzzy modelling problem for time-series data. The data we used in this paper is the land price movements of the three major metropolitan areas in Japan available from Ministry of Land, Infrastructure, Transport and Tourism webpage (http://www.mlit.go.jp/index_e.html). The data includes the land price movements from 1980 to 2000. In this period, the bubble economy was a big problem: the increased demand for office buildings in city centres due to internationalization and informatization.

The data is composed of 63 pairs of two inputs (i.e., *area* and *year*) and one output (*change of land price*). For simplicity, we normalized the input attributes into [0, 1]x[0, 1] space. We used seven categorical values (all possible combinations) for *area* attribute. For year attribute, we used 48 fuzzy membership functions shown in Fig. 4 and *don't care* condition. Each peak of triangular membership functions corresponds to one of years. These partitions could be understandable (e.g., 90's, Mid of 80's, around 1988).

From this data, 343 fuzzy if-then rules were generated as candidate rules. Thus, the search space is 2^{343}.

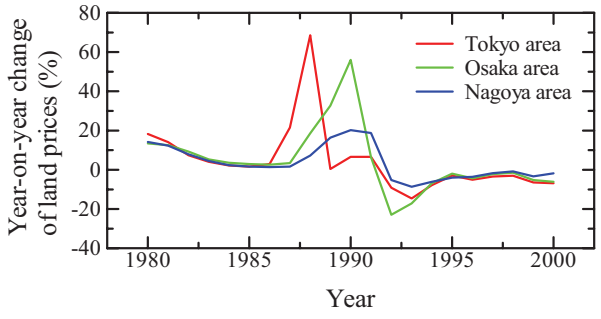


Figure 3: Year-on-year change of land prices of three major metropolitan areas in Japan from 1980 to 2000.

3.2 *Criteria for Representing User's Preference*

There are a lot of interpretability measures in the literature. In this case study, we used four simple criteria like:

- Maximum square error by S,
- Overlap among antecedent sets in S,
- Total square error by S,
- The number of fuzzy rules in S.

ISBN: 978-989-95079-6-8

The maximum square error is calculated by

$$g^1(S) = \max_{p=1,\ldots,m} (y_p - \hat{y}_p(\mathbf{x}_p))^2 . \quad (11)$$

When a user gives a high priority to this criterion, the larger changes could be fitted by a fuzzy rule set.

We normalized each value of four criteria within the valid ranges based on the distribution values in the pre-simulation without user's preference. The reason why we used *the total square error* and *the number of fuzzy rules* as the criteria for user's preference is to reduce the search space based on the user's preference.

The choice of interpretability measures is future research issues. The correlation among measures must be examined.

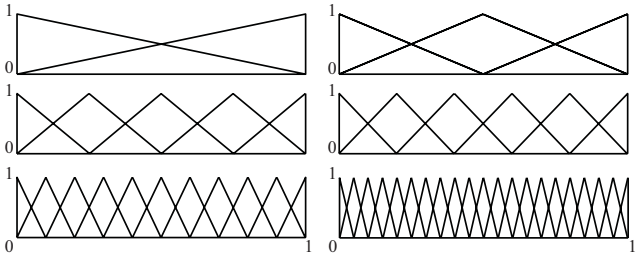


Figure 4: Fuzzy partitions with different granularities for *Year* attribute.

3.3 Prototype of User Interface

Figure 5 shows the prototype of our user interface. There are two windows. The left one represents the actual land price movement and the inferred land price movement calculated by the chosen fuzzy rule set.

At the middle of the left window, there are two graphs for representing non-dominated rule sets in terms of "the total square error and the number of rules" and "the total square error and the user preference value". Red open plots represent non-dominated rule sets. Blue closed plot means the chosen rule set. The above inferred land price movement corresponds to the chosen rule set. Users can choose one of the non-dominated rule sets by clicking any plot in the graphs. The right CUI window shows the rules in the chosen rule set.

At the bottom of the left window, there are four satisfaction level functions. Users can change the shapes of these functions by moving each point B in Fig. 2.

The button "Evolve" is a trigger to start internal evaluations. In this paper, we specified the number of generations for internal evaluations as 100.

3.4 Some Results

When a user specified the satisfaction level functions for the maximum error and the number of rules as in Fig. 6, the user obtained a rule set with a small number of rules. The rule set seems to represent the original characteristics of the data. From Table 1, we can see that there are some general rules and specific rules in the fuzzy rule set. The value in parentheses in Table 1 represents the range of 0.5-level set of used membership function. For example, "1990 [1]" means the smallest partition in which the peak is 1990.

Table 1: Obtained rule set in Fig. 6.

Area	Year	Change %
Osaka	1990 [1]	56.1
Osaka	1992 [1]	-22.9
Tokyo	1988 [1]	68.6
Tokyo & Osaka	1993 [1]	-15.9
Osaka & Nagoya	1990 [2]	28.4
Tokyo & Osaka & Nagoya	1988 [4]	17.2
Tokyo & Osaka & Nagoya	-	4.3

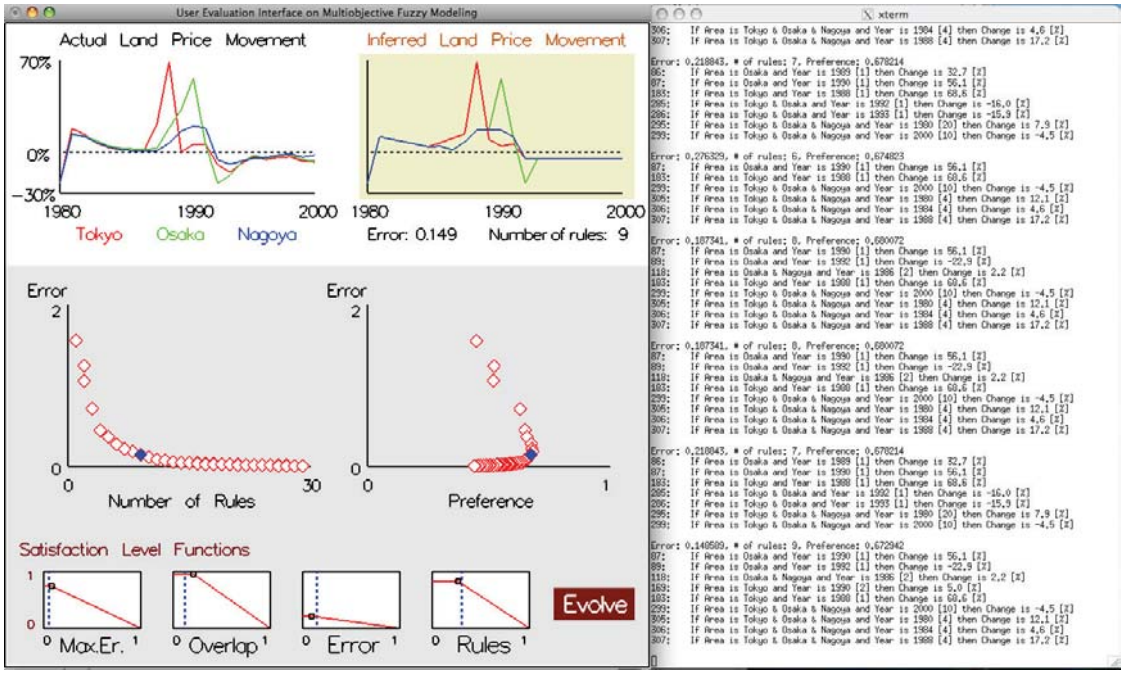


Figure 5: Prototype of user interface for interactive fuzzy modelling.

ISBN: 978-989-95079-6-8

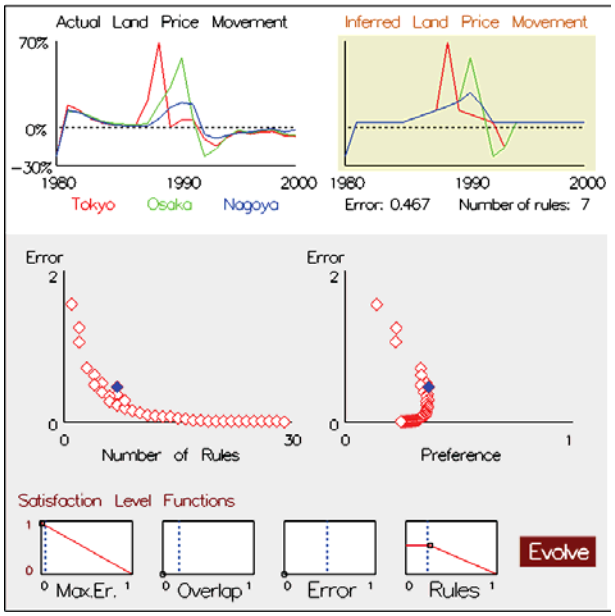


Figure 6: Example 1.

When a user gave a high priority to "Overlap among antecedent sets in S" as in Fig. 7, the user obtained a rule set with only one general rule and six specific rules (see Table 2).

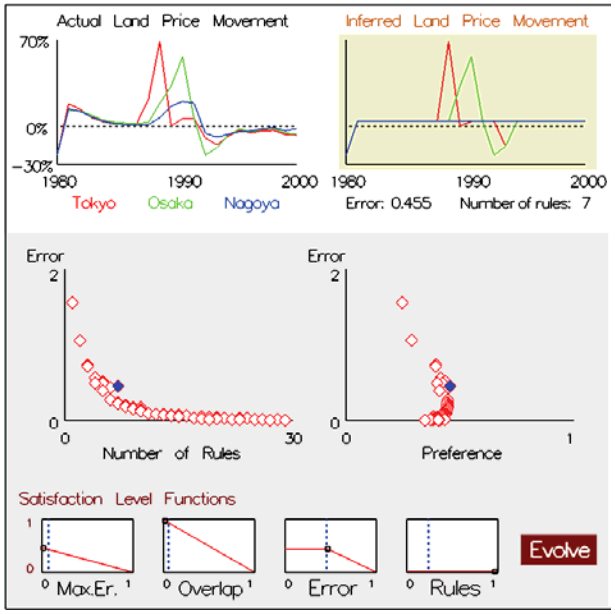


Figure 7: Example 2.

Table 2: Obtained rule set in Fig. 7.

Area	Year	Change %
Osaka	1989 [1]	32.7
Osaka	1990 [1]	56.1
Osaka	1992 [1]	-22.9
Tokyo	1988 [1]	68.6
Tokyo	1989 [1]	0.4
Tokyo & Osaka	1993 [1]	-15.9
Tokyo & Osaka & Nagoya	-	4.3

When a user gave high priorities to accuracy criteria (i.e., maximum error and total square error), a very accurate rule set was obtained as in Fig. 8.

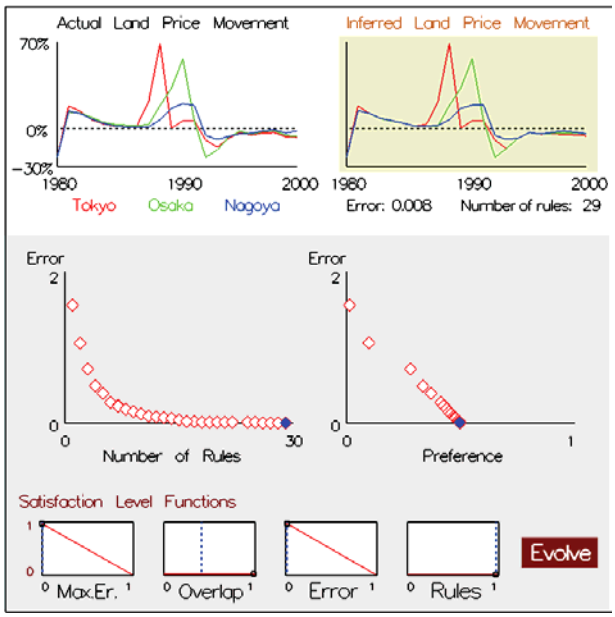


Figure 8: Example 3.

4 Conclusions

In this paper, we incorporated user's preference into multiobjective fuzzy modeling. We proposed a preference function composed of four satisfaction level functions. We utilized this preference function as an additional objective in an EMO algorithm. Through a case study, we demonstrated that a user can interactively specify satisfaction level functions during the evolution. We also showed that the user can obtain an accurate and interpretable fuzzy rule-based system based on his/her own preference.

In our case study, we intuitively selected four criteria to represent user's preference. Further studies are needed to choose appropriate criteria. We also have other interesting research issues to be discussed in future studies such as the visualization of multi-dimensional data and the minimization of human user's fatigue caused by the interaction with our system. The latter includes automated preference modeling.

Acknowledgment

This work partially supported by Grand-in-Aid for Young Scientists (B): KAKENHI (18700228).

References

[1] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka, Construction of fuzzy classification systems with rectangular fuzzy rules using genetic algorithms, *Fuzzy Sets and Systems* 65 (2/3) 237-253, 1994.

[2] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka, Selecting fuzzy if-then rules for classification problems using genetic algorithms, *IEEE Trans. on Fuzzy Systems* 3 (3) 260-270, 1995.

[3] H. Ishibuchi, T. Murata, and I. B. Turksen, Single-objective and two-objective genetic algorithms for selecting linguistic rules for pattern classification problems, *Fuzzy Sets and Systems* 89 (2) 135-150, 1997.

ISBN: 978-989-95079-6-8

[4] M. Setnes, R. Babuska, and B. Verbruggen, Rule-based modeling: Precision and transparency, *IEEE Trans. on Systems, Man, and Cybernetics - Part C* 28 (1) 165-169, 1998.

[5] J. Yen and L. Wang, G. W. Gillespie, Improving the interpretability of TSK fuzzy models by combining global learning and local learning, *IEEE Trans. on Fuzzy Systems* 6 (4) 530-537, 1998.

[6] D. Nauck and R. Kruse, Obtaining interpretable fuzzy classification rules from medical data, *Artificial Intelligence in Medicine* 16 (2) 149-169, 1999.

[7] Y. Jin, Fuzzy modeling of high-dimensional systems: Complexity reduction and interpretability improvement, *IEEE Trans. on Fuzzy Systems* 8 (2) 212-221, 2000.

[8] M. Setnes and H. Roubos, GA-based modeling and classification: Complexity and performance, *IEEE Trans. on Fuzzy Systems* 8 (5) 509-522, 2000.

[9] H. Ishibuchi, T. Nakashima, and T. Murata, Three-objective genetics-based machine learning for linguistic rule extraction, *Information Sciences* 136 (1-4) 109-133, 2001.

[10] L. Castillo, A. Gonzalez, and R. Perez, Including a simplicity criterion in the selection of the best rule in a genetic fuzzy learning algorithm, *Fuzzy Sets and Systems* 120 (2) 309-321, 2001.

[11] H. Roubos and M. Setnes, Compact and transparent fuzzy models and classifiers through iterative complexity reduction, *IEEE Trans. on Fuzzy Systems* 9 (4) 516-524, 2001.

[12] J. Abonyi, J. A. Roubos, and F. Szeifert, Data-driven generation of compact, accurate, and linguistically sound fuzzy classifiers based on a decision-tree initialization, *International Journal of Approximate Reasoning* 32 (1) 1-21, 2003.

[13] H. Ishibuchi and T. Yamamoto, Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining, *Fuzzy Sets and Systems* 141 (1) 59-88, 2004.

[14] J. Casillas, O. Cordon, 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. on Fuzzy Systems* 13 (1) 13-29, 2005.

[15] R. Alcala, J. Alcala-Fdez, J. Casillas, O. Cordon, and F. Herrera, Hybrid learning models to get the interpretability-accuracy trade-off in fuzzy modeling, *Soft Computing* 10 (9) 717-734, 2006.

[16] P. Pulkkinen and H. Koivisto, Identification of interpretable and accurate fuzzy classifiers and function estimators with hybrid methods, *Applied Soft Computing* 7 (2) 520-533, 2007.

[17] 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.

[18] M. Cococcioni, P. Ducange, B. Lazzerini, and F. Marcelloni, A Pareto-based multi-objective evolutionary approach to the identification of Mamdani fuzzy systems, *Soft Computing* 11 (11) 1013-1031, 2007.

[19] A. F. Gomez-Skarmeta, F. Jimenez, and G. Sanchez, Improving interpretability in approximative fuzzy models via multiobjective evolutionary algorithms, *International Journal of Intelligent Systems* 22 (9) 943-969, 2007.

[20] J. Gonzalez, I. Rojas, H. Pomares, L. J. Herrera, A. Guillen, J. M. Palomares, and 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* 44 (1) 32-44, 2007

[21] 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, 2009.

[22] L. Sanchez, J. Otero, and I. Couso, Obtaining linguistic fuzzy rule-based regression models from imprecise data with multiobjective genetic algorithms, *Soft Computing* 13 (5) 467-479, 2009.

[23] O. Cordon, F. Herrera, F. Hoffmann, and L. Magdalena, *Genetic Fuzzy Systems*, World Scientific, 2001.

[24] O. Cordon, F. Gomide, F. Herrera, 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.

[25] F. Herrera, Genetic fuzzy systems: Taxonomy, Current research trends and prospects, *Evolutionary Intelligence* 1, 27-46, 2008.

[26] K. Deb, *Multi-objective optimization using evolutionary algorithms*. John Wiley & Sons, Chichester, 2001.

[27] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Trans. on Evolutionary Computation* 6 (2) 182-197, 2002.

[28] C. A. C. Coello, D. A. van Veldhuizen, and B. B. Lamont, *Evolutionary algorithms for solving multi-objective problems*. Kluwer Academic Publishers, Boston, 2002.

[29] S. Guillaume, Designing fuzzy inference systems from data: An interpretability-oriented review, *IEEE Trans. on Fuzzy Systems* 9 (3) 426-443, 2001.

[30] R. Mikut, J. Jakel, and L. Groll, Interpretability issues in data-based learning of fuzzy systems, *Fuzzy Sets and Systems* 150 (2) 179-197, 2005.

[31] C. Mencar, G. Castellano, and A. M. Fanelli, On the role of interpretability in fuzzy data mining, *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems* 15 (5) 521-537, 2007.

[32] C. Mencar and A. M. Fanelli, Interpretability constraints for fuzzy information granulation, *Information Sciences* 178 (24) 4585-4618, 2008.

[33] S. M. Zhou and J. Q. Gan, Low-level interpretability and high-level interpretability: a unified view of data-driven interpretable fuzzy system modelling, *Fuzzy Sets and Systems* 159 (23) 3091-3131, 2008.

[34] J. M. Alonso, L. Magdalena, and S. 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.

[35] H. Ishibuchi and Y. Nojima, Discussions on interpretability of fuzzy systems using simple examples, *Proc. of 2009 IFSA World Congress and 2009 EUSFLAT conference* (to appear).

[36] Y. Nojima and H. Ishibuchi, Incorporation of user preference into multiobjective genetic fuzzy rule selection for pattern classification problems, *Proc. of 14th International Symposium on Artificial Life and Robotics 2009*, 186-189, 2009.

[37] Y. Nojima and H. Ishibuchi, Interactive genetic fuzzy rule selection through evolutionary multiobjective optimization with user preference, *Proc. of 2009 IEEE Symposium on Computational Intelligence in Multicriteria Decision-Making*, 141-148, 2009.

[38] H. Takagi, Interactive evolutionary computation: fusion of the capabilities of EC optimization an human evaluation, *Proceedings of the IEEE*, 81 (9) 1275-1296, 2001.

[39] S. B. Cho, Towards creative evolutionary systems with interactive genetic algorithm. *Applied Intelligence* 16: 129-138, 2002.

[40] A. M. Brintrup, J. Ramsden, H. Takagi and A. Tiwari, Ergonomic chair design by fusing qualitative and quantitative criteria using interactive genetic algorithms. *IEEE Trans. on Evolutionary Computation* 12 (3) 343-354, 2008.

[41] N. Kubota, Y. Nojima, F. Kojima, and T. Fukuda, Multiple fuzzy state-value functions for human evaluation through interactive trajectory planning of a partner robot, *Soft Computing* 10 (10), 891-901, 2006.

[42] H. Ichihashi, Iterative fuzzy modeling and a hierarchical network, *Proc. of 4th IFSA world congress*, 49-52, 1991.

[43] H. Ishibuchi, T. Yamamoto, and T. Nakashima, An approach to fuzzy default reasoning for function approximation, *Soft Computing* 10 (9), 850-864, 2006.

ISBN: 978-989-95079-6-8

Incremental possibilistic approach for online clustering and classification

Moamar Sayed Mouchaweh, Bernard Riera

Université de Reims Champagne-Ardenne, Centre de Recherche en STIC (URCA-CReSTIC)
Moulin de la Housse, BP 1039, 51687 Reims Cedex, France
Email: {moamar.sayed-mouchaweh,bernard.riera}@univ-reims.fr

Abstract— *In this paper, we propose to develop the supervised classification method Fuzzy Pattern Matching to be in addition a non supervised one. The goal is to monitor dynamic systems with a limited prior knowledge about their functioning. The detection of the occurrence of new states as well as the reinforcement of the estimation of their membership functions are performed online thanks to the combination of supervised and non supervised classification modes. No information in advance about the shape of classes or their number is required to achieve this detection and estimation reinforcement.*

Keywords— Classification, clustering, Sequential learning, Fuzzy Pattern Matching.

1 Introduction

Pattern Recognition (PR) [7] is the study of how machines can learn from experience to make sound decisions about the classes of patterns of interest. PR involves two stages: preprocessing and classification. The preprocessing includes feature extraction [1] and selection [12]. The classification stage is a mapping of a pattern from the feature space into the decision one. The latter is defined by a set of predefined classes. This mapping is done using a classifier. The latter is a method or algorithm which generates a class membership function in order to classify unlabelled incoming patterns into one of the predefined classes. Depending on the information available for classifier training, one can distinguish between supervised [10] and unsupervised [2] learning. In the first case, called also classification, there exists a set of patterns with their class assignment or label, called learning set. The goal of supervised learning is to learn a set of membership functions that allows the classification of new patterns into one of existing classes. The problem of unsupervised learning, also called clustering, arises if clusters' memberships of available patterns, and perhaps even the number of clusters, are unknown. In such cases, a classifier is learned based on similar properties of patterns. Hence, the clustering aims to partition a given set of patterns into clusters based on their similarity.

Semi supervised learning techniques use small or limited labelled patterns to estimate the classes' membership functions and the unlabelled ones to detect the occurrence of new classes and refine their membership functions estimation. Examples of semi supervised methods can be found in [4, 9, 11, 13] and the references therein. These methods are based either on the use of the Expectation Maximisation algorithm for maximum likelihood based parameters estimation [9], on the integration of an incremental algorithm for the update of classifiers' parameters [13], on the optimisation of an objective or learnable distance function [4] or on a classifier retraining to integrate new labelled points [11]. The popularity of these methods can be attributed to the fact that new information can be incorporated resulting better estimation of classes' membership functions and thus more prediction accuracies thanks to the unlabelled patterns. However, the representativeness of the labelled data is of crucial importance especially for small ratios of labelled to unlabelled patterns [9]. This is due to the fact that the clustering is guided by the labelled patterns.

One of the applications of PR methods is the monitoring of dynamic systems. System states, normal or faulty, are characterized by classes in the feature space. The performance of statistical PR methods depends on the prior knowledge, or learning set, about system behavior. The number of available learning patterns is often limited and small compared to the dimension of the feature space. Thus, it becomes hard to estimate the class membership function leading to a large variance in parameter estimates and thus higher classification error rates. Moreover, the behavior of a dynamic system can assume different operating states in the course of time. The learning set cannot contain patterns about all these states especially the faulty ones. Thus, the occurrence of these missing states must be anticipated online and integrated in the data base. In this paper, we propose a solution for these problems. This solution is based on the use of the supervised classification method Fuzzy Pattern Matching [3]. This method presents the advantage to process data with a low and constant classification time according to the size of data base. We propose to develop FPM as an unsupervised classification method. The goal is to combine the supervised and unsupervised learning strategies within a single algorithm leading to a semi supervised version of FPM.

The paper is organized as follows. Firstly, the functioning of FPM is illustrated. In the next section, the proposed solution to perform supervised and unsupervised classification using FPM is detailed. The performance of semi supervised FPM is illustrated and tested using some simulated examples.

ISBN: 978-989-95079-6-8

2 Supervised Fuzzy Pattern Matching

2.1 Learning phase

In the learning phase, the probability histograms are constructed for each class according to each attribute. The number of bins h for a histogram is experimentally determined. This number has an important influence on the performances of FPM [14]. The histogram upper and lower borders can be determined either as the maximal and minimal learning data coordinates or by experts. The height of each bin $b_k^j, k \in \{1,2,..,h\}$, according to each attribute j is the number of learning patterns $n_{ib_k^j}$ of the class C_i located in this bin. The probability distribution $\{ p_i^j(y_{b_k^j}), k \in \{1,2,..,h\}, j \in \{1,2,..,d\}\}$, of the class C_i, $i \in \{1,2,..,c\}$, according to the attribute j is obtained by dividing the height of each bin by the total number N_i of learning patterns belonging to the same class C_i. These probabilities are assigned to the bins' centres $y_{b_k^j}, k \in \{1,2,..,h\}$:

$$p_i^j(y_{b_k^j}) = \frac{n_{ib_k^j}}{N_i} \quad (1)$$

The Probability Density Function (PDF) is obtained by a linear linking between bins heights centres.

In order to take into account the uncertainty and the imprecision contained in the data, the probability distribution is converted into possibility one $\left\{\pi_i^j(y_{b_k^j}), k \in \{1,2,..,h\}, j \in \{1,2,..,d\}\right\}$. The conversion is performed using the transformation of Dubois and Prade [6] defined by:

$$\pi_i^j(y_{b_k^j}) = \sum_{f=1}^{h} \min(p_i^j(y_{b_k^j}), p_i^j(y_{b_f^j})) \quad (2)$$

A linear linking between bins heights centres converts the distribution of possibilities into density one Π_i^j. This operation is repeated for all the attributes of each class.

2.2 Classification Phase

The membership function for each class C_i according to each attribute j is considered to be numerically equivalent to the possibility distribution [15]. Thus, the classification of a new pattern x, whose values of the different attributes are x^1, .., x^j, .., x^d, is made in two steps :

Determination of the possibility membership value $\pi_i^j(x^j)$ of x^j to each class C_i according to the attribute j by a projection on the corresponding possibility density Π_i^j,

Merging all the possibility values $\pi_i^1(x^1)$, $\pi_i^2(x^2)$,.., $\pi_i^d(x^d)$ concerning the class C_i, into a single one by the aggregation operator "minimum" :

$$\pi_i(x) = \min(\pi_i^1(x^1), \pi_i^2(x^2),.., \pi_i^d(x^d)) \quad (3)$$

The result π_i of this fusion corresponds to the global possibility value that x belongs to the class C_i. Finally, x is assigned to the class for which it has the maximum membership value.

3 Semi Supervised Fuzzy Pattern Matching

The proposed semi supervised FPM has an agglomerative characteristic. Thus, it does not require any prior information about the number of classes. The classes' membership functions are constructed sequentially with the patterns' arrival. According to the ratio $r = \frac{L}{UL + L}$ of the number L of labelled points to the one UL of unlabelled points, the proposed method can be totally supervised, $r = 1$, or totally unsupervised, $r = 0$. Let $r_i = \frac{L_i}{UL_i + L_i}$ be the ratio of labelled points L_i belonging to the class C_i to the unlabelled ones UL_i which will be assigned to C_i. In the case that $0 \le r_i < 1$, the benefit of semi supervised FPM is to enhance the quality of class's membership estimation thanks to the incorporation of the unlabelled points in this class. This enhancement is performed online thanks to the use of an incremental approach as we can see later. While if $r_i = 0$, the benefit of semi supervised FPM is to detect this new class and to learn its membership function online. Thus, semi supervised FPM presents benefits in both classification and clustering.

In the case of $r_i = 0$, the first incoming unlabelled pattern is considered as the point prototype of a new class and its possibilistic membership function according to each attribute is computed as in supervised FPM based on this only pattern. The next unlabelled pattern is either classified in this created class, if it has a membership value according to this class, or considered as a point prototype of a new class. After the classification of each new pattern, the membership function of the corresponding class is updated online using an incremental algorithm. Due to the initialisation, created classes may need to be merged. This merging is performed using a similarity measure. The functioning of semi supervised FPM involves the following two steps.

3.1 Classes detection and local adaptation step

Let $x = (x^1, x^2,.., x^d) \in IR^d$ be a given pattern vector in a feature space constituted of d parameters or attributes. There is no learning set containing labelled patterns, nor a prior information about classes' probability density shape or their number. Each attribute is divided into equal intervals defining the bins of the histogram according to this attribute. This histogram is used to estimate the conditional probability density for the class that x is driven from. Let $X_{\min}^j$ and $X_{\max}^j$ be respectively the lower and upper borders of the histogram according to the attribute j. These borders can be determined by expert as the minimal and maximal values that an attribute can reach. Let h be the number of histogram's bins, then each bin according to the attribute j has the width :

ISBN: 978-989-95079-6-8

$$\Delta^j = \frac{X^j_{\max} - X^j_{\min}}{h},\ j \in \{1,2,..,d\} \quad (4)$$

Thus the limits of these bins are defined as follows :

$$b^j_1 = [X^j_{\min}, X^j_{\min} + \Delta^j],\ b^j_2 = [X^j_{\min} + \Delta^j, X^j_{\min} + 2\Delta^j],$$
$$.., b^j_h = [X^j_{\min} + (h-1)\Delta^j, X^j_{\max}],\ j \in \{1,2,..,d\} \quad (5)$$

The classes detection and local adaptation step involves two strategies : detection of new classes and local adaptation of their membership functions. The local adaptation strategy is based on an update of classes' possibility densities after the classification of each new pattern so that classifier can follow online gradual temporal, or local, changes of classes' membership functions. This online update requires a recursive representation of classes' possibility densities. However the incremental updating cannot detect abrupt changes as changes in the number of clusters. This abrupt change is followed up by the detection strategy which is based on the fact that each new rejected pattern by all the learned classes is considered as a point prototype of a new class. The detection strategy is a mechanism for adjusting the number of clusters online, which is incremented after the detection of each new cluster or class.

A) Detection of new classes strategy

The first rejected pattern x according to all the known c classes is considered as the point prototype of the first new class: $C_c \leftarrow x, c \leftarrow c+1$. The PDF is obtained as in supervised FPM. If x is located in the bin $b^j_k, k \in \{1,2,..,h\}$, then the probability histogram of C_c according to the attribute j is : $p^j_c = \{p^j_{c1} = 0, p^j_{c2} = 0, .., p^j_{ck} = 1, .., p^j_{ch} = 0\}$. The possibility histogram will then be computed using (2). Since there is just one pattern, the possibility histogram is equal to the probability one. The possibility density of the class C_c is obtained by a linear linking between the centre of the bin b^j_k , which has the height 1, and the ones of its left b^j_{k-1} and right b^j_{k+1} neighbours, which have both at present the height 0. Generally, if $C = \{C_1, C_2, .., C_c\}$ is the set of learned classes at present, x a new pattern which is rejected by all the learned classes. The detection strategy is defined as follows :

$$\pi_i(x) = 0, \forall i \in \{1,2,..,c\} \Rightarrow c \leftarrow c+1,$$
$$C_c = \{x\}, \pi_c = \{\pi^1_c, .., \pi^j_c, .., \pi^d_c\} \quad (6)$$

B) Local adaptation strategy

For a next pattern x', the membership value to each class C_i, $\forall i \in \{1,2,..,c\}$, will be obtained by a projection on its possibility density Π^j_i according to each attribute j and then merging the values according to all the attributes using the aggregation operator "minimum" as in supervised FPM. If the membership value $\pi_i(x')$ of x' to the class C_i is different of zero, then this pattern will be assigned to the class C_i and the possibility densities of this class according to each attribute will be incrementally updated. To establish an incremental update of possibility densities, let $p^j_i = \{p^j_{i1}, p^j_{i2}, .., p^j_{ik}, .., p^j_{ih}\}$ and $\pi^j_i = \{\pi^j_{i1}, \pi^j_{i2}, .., \pi^j_{ik}, .., \pi^j_{ih}\}$ define respectively the probability and possibility histograms of the class C_i according to the attribute j. Let $p'^j_i = \{p'^j_{i1}, p'^j_{i2}, .., p'^j_{ik}, .., p'^j_{ih}\}$ and $\pi'^j_i = \{\pi'^j_{i1}, \pi'^j_{i2}, .., \pi'^j_{ik}, .., \pi'^j_{ih}\}$ define respectively the updated probability and possibility histograms of the class C_i according to the attribute j after the assignment of x' to the class C_i. Let suppose for the simplicity that : $p^j_{ih} < p^j_{i(h-1)} < .. < p^j_{i1}$, then these new probabilities can be computed incrementally by [14] :

$$x'^j \in b^j_k, \forall k \in \{1,..,h\} \Rightarrow p'^j_{ik} = p^j_{ik} * \frac{N_i}{N_i+1} + \frac{1}{N_i+1},$$
$$p'^j_{iz} = p^j_{iz} * \frac{N_i}{N_i+1}, \forall z \in \{1,..,h\}, z \neq k \quad (7)$$

Then the new possibilities can be computed using Dubois and Prade transformation defined by (2). Thus, the local adaptation strategy is defined as follows :

$$\pi_i(x') = \max_{z \in \{1,..,c\}} (\pi_z(x')) \Rightarrow C_i \leftarrow \{C_i, x'\},$$
$$\pi'_i = \{\pi'^1_i, \pi'^2_i, .., \pi'^j_i, .., \pi'^d_i\} \quad (8)$$

3.2 Classes merging step

The occurrence order of incoming patterns influences the final constructed clusters. This entails the possibility to obtain several different partitions or number of clusters. Thus, several clusters can represent the same class. These clusters must be merged into one cluster to obtain one partition and one membership function. This fusion can be done either by expert or by a merging measure. The later measures the overlap or closeness between constructed clusters. There are different measures for merging clusters in the literature. Most of them are based on a similarity measure between clusters, which takes into account either the degree of overlapping of clusters or the distance between clusters' centres. The clusters overlapping degree is based on the number of ambiguous patterns, belonging to several clusters, and their membership values to these clusters. If the number of these ambiguous patterns is large enough and their membership values to several clusters are high then these clusters cannot be considered as heterogeneous anymore and must be merged. An interesting similarity criterion which takes into account at the same time the number of ambiguous patterns as well as their membership values is defined by (Frigui *el al.* 1996) :

$$\delta_{iz} = 1 - \frac{\sum_{x \in C_i \text{ or } x \in C_z} |\pi_i(x) - \pi_z(x)|}{\sum_{x \in C_i} \pi_i(x) + \sum_{x \in C_z} \pi_z(x)} \quad (9)$$

δ_{iz} is the fuzzy similarity measure between the classes C_i and C_z. More this measure is close to one, more the two clusters are overlapped. We adopt this measure for the merging step of semi supervised FPM.

ISBN: 978-989-95079-6-8

The merging criterion can be applied offline or online. In the first case, several iterations of clustering with a variable number of clusters start with a high over-specified number (an upper bound). Then, the clusters number is reduced gradually until an appropriate number is found. In each iteration, similar clusters are merged and this procedure is repeated until no more clusters can be merged and finished with the optimal number of clusters. In dynamic applications, the merging operation must be done online because the whole unlabelled patterns are not available *a priori*. Thus, the merging criterion can be tested either after the classification of each pattern or in each time window. The problem of the testing after each classification pattern is the calculation complexity which depends on the cardinality of clusters to be merged. However, the clusters are merged online when this measure reaches a predefined threshold. While in the other case, the merging can be delayed according to the size of the time window, but the calculation complexity is less than the one of the first case. We propose to update the fuzzy similarity measure within a time window. This update requires the calculation of the new membership values for all the patterns of all the classes inside which new patterns have been assigned. Indeed, if a new incoming pattern x is assigned to the cluster C_i, then the fuzzy similarity measure must be computed only between C_i and the other clusters C_z, $z \in \{1,..,c\}, z \neq i$.

When two clusters are merged, their membership functions must also be merged. We propose to merge the membership functions online using an incremental approach. Let $p_i^j = \{p_{i1}^j, p_{i2}^j,.., p_{ik}^j,.., p_{ih}^j\}$, $p_z^j = \{p_{z1}^j, p_{z2}^j,.., p_{zk}^j,.., p_{zh}^j\}$ be respectively the probability distributions of the two clusters C_i and C_z to be merged. Each bin probability is computed as (1) : $p_{ik}^j = \frac{n_{ib_k^j}}{N_i}$, $p_{zk}^j = \frac{n_{zb_k^j}}{N_z}, \forall k \in \{1,..,h\}$. The probability of the bin b_k^j after the merging of the patterns of the two clusters C_i and C_z is equal to :

$$p_{izk}^j = \frac{n_{ib_k^j} + n_{zb_k^j}}{N_i + N_z} \tag{10}$$

Where $n_{ib_k^j}$ and $n_{zb_k^j}$ are respectively the number of patterns of the classes C_i and C_z located in the bin b_k^j according to the attribute j. N_i and N_z are respectively the number of patterns of the classes C_i and C_z. We can rewrite (10) as follows:

$$p_{izk}^j = \frac{n_{ib_k^j}}{N_i} * \frac{N_i}{N_i + N_z} + \frac{n_{zb_k^j}}{N_z} * \frac{N_z}{N_i + N_z} = p_{ik}^j * \frac{N_i}{N_i + N_z} + p_{zk}^j * \frac{N_z}{N_i + N_z} \tag{11}$$

Using (11), the probability distribution of the class after merging is obtained incrementally according to each attribute. Based on (2), the corresponding possibility distributions defining the membership functions can be obtained.

4 Experimental Results

Figure 1 presents a simulated data base of two classes in a feature space of two attributes. The two classes are of different sizes, the first has 200 patterns and the second has 100 patterns. The distribution of each class has two independent normal variables with different standard deviations and means. Anyway, the two classes are not overlapped.

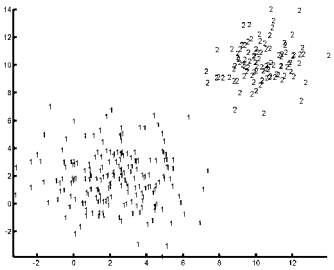


Figure 1: Example of two classes with different sizes, standard deviations and means.

We apply the proposed semi supervised FPM by selecting the patterns with a complete random order. The goal is to test the robustness of our algorithm against the initialisation problem. The experience is repeated several times with a different random pattern's occurrences at each time. We start the experience with a time window which is equal to the size of the data set, i.e., the merging step is applied after the reception of all the available patterns in the data base. Semi supervised FPM detects between 2 to 5 clusters according to the initialisation. Figure 2 shows the result of our algorithm in the case of the detection of 5 clusters.

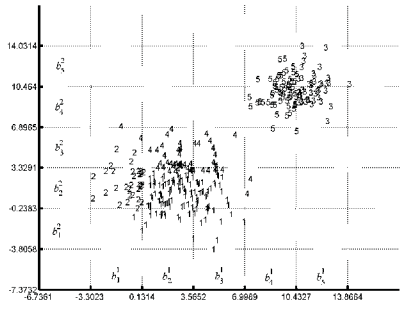


Figure 2: Results of the detection and local adaptation step of semi supervised FPM for the example of Figure 1.

The fuzzy similarity measures between the classes in this case are :

$\delta_{12} = 0.23$, $\delta_{13} = 0$, $\delta_{14} = 0.69$, $\delta_{15} = 0$, $\delta_{23} = 0$,
$\delta_{24} = 0.42$, $\delta_{34} = 0.04$, $\delta_{35} = 0.52$, $\delta_{45} = 0.04$

The merging threshold $\lambda_{\min} = 0.23$ is sufficient to merge the classes C_1, C_2 and C_4 into one cluster and C_3 with C_5 into one another cluster in order to obtain at the end the two necessary clusters. Based on the experimentation of several time windows' sizes, a time window of 30 patterns is sufficient to

ISBN: 978-989-95079-6-8

well estimate the correct number of clusters. Anyway, the width of the time window depends of the application: its dynamic, and the initialisation. If this width is too large, then the clusters merging will be delayed, while if it is too small the fuzzy similarity measure value may not be enough to validate the merging.

4.1 Classes of non convex shape

Supervised FPM works well if the classes are separated by at least one attribute. This is due to the fact that the classification of a pattern by FPM is based on a selection of one attribute. Another consequence of the selection of one attribute is that FPM does not respect the shape of classes if this shape is not convex, which is the case of the majority of real applications. Indeed, FPM provides always rectangular membership level curves for all the classes. Figure 3 presents a case for which the classes are not separated by at least one attribute. In addition, the classes C_1 and C_3 are of non convex shape. Figure 4 presents the results of the application of semi supervised FPM on this data, with a histogram containing 8 bins. Indeed, a higher number of histogram's bins is necessary when the classes are not separated by at least one attribute. A threshold equal to 0.14 is the required one for the merging measure to obtain the three classes.

Figure 5 shows the membership level curves obtained for this data set after the merging of clusters. Here the application of (11) leads to obtain one membership function for each class according to each attribute equivalent to the one resulting by the application of supervised FPM on each class after merging. We can see that these curves do not respect the shape non convex of classes C_1 and C_3. This, as we said before, is due to the classification decision based on the selection of one attribute. Inspired of the multi-prototypes approach [5] used in the literature to respect the shape of classes, we propose to merge the membership functions of the classes as follows :

$$\pi_{iz} = \min(1, \pi_i + \pi_z) \quad (12)$$

Where π_{iz} is the membership function after the fusion of classes C_i and C_z.

Using (12) means that each class is composed of several subclasses. Each subclass keeps its membership function. The application of (12) provides the membership level curves of Figure 6. We can see that these curves respect the classes shape. In addition, no supplementary computation is required to obtain this fusion. When the number of bins h increases, the number of subclasses increases also. This leads to obtain membership level curves which respect more precisely the classes' shape. However, increasing too much h entails the appearance of some membership peaks in the centre of classes.

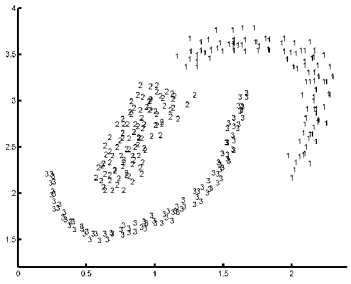


Figure 3: Classes which are not separated by at least one attribute and their shape is not convex (classes C_1 and C_3).

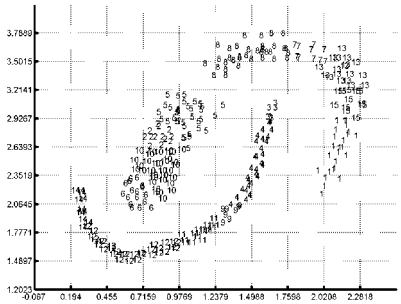


Figure 4: Clusters obtained by the application of semi supervised FPM on the data set of Figure 3.

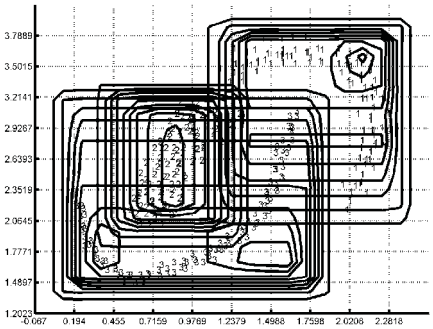


Figure 5: Membership level curves for the example of Figure 3. The membership functions of classes are obtained using (11).

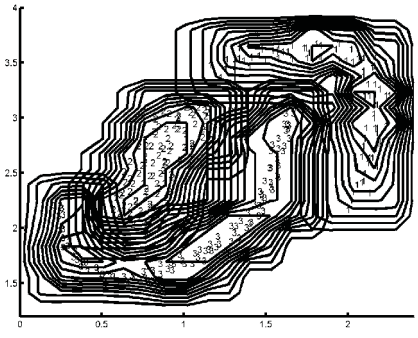


Figure 6: Membership level curves for the example of Figure 3 using (12).

ISBN: 978-989-95079-6-8

4.2 Overlapped classes

In the case of overlapped classes, the similarity measure becomes useless since the classes are not well separated. In addition, some points can be misclassified according to the ratio of labelled points to the total number of points in data sets. To allow testing the performance of semi supervised FPM in the case of overlapped classes, we take the following artificial 2-dimensioal dataset (Figure 7) available at http://www.stats.ox.ac.uk/pub/PRNN/. It is a normal mixtures data set. The training data consists of two classes with 125 patterns in each class. Each of the two classes has bimodal distribution. The testing set is an independent set of 1000 patterns drawn from the same distribution. The reported misclassification error is based on this testing set. The patterns were selected randomly which means that labelled patterns about some classes may not be presented at all. The experience is repeated 50 times to take into account the effect of initialisation. Table 1 presents the obtained results using semi supervised FPM. We can see that when *r* increases the misclassification error decreases thanks to the existence in advances of labelled patterns about some classes.

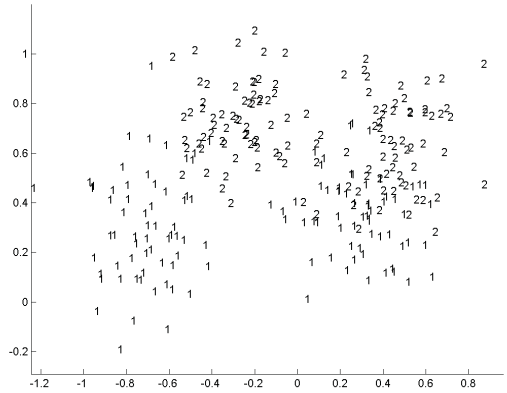


Figure 7: Normal mixtures data set.

Table 1: Misclassification Rate (*MR*) and its STandard Deviation (*STD*) in % according to different values of *r* for the normal mixture data set.

r %	0%	10%	20%	30%	40%
MR %	13.49	11.89	9.91	9.39	9.19
STD %	12.52	3.71	1.27	1.03	0.62
50%	60%	70%	80%	90%	100%
8.84	8.61	8.43	8.26	8.16	8.10
0.39	0.32	0.30	0.27	0.2	0

5 Conclusions

In this paper, the supervised classification method Fuzzy Pattern Matching (FPM) is developed to be also an unsupervised classification one. The goal is to obtain a semi supervised classification method adapted to dynamic systems for which a limited prior knowledge is available. Since the unsupervised learning technique is not based on a distance measure, the proposed algorithm will not favour the smaller sized clusters. In addition, it can start with no prior information. Finally, the membership functions can be adapted to elongated and non convex clusters.

We are developing FPM to be operant in the case of non stationary data. Indeed, in many practical situations, the environment changes. A learning data set used to construct the membership functions will be no more valid after a certain time. Thus, the classification method must be able to forget the information which is no more valid or representative of classes and adapt the membership functions based only on the recent and useful one.

Acknowledgements

This work is integrated in the regional project MOSYP (Performances Measurements and Optimization of Production Systems). The authors thank the region Champagne-Ardennes within this project (CPER ICOS).

References

[1] Benediktsson, J. A., J. R. Sveinsson, Feature extraction for multisource data classification with artificial neural networks, Int. journal of remote sensing 18(4) (1997) 727-740.

[2] Bezdek, J. C., N. R. Pal, Some new indexes of cluster validity, IEEE Tran. on Systems, Man and Cybernetics 28(3) (1998) 301-315.

[3] Cayrol, M., H. Farreny, H. Prade, Fuzzy Pattern Matching, Kybernetes 11, (1982) 103-116.

[4] Demiriz, K. P. B., M. J. Embrechts, Semi-supervised clustering using genetic algorithms, Artificial Neural Networks in Engineering, (1999) 809–814.

[5] Devillez, A., M. Sayed Mouchaweh, P. Billaudel, A process monitoring module based on fuzzy logic and pattern recognition, International Journal of Approximate Reasoning 37 (2004) 43-70.

[6] Dubois, D., H. Prade, On possibility/probability transformations, Fuzzy Logic (1993) 103-112.

[7] Duda, R. O., P. E. Hart, D. E. Stork, Pattern Classification second edition, Wiley, New York, (2001).

[8] Frigui, H., R. Krishnapuram, Robust algorithm for automatic extraction of an unknown number of clusters from noisy data, Pattern Recognition Letters 17 (1996) 1223-1232.

[9] Gabrys, B., L. Petrakieva, Combining labelled and unlabelled data in the design of pattern classification systems, International Journal of Approximate Reasoning 35(3), (2004) 251-273.

[10] Jain, A. K., P. Devijver, J. Kittler, Advances in statistical pattern recognition, Pattern Recognition Theory and Applications, Springer-Verlag, Berlin, Heidelberg, (1987) 1-19.

[11] Kate, R. J., R. J. Mooney, Semi-Supervised Learning for Semantic Parsing using Support Vector Machines, Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics, Rochester, NY, (2007) 81-84.

[12] Liu, H., L. Yu, Toward Integrating Feature Selection Algorithms for Classification and Clustering, IEEE Trans. on knowledge and data engineering 17 (4) (2005).

[13] Marsili-Libelli, S., Adaptive Fuzzy Monitoring and Fault Detection.Int. Journal of COMADEM 1(3) (1998) 31-37.

[14] Sayed Mouchaweh, M., A. Devillez, G. V. Lecolier, P. Billaudel, Incremental learning in real time using fuzzy pattern matching, Fuzzy Sets and Systems 132(1) (2002) 49-62.

[15] Zadeh, L. A., Fuzzy sets as a basis for a theory of possibility, Fuzzy sets and systems 1(3-2) (1978).

ISBN: 978-989-95079-6-8

Optimistic Fuzzy Weighted Average

Sylvie Galichet, Reda Boukezzoula[1]

1. Polytech'Savoie, LISTIC, Université de Savoie, Domaine Universitaire, BP 80439,
74944 Annecy le Vieux Cedex, France
Email: {sylvie.galichet, reda.boukezzoula}@univ-savoie.fr

Abstract *– The fuzzy weighted average (FWA) is used in many engineering problems where aggregation of fuzzy information is dealt with. In this framework, many algorithms have been proposed to compute efficiently the FWA according to Zadeh's extension principle. However, due to fuzzy interval calculus, the exact solution presents a characteristic that may be viewed as an important drawback. Indeed, replacing each individual by the fuzzy weighted average in the assessment of the population score leads to a result different from the original one. The presented work is an attempt to propose an optimistic counterpart for the FWA that eliminates the mentioned characteristic, possibly undesirable. Actually, the optimistic FWA is computed by using a modified division operator. When existing, the latter is the inverse operator of the fuzzy multiplication. Contrary to the conventional FWA, the optimistic one can be computed by a sequence of elementary arithmetic operations.*

Keywords *– Fuzzy arithmetic, Fuzzy weighted average, Gradual real numbers*

1 Introduction

The average is probably the easiest and the most widespread solution to aggregate information. The average is defined as the single value which all the individuals of a population should have so that their total is unchanged. When the sample contains several times the same individuals, it is possible to reformulate the average as a weighted average by introducing coefficients related to each individual. So a weighted average is an average in which a weight w_i is assigned to each quantity x_i to be averaged. The weights determine the relative importance of each quantity on the average. Considering N numbers x_i with associated weights w_i, the weighted average y is expressed as follows:

$$y = f(x_1, x_2, ..., x_N, w_1, w_2, ..., w_N) = \frac{x_1 w_1 + x_2 w_2 + ... + x_N w_N}{w_1 + w_2 + ... + w_N} \quad (1)$$

In an imprecise environment where information are poorly defined, it may be appropriate to represent scores and weighting coefficients by fuzzy numbers $\tilde{x}_i$ and $\tilde{w}_i$. In this case, the weighted average becomes a fuzzy number $\tilde{y}$ as well, i.e. the fuzzy weighted average (FWA).

If we follow the traditional fuzzy set theory, the fuzzy average $\tilde{y} = f(\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_N, \tilde{w}_1, \tilde{w}_2, ..., \tilde{w}_N)$ is obtained according to the extension principle that is:

$$\mu_{\tilde{y}}(y) = \sup_{\substack{x_1, .., x_N, w_1, .., w_N \ / \\ y = f(x_1, .., x_N, w_1, .., w_N)}} \min(\mu_{\tilde{x}_1}(x_1), .., \mu_{\tilde{x}_N}(x_N), \mu_{\tilde{w}_1}(w_1), .., \mu_{\tilde{w}_N}(w_N)) \quad (2)$$

Many authors proposed computational algorithms providing a discrete but exact solution of (2). All methods are based on the α-cut representation of fuzzy sets and interval analysis as initially suggested by Dong and Wong [5]. In interval computation, efficient algorithms exist for specific classes of functions, especially for the class of fractionally-linear functions [12]. The weighted average is an example of such functions and the problem of computing exactly the FWA is thus practically solvable. Liou and Wang [14] were the first to observe that since the x_i appear only in the numerator of (1), only the smallest values of the x_i are used to find the smallest value of (1), and only the largest values of the x_i are used to find the largest value of (1). This decomposition is used to reduce the algorithmic complexity in recent proposed approaches ([4], [8], [9], [12], [13], [15]).

Even if the above mentioned iterative procedures determine the exact solution for the FWA and so cope with multiple appearance of variables in the expression of the function f in (1), there is no known closed-form formula for computing $\tilde{y}$. Moreover, there is no available algorithm that could be directly implemented using elementary arithmetic operations between fuzzy operands. This aspect is studied in [16] where the authors are interested in designing a symbolic engine that could transform any given function into a sequence of elementary operations for which fuzzy interval computation would achieve exact result (without overestimation). Unfortunately, it was proven by Nguyen *et al.* [16] that operations with one or two fuzzy operands are not sufficient to describe generic functions on fuzzy sets.

In addition to the difficulty of obtaining the exact solution, the fuzzy weighted average has an important drawback directly related to fuzzy interval calculus. Indeed, replacing each individual by the fuzzy weighted average in the assessment of the population score leads to a result different from the original one. Actually, when the fuzzy average is viewed as expressing a requirement of maximal tolerance on a variable y which is itself the result of a computation involving quantities x_1, x_2, ..., x_N whose values are implicitly constrained by this calculation, it is desirable to solve the fuzzy

ISBN: 978-989-95079-6-8

equation:

$$B \otimes \tilde{y} = A \tag{3}$$

with $B = \tilde{w}_1 \oplus \tilde{w}_2 \; .. \oplus .. \; \tilde{w}_N$ (4)

and $A = (\tilde{w}_1 \otimes \tilde{x}_1) \oplus (\tilde{w}_2 \otimes \tilde{x}_2) \oplus \ldots \oplus (\tilde{w}_N \otimes \tilde{x}_N)$

for determining $\tilde{y}$. For example, such a procedure would be probably well suited to the case where the FWA is used for aggregating rule contributions in a Sugeno-like fuzzy system that would deal with fuzzy inputs.

In this context of «optimistic fuzzy interval calculus» [6], standard fuzzy operations can not be used directly for computing $\tilde{y}$. It means that a modified division denoted $\ominus_{\approx}$ must be searched for so that computing $\tilde{y}$ as $A \ominus_{\approx} B$ guarantees the satisfaction of the equality constraint (3). In the chosen context, both fuzzy quantities A and B are computed with the same weights $\tilde{w}_i$ according to (4) but the computation of $\tilde{y}$ does not handle explicitly weight interaction. The proposed approach is thus different from constrained fuzzy arithmetics as dealt with by Klir in [10] whose objective is an appropriate handling of interactive variables.

The motivation of this paper is twofold. Indeed, the presented work is an attempt:

- to propose an optimistic counterpart for the fuzzy weighted average,
- to compute the latter by a sequence of elementary operations.

This paper is organized as follows. Section 2 introduces fuzzy intervals with emphasis on their profile representation and their combination by arithmetic operators. In section 3, a new modified division operator is proposed for directly solving fuzzy equations. Then, section 4 is devoted to the implementation of the optimistic fuzzy weighted average with the presentation of two examples taken from the literature.

2 Fuzzy arithmetic operators

2.1 Fuzzy Intervals and Profiles

Let us consider a unimodal fuzzy interval A with kernel value K_A and support $S_A = [S_A^-, S_A^+]$ where S_A^- and S_A^+ are respectively the lower and upper bounds, of the interval S_A.

In order to specify the fuzzy interval shape, two additional functions are used to link the support with the kernel value. These functions, called left and right profiles, respectively denoted A^- and A^+, are defined from the membership function μ_A by:

$$\begin{aligned} A^-(\lambda) &= \text{Inf } \{x \mid \mu_A(x) \geq \lambda; x \geq S_A^-\} \\ A^+(\lambda) &= \text{Sup } \{x \mid \mu_A(x) \geq \lambda; x \leq S_A^+\} \end{aligned} \tag{5}$$

where $\lambda \in [0, 1]$. The profile A^- (resp. A^+) is an increasing (resp. decreasing) mapping that corresponds to the left (resp. right) part of the fuzzy interval A. It can be easily stated that:

$$K_A = A^-(1) = A^+(1), \tag{6}$$

$$S_{A=} [A^-(0), A^+(0)]. \tag{7}$$

Finally, the fuzzy interval A is univoquely defined by its left and right profiles. Thus, in the same way that the conventional interval S_A is denoted $[S_A^-, S_A^+]$, the fuzzy interval A will be denoted $[A^-, A^+]$. Equivalently, A can be viewed as the family of nested intervals $A(\lambda) = [A^-(\lambda), A^+(\lambda)]$, $\lambda \in [0, 1]$, when an explicit formulation with respect to λ is preferred.

For simplicity, the following additional notations are used in the remaining part of the paper. Given the interval $S = [S^-, S^+]$, its midpoint $M(S)$, its radius $R(S)$ and its relative extent $Rex(S)$ are respectively defined by:

$$M(S) = (S^+ + S^-) / 2, \tag{8}$$

$$R(S) = (S^+ - S^-) / 2, \tag{9}$$

$$Rex(S) = R(S) / M(S). \tag{10}$$

2.2 Arithmetic operators

Let $A = [A^-, A^+]$ and $B = [B^-, B^+]$ be two fuzzy intervals, the classical four arithmetic operations are expressed by:

$$A \oplus B = [A^- + B^-, A^+ + B^+] \tag{11}$$

$$A \ominus B = [A^- - B^+, A^+ - B^-] \tag{12}$$

$$A \otimes B = [\min Z, \max Z] \quad \text{where } Z = \{A^-B^-, A^-B^+, A^+B^-, A^+B^+\} \tag{13}$$

$$A \ominus B = [A^-, A^+] \otimes [1/B^+, 1/B^-] \quad \text{for } B \text{ such that } 0 \notin S_B \tag{14}$$

In order to cope with the twofold objective of solving equation (3) for determining the optimistic fuzzy average $\tilde{y}$ and of computing the fuzzy solution by a sequence of elementary operations, it is quite natural to search for a representation of $\tilde{y}$ in the form:

$$\tilde{y} = A \ominus B \tag{15}$$

where A and B are defined according to equation (4).

However, using the usual division and multiplication operators given by (14) and (13), it can be easily stated that substituting $\tilde{y}$ computed according to (15) into (3) gives a result more imprecise than the original A. Figure 1 illustrates this characteristic using two triangular fuzzy numbers A and B, with $K_A = 2.4$, $S_{A=} [0.8, 5.8]$ and $K_B = 1$, $S_{B=} [0.4, 1.9]$.

At best $A \subseteq B \otimes (A \ominus B)$ which means that the desired equality is generally not achieved. This problem is related to the lack of inverses in the calculus of fuzzy quantities.

Thus, a way around overestimation problems must be searched for outside standard arithmetic operations. One may think of using fuzzy arithmetic with requisite equality constraints as proposed in [10]. Klir's idea consists in doing fuzzy arithmetic with constraints dictated by the context of the problem. In practice, constraints are achieved by a-priori knowing that the α-cuts from two variables are the same. The approach is thus efficient for avoiding overestimation due to the occurrence of interactive variables. However, even when using constrained arithmetic, the calculus of fuzzy quantities is still pessimistic about the precision.

In the context of optimistic fuzzy calculus, we propose to use the modified division operator presented in next section.

ISBN: 978-989-95079-6-8

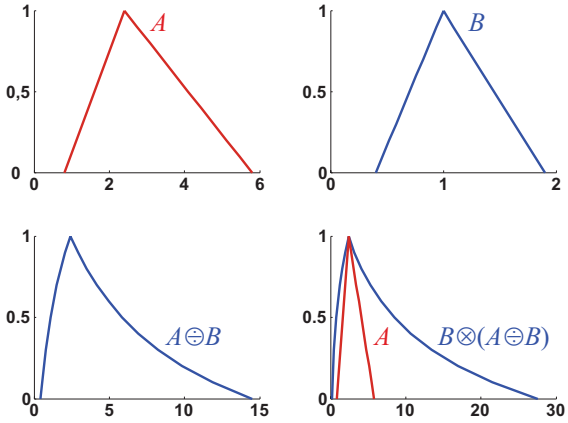


Figure 1: Fuzzy division and multiplication

3 The modified $\ominus_{\approx}$ Operator

3.1 Definition

Let $A = [A^-, A^+]$ and $B = [B^-, B^+]$ be two unimodal fuzzy intervals with $0 \notin S_B$. The result of the modified division $A \ominus_{\approx} B$ is defined from two gradual real numbers [7], denoted Φ^- and Φ^+, defined by their assignment functions from [0, 1] to the reals, as :

$$\Phi^-(\lambda) = \frac{Num^-(\lambda)}{Den^-(\lambda)} = \frac{M(A(\lambda)) - R(A(\lambda)) \cdot sign(M(S_B))}{M(B(\lambda)) - R(B(\lambda)) \cdot sign(Num^-(\lambda))} \quad (16)$$

and

$$\Phi^+(\lambda) = \frac{Num^+(\lambda)}{Den^+(\lambda)} = \frac{M(A(\lambda)) + R(A(\lambda)) \cdot sign(M(S_B))}{M(B(\lambda)) + R(B(\lambda)) \cdot sign(Num^+(\lambda))} \quad (17)$$

It is clear that, given A and B, both gradual numbers Φ^- and Φ^+ can always be computed. It is thus always possible to define $A \ominus_{\approx} B$ as the ordered pair (Φ^-, Φ^+) considering Φ^- as a left profile and Φ^+ as a right profile. Such entities are studied in [11] where it is proposed to add an extra feature, called "orientation", to pairs of profiles so as to build "ordered fuzzy numbers".

It is shown in [2] that the modified division $\ominus_{\approx}$ is the inverse operator of $\otimes$, that is $B \otimes (A \ominus_{\approx} B) = (A \ominus_{\approx} B) \otimes B = A$. Further details on the way equations (16) and (17) were established can be found in [3].

There is no guarantee that the ordered pair (Φ^-, Φ^+) is a gradual interval [7]. In other words, computations of Φ^- and Φ^+ according to (16) and (17) do not guarantee that $\forall \lambda \in [0, 1], \Phi^-(\lambda) \le \Phi^+(\lambda)$. It is however possible to determine a necessary and sufficient condition on operands A and B, under which $A \ominus_{\approx} B = (\Phi^-, \Phi^+)$ is the gradual interval $[\Phi^-, \Phi^+]$.

For unimodal operands, requiring that $\Phi^- \le \Phi^+$ at the kernel and support levels, that is :

$$\Phi^-(1) = \Phi^+(1) = K_\Phi$$
$$\Phi^-(0) \le K_\Phi \le \Phi^+(0)$$

leads to the following condition on A and B:

$$\frac{1 - |Rex(S_A)|}{1 - S \cdot |Rex(S_B)|} \le \frac{\delta K}{\delta M} \le \frac{1 + |Rex(S_A)|}{1 + |Rex(S_B)|} \quad (18)$$

where $\delta K = K_A / K_B$, $\delta M = M(S_A) / M(S_B)$ and $S = sign(1 - |Rex(S_A)|)$. Extending (18) to all λ-cuts leads to a necessary and sufficient condition under which $[\Phi^-, \Phi^+]$ is a gradual interval. Without any additional monotonicity assumption, the obtained interval is not always a fuzzy interval in the sense that λ-cuts are not necessarily nested.

Next subsection illustrates different behaviors of the pair (Φ^-, Φ^+).

3.2 Modified division of fuzzy triangular sets

Using the modified division operator for computing the division $A \ominus_{\approx} B$ where A and B are the triangular fuzzy subsets dealt with previously in figure 1, the fuzzy subset plotted in figure 2 is obtained. It can be stated that the latter is less imprecise than the standard division $A \ominus B$. As desired, the computed result is the exact solution of the fuzzy equation $B \otimes X = A$. It induces that X is implicitly less imprecise than A which is clearly verified in figure 2.

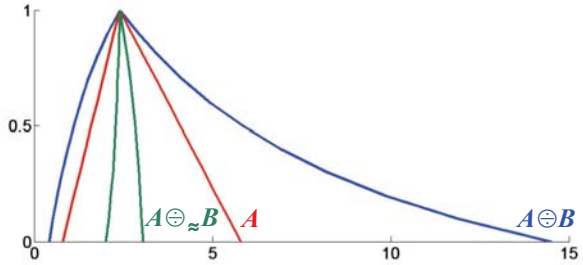


Figure 2: Modified fuzzy division

One may try to compute $1 \ominus_{\approx} A$ in order to determine the inverse of A according to $\ominus_{\approx}$ where 1 is the degenerated fuzzy subset associated with the crisp value 1. Considering the fuzzy subset A defined previously in figure 1, the result of the computation using equations (16) and (17) is given in figure 3. It appears clearly that (Φ^-, Φ^+) is not a gradual interval since the left and right profiles are exchanged, that is: $\Phi^+(\lambda) < \Phi^-(\lambda)$ for $\lambda \neq 1$.

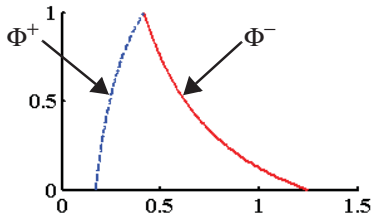


Figure 3: Computation of Φ^- and Φ^+ for $1 \ominus_{\approx} A$

It would have been possible to detect the «non-inversibility» of A prior to calculus according to the violation of condition (18).

The «non-inversibility» statement can be generalized to any

ISBN: 978-989-95079-6-8

non degenerated fuzzy subset. Indeed, let us suppose a fuzzy unimodal interval A, $0 \notin S_A$, such that $1 \ominus_{\approx} A$ exists. In this case, condition (18) has to be verified. As $K_1 = M(S_1) = 1$ and $R(S_1) = Rex(S_1) = 0$, condition (18) becomes:

$$\frac{1}{1-\left|Rex(S_A)\right|} \leq \frac{M(S_A)}{K_A} \leq \frac{1}{1+\left|Rex(S_A)\right|} \qquad (19)$$

As $0 \notin S_A$, $M(S_A)$ and K_A have the same sign. Let us first consider that $M(S_A)$ and K_A are both positive. Then, inequality (19) is rewritten as follows:

$$\frac{M(S_A)}{M(S_A)-R(S_A)} \leq \frac{M(S_A)}{K_A} \leq \frac{M(S_A)}{M(S_A)+R(S_A)} \qquad (20)$$

otherwise expressed as:

$$S_A^+ = M(S_A)+R(S_A) \leq K_A \leq M(S_A)-R(S_A) = S_A^- \qquad (21)$$

The single case where condition (21) holds, i.e. A is invertible, is for A being a crisp positive number. The case when $M(S_A)$ and K_A are both negative leads to a similar conclusion.

3.3 Modified division for computing FWA

In the context of computing weighted average, weights are usually supposed to be positive numbers. Extending to fuzzy weighted average, fuzzy positive weights $\tilde{w}_i$ are also assumed. Thus, computing the FWA $\tilde{y}$ as $A \ominus_{\approx} B$ with B being the sum of the weights, i.e. $B = \tilde{w}_1 \oplus \tilde{w}_2 \,..\oplus..\, \tilde{w}_N$, one can restrict the closed form of the modified division to the case of B positive. Equations (16) and (17) are then rewritten as $A \ominus_{\approx} B = [\Phi^-, \Phi^+]$, with

$$\Phi^-(\lambda) = \frac{A^-(\lambda)}{M(B(\lambda)) - R(B(\lambda)) \cdot sign(A^-(\lambda))} \qquad (22)$$

and

$$\Phi^+(\lambda) = \frac{A^+(\lambda)}{M(B(\lambda)) + R(B(\lambda)) \cdot sign(A^+(\lambda))}. \qquad (23)$$

4 Examples

4.1 Two-term average

To illustrate the proposed method for computing the fuzzy weighted average, a two-term example discussed in [9] is considered. Hence, we focus on the computation of:

$$\tilde{y} = (\tilde{x}_1 \otimes \tilde{w}_1 \oplus \tilde{x}_2 \otimes \tilde{w}_2) \ominus_{\approx} (\tilde{w}_1 \oplus \tilde{w}_2) \qquad (24)$$

where fuzzy triangular scores $\tilde{x}_1$, $\tilde{x}_2$ and fuzzy triangular weights $\tilde{w}_1$, $\tilde{w}_2$ are illustrated in figure 4. Using the profile representation, it follows:

$\tilde{x}_1 = [\lambda, 2-\lambda]$,
$\tilde{x}_2 = [\lambda+2, 4-\lambda]$,
$\tilde{w}_1 = [0.3\,\lambda, 0.9-0.6\lambda]$,
$\tilde{w}_2 = [0.3\lambda+0.4, 1-0.3\lambda]$.

The computation of $\tilde{y}$ according to (24) leads to the fuzzy result plotted in figure 5. As both fuzzy variables $\tilde{x}_1$ and $\tilde{x}_2$ are positive, according to (22) and (23) the developed forms of the left and right profiles of $\tilde{y}$ are given by :

$$\tilde{y}^- = \frac{\tilde{x}_1^- \cdot \tilde{w}_1^- + \tilde{x}_2^- \cdot \tilde{w}_2^-}{\tilde{w}_1^- + \tilde{w}_2^-} \qquad (25)$$

$$\tilde{y}^+ = \frac{\tilde{x}_1^+ \cdot \tilde{w}_1^+ + \tilde{x}_2^+ \cdot \tilde{w}_2^+}{\tilde{w}_1^+ + \tilde{w}_2^+} \qquad (26)$$

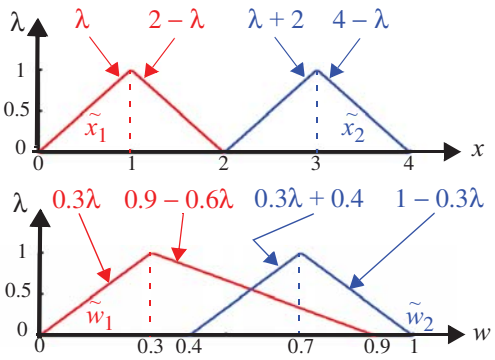


Figure 4: Fuzzy quantities and weights

Althrough the λ parametrization has been omitted in equations (25) and (26), all involved profiles are fonctions from the unit interval [0, 1] to the real line. It means that the addition, multiplication and division operations are function operations.

It can be observed that the computed optimistic FWA is a gradual interval. However, that one is not a fuzzy interval since its left profile is not an increasing function with respect to λ.

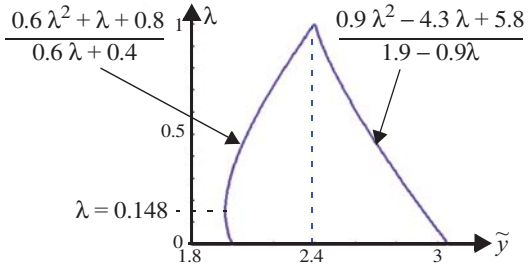


Figure 5: Optimistic fuzzy weighted average

For comparison purpose, figure 6 regroups the exact FWA according to Zadeh's extension principle with its lower and upper approximations, respectively obtained with the modified division operator and the conventional one. At the kernel level (λ = 1), all approaches give the same result, i.e. the weighted average for precise weights and numbers. At the other levels, the following inclusion is obtained:

$$\mathrm{FWA}_{\mathrm{optimistic}} \subseteq \mathrm{FWA}_{\mathrm{exact}} \subseteq \mathrm{FWA}_{\mathrm{arithmetic}}. \qquad (27)$$

In the very simple example under consideration, it can be observed that the $\tilde{x}_i$ variables are sorted. Then, according to [13], the analytical closed form of the exact FWA is expressed by:

$$\tilde{y}^- = \frac{\tilde{x}_1^- \cdot \tilde{w}_1^+ + \tilde{x}_2^- \cdot \tilde{w}_2^-}{\tilde{w}_1^+ + \tilde{w}_2^-} \qquad (28)$$

$$\tilde{y}^+ = \frac{\tilde{x}_1^+ \cdot \tilde{w}_1^- + \tilde{x}_2^+ \cdot \tilde{w}_2^+}{\tilde{w}_1^- + \tilde{w}_2^+} \qquad (29)$$

ISBN: 978-989-95079-6-8

However, in the general case it may not be possible to order the $\tilde{x}_i$ fuzzy variables and numerical methods have to be applied for discretized λ values. In this context, mathematical programming techniques may be preferred ([9], [8]). Methodologies based on the handling of gradual numbers are also sensitive to the ordering problem. It is thus suggested in [7] to divide the λ-range in several parts for which specific orders are defined.

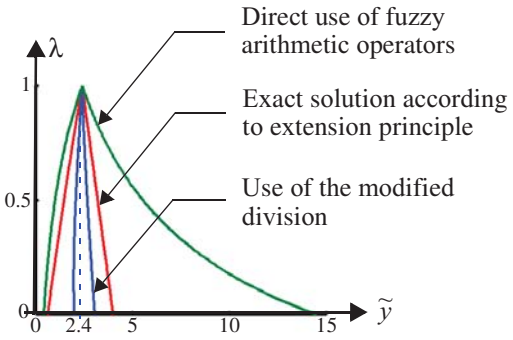


Figure 6: Fuzzy weighted average

4.2 Five-term example

This example taken from [13] consists in computing a five-term FWA, with the fuzzy triangular numbers $\tilde{x}_i$ and $\tilde{w}_i$, i = 1, ..., 5, given in table 1 and illustrated in figure 7.

Table 1: Fuzzy numbers and fuzzy weights

$[\, S_{\tilde{x}}^{-} \;\; K_{\tilde{x}} \;\; S_{\tilde{x}}^{+} \,]$	$[\, S_{\tilde{w}}^{-} \;\; K_{\tilde{w}} \;\; S_{\tilde{w}}^{+} \,]$
[1 2 3]	[1 2 5]
[2 5 7]	[2 2.5 3]
[6 8 9]	[4 7 9]
[7 9 10]	[3 4 7]
[10 11 12]	[2 3 4]

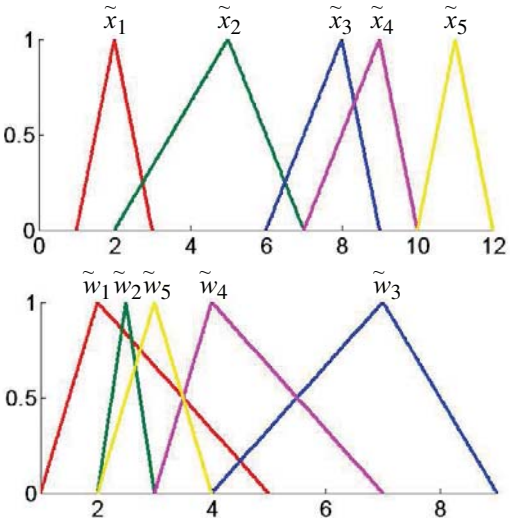


Figure 7: Fuzzy numbers and fuzzy weights

Using the modified division operator to compute the FWA, the result plotted in figure 8 is obtained. It can be observed that the optimistic FWA is here a fuzzy number, i.e. both left and right profiles are monotonic. As expected, the inclusion property (27) is valid.

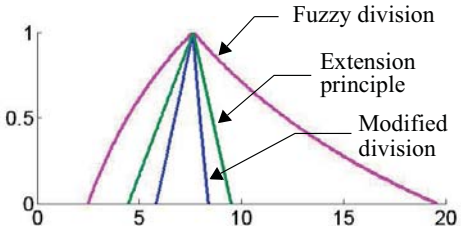


Figure 8: Five-term FWA

5 Conclusion

Based on the use of a modified fuzzy division operator, an optimistic counterpart of the usual fuzzy weighted average was proposed. The modified division is the inverse operation of the fuzzy multiplication. Consequently, the computed optimistic fuzzy average can replaced all individuals without modifying the weighted sum of the population. Moreover, the computation can be achieved by a sequence of elementary arithmetic operations. Unfortunately, the optimistic average may not be a fuzzy number but only an interval of gradual numbers. Actually, the modified division definition only guarantees that the result can be expressed in the form of an ordered pair of gradual numbers without further requirement on the profile monotonicity.

Two simple examples from the fuzzy literature have been used for illustration. More complicated and realistic cases must be further tested. For example, the proposed optimistic average may be used in the context of determining cluster centers for linguistic fuzzy C-means ([1]). Another possible use may be found in the aggregation of Sugeno-like rule consequents. More conceptual works have also to be developed in order to correctly position the optimistic FWA with respect to the exact one.

References

[1] S. Auephanwiriyakul, J. M. Keller, "Analysis and Efficient Implementation of a Linguistic Fuzzy C-Means", IEEE Trans. on Fuzzy Systems, Vol. 10, no. 5, pp. 563-582, 2002.

[2] R. Boukezzoula, S. Galichet, L. Foulloy, "Inverse arithmetic operators for fuzzy intervals", EUSFLAT'07, Ostrava, 2007.

[3] R. Boukezzoula, L. Foulloy, S. Galichet, "Inverse controller design for interval fuzzy systems, IEEE Trans. on fuzzy systems, Vol. 14, no. 1, pp. 111-124, 2006.

[4] P.T. Chang, K.C. Hung, K.P. Lin, C.H. Chang, "A comparison of discrete algorithms for fuzzy weighted average", IEEE Trans. on Fuzzy Systems, Vol. 14, no. 5, pp. 663-675, 2006.

[5] W.M. Dong, F.S. Wong, "Fuzzy weighted averages and implementation of the extension principle", Fuzzy Sets

ISBN: 978-989-95079-6-8

and Systems, no. 21, pp. 183-199, 1987.

[6] D. Dubois, E. Kerre, R. Mesiar, H. Prade, "Fuzzy interval analysis", in Fundamentals of Fuzzy Sets, The handbooks of fuzzy sets series, Eds. D. Dubois and H. Prade, 2000.

[7] J. Fortin, D. Dubois, H. Fargier, "Gradual numbers and their application to fuzzy interval analysis", IEEE Trans. on Fuzzy Systems, Vol. 16, no. 2, pp. 388-402, 2008.

[8] Y.Y. Guh, C.C. Hon, E.S. Lee, "Fuzzy weighted average: the linear programming approach via Charnes and Cooper's rule", Fuzzy Sets and Systems, no. 117, pp. 157-160, 2001.

[9] C. Kao, S.T. Liu, "Fractional programming approach to fuzzy weighted average", Fuzzy Sets and Systems, no. 120, pp. 435-444, 2001.

[10] G.J. Klir, "Fuzzy arithmetic with requisite constraints", Fuzzy Sets and Systems, no. 91, pp. 165-175, 1997.

[11] W. Kosinski, "On fuzzy number calculus", Int. J. Appl. Math. Comput. Sci., Vol. 16, no. 1, pp. 51-57, 2006.

[12] R. Lea, V. Kreinovich, R. Trejo, "Optimal interval enclosures for fractionally-linear functions, and their application to intelligent control", Reliable Computing, Vol. 2, no. 3, pp. 265-286, 1996.

[13] D.H. Lee, D. Park, "An efficient algorithm for fuzzy weighted average", Fuzzy Sets and Systems, no. 87, pp. 39-45, 1997.

[14] T.-S. Liou, M.J.J. Wang, "Fuzzy weighted average: an improved algorithm", Fuzzy Sets and Systems, no. 49, pp. 307–315, 1992.

[15] F. Liu, J. Mendel, "Aggregation Using the Fuzzy Weighted Average as Computed by the Karnik–Mendel Algorithms", IEEE Trans. on Fuzzy Systems, Vol. 16, no. 1, pp. 1-12, 2008.

[16] H.T. Nguyen, V. Kreinovich, V. Nesterov, M. Nakamura, "On hardware support for interval computations and soft computing: Theorems", IEEE Trans. on Fuzzy Systems, Vol. 5, no. 1, pp. 108-127, 1997.

ISBN: 978-989-95079-6-8

Evolving Inverse Fuzzy Models for Uncalibrated Visual Servoing in 3D Workspace

P. J. S. Gonçalves[1,2] J.M.C. Sousa[2] J.R. Caldas Pinto[2]

1.Instituto Politécnico de Castelo Branco, Escola Superior de Tecnologia
Av. Empresário, 6000-767 Castelo Branco, Portugal
2.IDMEC/IST,Technical University of Lisbon (TU Lisbon)
Av. Rovisco Pais, 1049-001 Lisboa, Portugal
Email: pgoncalves@est.ipcb.pt, {j.sousa,jcpinto}@dem.ist.utl.pt

Abstract— *In this paper, evolving inverse fuzzy models obtained online for uncalibrated visual servoing in 3D workspace, are developed and validated in a six degrees of freedom robotic manipulator. This approach will recursively update the inverse fuzzy model based only on measurements at a given time instant. The uncalibrated approach does not require calibrated kinematic and camera models, as needed in classical visual servoing to obtain the Jacobian. Experimental results obtained in a PUMA robot performing eye-to-hand visual servoing in 3D workspace are used to demonstrate the validity of the proposed approach, when compared to the previous developed off-line learning.*

Keywords— Visual Servoing, Fuzzy Control, Robotics, Evolving Fuzzy Systems.

1 Introduction

Industrial robotic manipulators use sensor-based control to perform tasks, either in structured or unstructured environments. Vision sensors provide a vast amount of information on the environment in which robots move. Thus, vision is essential for robots working in unstructured environments. This type of sensors definitely enlarges the potential applications of the actual robotic manipulators. Visual servoing can be defined as a method to control dynamic systems using the information provided by visual sensors. In this paper, the control of a robot manipulator end-effector using a single camera looking to the robot (eye-to-hand) is addressed [1].

Early approaches for visual servoing are based on a model already known, i.e. the robot-camera model, from which the relation between the features and the robot kinematics is analytically obtained [2]. Apart from the stated approaches where the robot-camera model is already known, it can also be estimated [3, 2]. This type of systems, *Uncalibrated Visual Servoing*, can deal with unknown robot kinematics and/or unknown camera parameters. By using this type of robot-camera models, the system becomes independent of robot type, camera type or even camera location.

In this paper, the robot-camera model estimation, from motions in the 3D workspace, by learning is addressed both online and off-line using fuzzy techniques to obtain a controller capable of controlling the system. An inverse fuzzy model is used to derive the inverse robot-camera model, i.e. the Jacobian, in order to compute the joints and end-effector velocities in a straightforward manner. The inverse fuzzy model cab be applied directly as a controller, which is a simple way to implement a controller in real-time. Note that this feature is very important in robotic systems.

From the modeling techniques based on soft computing, fuzzy modeling is one of the most appealing. As the robotic manipulator together with the visual system is a nonlinear system, which is only partly known, it is advantageous to use fuzzy modeling as a way to derive a model (or an inverse model as in this case) based only on measurements. In this paper will be applied and compared two approaches, off-line [2], and on-line [4, 5]. The first approach have already proven to perform well in visual servoing systems. It's main drawback is that the learning must be performed off-line. Since the environment in which the robot operates can vary importantly, new models should be derived very often, which take time and also the robot must stop the task that is performing. That is why an on-line modeling of the visual servoing system is preferable. This approach was not yet applied to control robotic manipulators using vision and is the objective of the present paper.

The paper is organized as follows. Section 2 describes briefly the concept of visual servoing and presents the uncalibrated visual servoing approach. Section 3 presents very briefly on-line and off-line fuzzy modeling, and discusses the problem of identifying inverse fuzzy models directly from data measurements. Section 4 describes the experimental setup and presents the obtained results, where the identified inverse fuzzy models are discussed. Finally, Section 5 presents the conclusions and the future work.

2 Visual Servoing

Machine Vision and Robotics can be used together to control a robotic manipulator. This type of control, which is defined as Visual Servoing, uses visual information from the work environment to control a robotic manipulator performing a given task. In the following sub-sections classical and uncalibrated visual servoing are presented, along with their main goals.

2.1 Classical Visual Servoing

Within the visual servoing framework, there are three main categories related with the information obtained from the image features:

- *image-based visual servoing* [1, 6], which uses direct information from the object in the image, i.e. image features,

ISBN: 978-989-95079-6-8

- *position-based visual servoing* [1], which uses 3D information of the object from the image(s), e.g. the 3D coordinates of the object or the rotation and translation between the camera and object frames, obtained using a CAD model of the object to perform an on-line pose estimation,
- *hybrid visual servoing* [7], which combines the first two approaches and is a possible solution to some drawbacks of image-based and position-based methods [8].

In the following, the classical approaches to modeling and control of position-based visual servoing are presented.

2.1.1 Modeling the Position-Based Visual Servoing System

In this paper position-based visual servoing with 3D features [9] is used in an eye-to-hand system [1], where the camera is fixed and looking the robot and object. The 3D image features, s are 3D points of the object in the camera frame, p. The kinematic modeling of the transformation between the image features velocities, $\dot{s}$, and the joints velocities $\dot{q}$ is defined as follows [9]:

$$\dot{s} = \begin{bmatrix} -I_3 & S(p) \end{bmatrix} \cdot {}^{c}W_e \cdot {}^{e}J_R \cdot \dot{q} = J \cdot \dot{q}\,, \qquad (1)$$

where I_3 is the 3×3 identity matrix, $S(p)$ is the skew-symmetric matrix of the 3D point p, ${}^{c}W_e$ is defined as the transformation between the camera and end-effector frames velocities, and ${}^{e}J_R$ is the robot Jacobian. The 3D point is obtained from the captured image using a pose estimation algorithm [9].

2.2 Uncalibrated Visual Servoing

To derive an accurate Jacobian, J, a perfect modeling of the camera, the chosen image features, the position of the camera related to the world, and the depth of the target related to the camera frame must be accurately determined. Even when a perfect model of the Jacobian is available, it can contain singularities, which hampers the application of a control law. Remind that the Jacobian must be inverted to send the camera velocity to the robot inner control loop. When the Jacobian is singular, the control cannot be correctly performed.

There are visual servoing systems that obviate the calibration step and estimate the robot-camera model either online or offline. The robot-camera model may be estimated:

- analytically, using nonlinear least square optimization [10], and
- by learning or training, using fuzzy membership functions and neural networks [11, 2].

In addition, the control system may estimate an image Jacobian and use the known robot model, or a coupled robot-camera Jacobian may be estimated.

To overcome the difficulties regarding the Jacobian, a new type of differential relationship between the features and camera velocities was proposed in [11]. This approach estimates the variation of the image features, when an increment in the camera position is given, by using a relation G. This relation is divided into G_1 which relates the position of the camera and the image features, and F which relates their respective variation:

$$s + \delta s = G(q + \delta\, q) = G_1(q) + F(q, \delta\, q). \qquad (2)$$

Considering only the variations in (2):

$$\delta s = F(q, \delta\, q), \qquad (3)$$

the inverse function F^{-1} is given by:

$$\delta q = F^{-1}(\delta s, q)\,, \qquad (4)$$

and it states that the joint variation depends on the image features variation and the previous position of the robot manipulator. Equation (4) can be discretized as

$$\delta q(k) = F_k^{-1}(\delta s(k+1), q(k)). \qquad (5)$$

In image-based visual servoing, the goal is to obtain a joint velocity, $\delta q(k)$, capable of driving the robot according to a desired image feature position, $s(k+1)$, with an also desired image feature error, $\delta s(k+1)$, from any position in the joint spaces. This goal can be accomplished by modeling the inverse function F_k^{-1}, using inverse fuzzy modeling as proposed in this paper and presented in Section 3. This new approach to image-based visual servoing allows to overcome the problems stated previously regarding the Jacobian and the calibration of the robot-camera model.

3 Inverse Fuzzy Modeling

3.1 Off-Line Fuzzy Modeling

Fuzzy modeling often follows the approach of encoding expert knowledge expressed in a verbal form in a collection of if–then rules. Parameters in this structure can be adapted using input-output data. When no prior knowledge about the system is available, a fuzzy model can be constructed entirely on the basis of system measurements. In the following, we consider data-driven modeling based on fuzzy clustering [12, 13].

We consider rule-based models of the Takagi-Sugeno (TS) type. TS models consist of fuzzy rules describing a local input-output relation, typically in an affine form:

$$\begin{aligned} R_i : \textbf{If } & x_1 \text{ is } A_{i1} \textbf{ and } \ldots \textbf{and } x_n \text{ is } A_{in} \\ \textbf{then } & y_i = \mathbf{a}_i \mathbf{x} + b_i\,, \quad i = 1, 2, \ldots, K. \end{aligned} \qquad (6)$$

Here R_i is the ith rule, $\mathbf{x} = [x_1, \ldots, x_n]^T$ are the antecedent variables, $A_{i1}, \ldots, A_{in}$ are fuzzy sets defined in the antecedent space, and y_i is the rule output variable. K denotes the number of rules in the rule base, and the aggregated output of the model, $\hat{y}$, is calculated by taking the weighted average of the rule consequents:

$$\hat{y} = \frac{\sum_{i=1}^{K} \beta_i y_i}{\sum_{i=1}^{K} \beta_i}\,, \qquad (7)$$

where β_i is the degree of activation of the ith rule: $\beta_i = \Pi_{j=1}^{n} \mu_{A_{ij}}(x_j)$, $i = 1, \ldots, K$, and $A_{ij}(x_j) : \mathbb{R} \rightarrow [0, 1]$ is the membership function of the fuzzy set A_{ij} in the antecedent of R_i.

To identify the model in (6), the regression matrix X and an output vector $\mathbf{y}$ are constructed from the available data: $\mathrm{X}^T = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$, $\mathbf{y}^T = [y_1, \ldots, y_N]$, where $N \gg n$ is the number of samples used for identification. The number of rules, K, the antecedent fuzzy sets, A_{ij}, and the consequent parameters, $\mathbf{a}_i, b_i$ are determined by means of fuzzy clustering

ISBN: 978-989-95079-6-8

in the product space of the inputs and the outputs [13]. Hence, the data set Z to be clustered is composed from X and $\mathbf{y}$: $\mathrm{Z}^T = [\mathrm{X}, \mathbf{y}]$. Given Z and an estimated number of clusters K, the Gustafson-Kessel fuzzy clustering algorithm [14] is applied to compute the fuzzy partition matrix U.

The fuzzy sets in the antecedent of the rules are obtained from the partition matrix U, whose ikth element $\mu_{ik} \in [0, 1]$ is the membership degree of the data object $\mathbf{z}_k$ in cluster i. One-dimensional fuzzy sets A_{ij} are obtained from the multi-dimensional fuzzy sets defined point-wise in the ith row of the partition matrix by projections onto the space of the input variables x_j. The point-wise defined fuzzy sets A_{ij} are approximated by suitable parametric functions in order to compute $\mu_{A_{ij}}(x_j)$ for any value of x_j.

The consequent parameters for each rule are obtained as a weighted ordinary least-square estimate. Let $\theta_i^T = [\mathbf{a}_i^T; b_i]$, let X_e denote the matrix $[\mathrm{X}; \mathbf{1}]$ and let W_i denote a diagonal matrix in $\mathbb{R}^{N \times N}$ having the degree of activation, $\beta_i(\mathbf{x}_k)$, as its kth diagonal element. Assuming that the columns of X_e are linearly independent and $\beta_i(\mathbf{x}_k) > 0$ for $1 \leq k \leq N$, the weighted least-squares solution of $\mathbf{y} = \mathrm{X}_e\theta + \epsilon$ becomes

$$\theta_i = \left[\mathrm{X}_e^T \mathrm{W}_i \mathrm{X}_e\right]^{-1} \mathrm{X}_e^T \mathrm{W}_i \mathbf{y}\,. \tag{8}$$

3.2 On-Line Fuzzy Modeling

The model obtained from the techniques presented in the previous section is assumed to be fixed, since it is learned in off-line mode. Recently attention is focused in on-line learning [4], where in a first phase input-output data is partitioned using unsupervised clustering methods and in a second phase, parameter identification is performed using a supervised learning method.

In On-Line Fuzzy Modeling and according to [4], also rule-based models of the TS type, are considered. Typically in the affine form described in (6), where the input-output data is acquired continuously. The new data, arriving at some time instant, can bring new information from the system, which could indicate a change in its dynamics. This information may change an existing rule, by changing the spread of the membership functions, or even introduce a new one. To achieve this, the algorithm must be able to judge the informative potential and the importance of the new data.

In the following are briefly presented the several steps of the algorithm used for on-line fuzzy modeling, proposed in [4], evolving fuzzy systems. The first step is based on the subtractive clustering algorithm [15], where the input-output data is partitioned. The procedure used must be initialized, i.e. the focal point of the first cluster is equal to the first data point and its potential is equal to one. Starting from the first data point, the potential of the next data point is calculated recursively using a Cauchy type function of first order:

$$P_k(z_k) = \frac{1}{1 + \frac{1}{k-1}\sum_{i=1}^{k-1}\sum_{j=1}^{n+1}(d_{ik}^j)^2}, \quad k = 2, 3, \ldots \tag{9}$$

where $P_k(z_k)$ denotes the potential of the data point z_k calculated at time k; $d_{ik}^j = z_i^j - z_k^j$, denotes projection of the distance between two data points (z_i^j and z_k^j) on the axis z^j.

When a new data point arrives it also influences the potential of the already defined center of the clusters ($z_i^*, i = 1, 2, ..., K$). A recursive formula for the update of the cluster centers potential is defined in [4]:

$$P_k(z_i^*) = \frac{(k-1)P_{(k-1)}(z_i^*)}{k - 2 + P_{(k-1)}(z_i^*) + P_{(k-1)}(z_i^*) + \sum_{j=1}^{n+1}(d_{ik}^j)^2},$$

where $P_k(z_i^*)$ is the potential at time k of the cluster center, related to the rule i.

The next step of the algorithm is to compare the potential of the actual data point to the potentials of the existing cluster centers. According to the two approaches tested in this paper [4, 5] a crucial part of the evolving fuzzy systems is the rule creation and modification. The first approach is called Evolving Takagi-Sugeno (eTS) and the second eXtended Takagi-Sugeno (xTS).

In eTS, if the potential of a new data point is higher than the potential of the existing cluster centers, then the new data point is accepted as a new cluster center and a new rule is formed. If in addition to the previous condition the new data point is close to an old cluster center, the old cluster center is replaced. This last condition is defined by:

$$\frac{P_k(Z_k)}{\underset{i=1}{\overset{K}{max}} P_k(z_i^*)} - \frac{\delta_{min}}{r} \geq 1, \tag{10}$$

where $r \in [0.3; 0.5]$ is the spread of the antecedent [4], and δ_{min} is the shortest distance between the new candidate Z_k and all the existing cluster centers z_i^*.

In xTS, if the potential of a new data point is higher than the maximum potential of all the existing clusters centers or is lower than the minimum potential of all the existing clusters centers, then the new data point is accepted as a new cluster center and a new rule is formed. In addition to the previous condition, if the next condition is true then the old cluster is replaced because the candidate is close to an old cluster center:

$$\underset{i=1}{\overset{K}{min}} \|z_k - z_i^*\|_j \geq \frac{r_j^i}{2}, \tag{11}$$

where r_j^i is the spread that is not fixed, like in the previous approach. In this approach the spread is adaptive and can vary in each element j of every cluster (rule) i, making possible to obtain hyper-ellipsoidal clusters. The recursive formula to obtain the adaptive spread is as follows:

$$r_{jk}^p = \rho \cdot r_{j(k-1)}^p + 0.5\sqrt{D_{jk}^p}; p = \underset{i=1}{\overset{K}{argmin}} \|z_k - z_i^*\|, \tag{12}$$

where D_{jk}^p is the local scatter and $\rho = 0.75$, both defined in [5].

For both approaches the fuzzy sets in the antecedent of the rules are gaussian, with the form:

$$\mu_{ij} = e^{-r\|x_j - x_{ij}^*\|^2}, \tag{13}$$

The consequents of the fuzzy rules are obtained using the global parameter estimation procedure based on the weighted recursive least squares, presented in [4].

3.3 Inverse modeling

For the robotic application in this paper, the inverse model is identified using input-output data from the inputs $\dot{q}(k)$ and

ISBN: 978-989-95079-6-8



Figure 1: The eye-to-hand experimental setup.

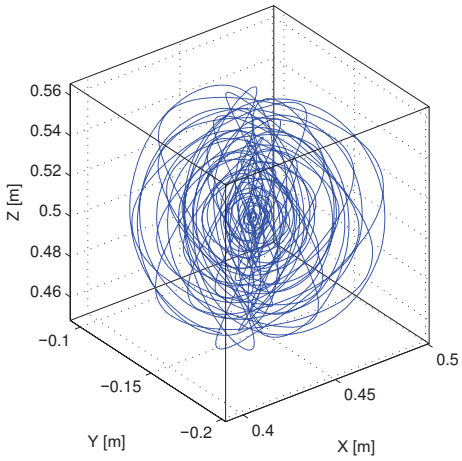


Figure 2: Robot end-effector 3D spiral trajectory.

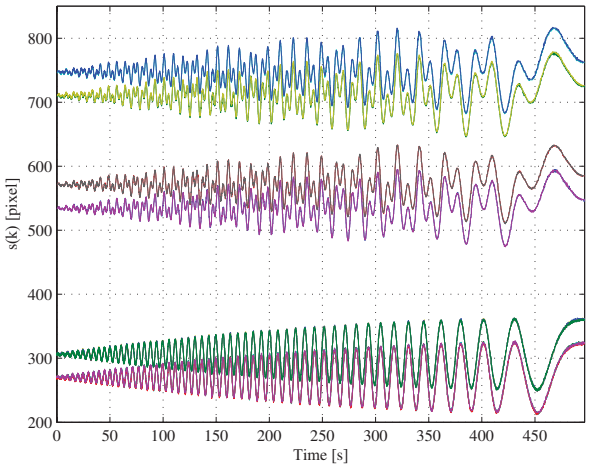


Figure 3: Input data for inverse model identification, image features $s(k)$.

the outputs $\delta s(k+1)$, following the procedure described in [2]. In this paper, the approach presented in [3] to obtain the 3D training set, was used. Note that we are interested in the identification of an inverse model as in (5).

4 Results

4.1 Experimental Setup

A Puma 560 Robotic Manipulator and a Vector CCi4 camera, in eye-to-hand configuration, were used to demonstrate the validity of the approach proposed in this paper. The experimental setup is presented in Fig. 1. The visual control algorithms were implemented in real-time using MatLab Simulink, and the xPC Target toolbox. The robot inner loop velocity control, performs at 1 KHz and the visual loop control at 12.5 Hz. A planar object is rigidly attached to the robot end-effector and is described with eight points, which centroids were used as the sixteen image features. In this paper, the robot moves in its 3D workspace (moving joints 1, 2 and 3). To maintain the image features in the camera field of view during the servoing, the planar target is set to be perpendicular to the camera optical axis by moving joints 4 and 5, as shown in Fig. 1.

4.2 Inverse Modeling Results

Following the work in [3], to obtain the identification data, the robot swept the 3D workspace in the camera field of view in a 3D spiral path, starting in the spiral center, Fig. 2, which allows to obtain the model for the end-effector position.

The variables needed for identification, $\delta q(k)$ and $\delta s(k+1)$, are obtained from the spiral when setting the desired position to the spirals center, by:

$$\delta q(k) = \frac{q^* - q(k)}{\Delta t} \tag{14}$$

$$\delta s(k+1) = s^* - s(k+1) \tag{15}$$

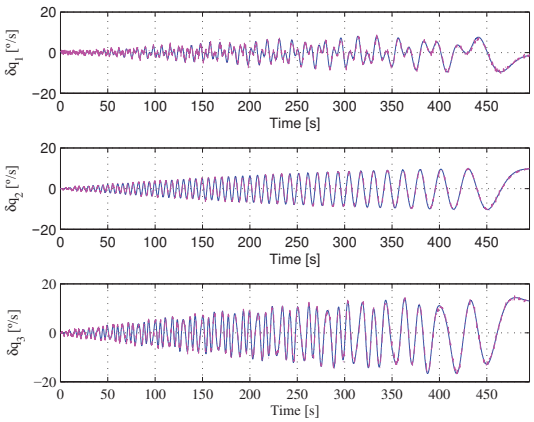
This allows to cover a wide range of values for $\delta q(k)$ and $\delta s(k+1)$. The 3D spiral also allows to control the precision of the model, by increasing or decreasing its parameters. Moving the first, the second and third joints of the PUMA Robotic Manipulator, the end-effector position follows the desired 3D spiral. The end-effector orientation is set to maintain the target parallel to the image plane.

The inverse fuzzy model was identified from the 3D spiral trajectory data, i.e. the joint velocities (see Fig. 4) and the sixteen image features obtained (see Fig. 3) using (14).

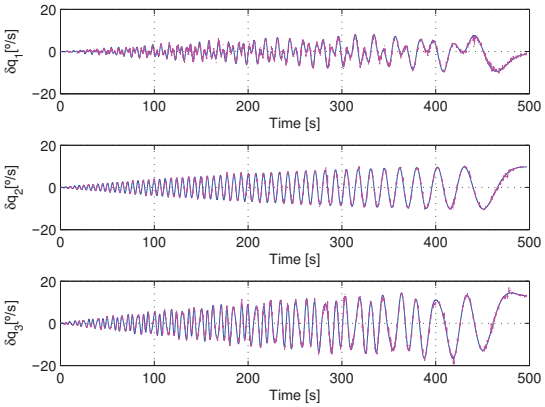
Note that to identify the inverse model, one cannot simply feed the inputs as outputs and the outputs as inputs. Since the inverse model (5) is a non-causal model, the output of the inverse model must be shifted one step, see [16].

For the identification of joints 1, 2 and 3, the obtained joint velocities $\dot{q}(k)$, are shown in Fig. 4, solid lines. Note that three fuzzy models are identified, one for each joint velocity. In the plots it is hard to see the difference between the real output data and the output of the inverse fuzzy model, because they are very similar, output of the models in dash-dotted lines. In order to measure modeling accuracy, this paper uses the Variance Accounted For (VAF) [2].

ISBN: 978-989-95079-6-8



(a) Comparison of the output data of the off-line inverse fuzzy model, $\dot{q}(k)$, joints 1 to 3. Solid – real output data, and dash-dotted – output of the off-line inverse fuzzy model.



(b) Comparison of the output data of the on-line inverse fuzzy model (xTs), $\dot{q}(k)$, joints 1 to 3. Solid – real output data, and dash-dotted – output of the on-line inverse fuzzy model.

Figure 4: Output data for inverse model identification, joints 1 to 3.

In Table 1 are presented the VAF's for the off-line modeling (oTS) and for the two on-line modeling approaches (eTS and xTS). The best result is for the off-line case as expected, although the on-line case also performs very well. The best result for on-line modeling is presented by xTS but it costs more rules, 110, and consequently more computational time to update all the cluster centers and rules.

Note that the number of clusters is pre-defined to four in oTS and that the spread $r = 0.4$ for eTS. In xTS the spread is adaptive.

5 Conclusions and Future Works

This paper introduces a novel contribution to eye-to-hand visual servoing, based on on-line fuzzy modeling to obtain an uncalibrated visual servoing system. Two methods for on-line fuzzy modeling (eTS) and (xTS) have showed excellent results when compared with the results obtained for the off-line approach. This proves the validity of the proposed approach

Table 1: Results of the inverse fuzzy model, obtained for each joint with the three different approaches.

	Joint 1		Joint 2		Joint 3	
	VAF	Rules	VAF	Rules	VAF	Rules
oTS	96.4%	4	98.5%	4	98.5%	4
eTS	95.6%	50	98.4%	50	98.5%	50
xTS	95.6%	110	98.4%	110	98.6%	110

to uncalibrated visual servoing. With the robot-camera model being able to adapt on-line to changes in the environment, the robot can now operate on different conditions without the need for re-learning off-line the model or perform any kind of calibration.

As expected the off-line modeling produces better results when compared to both on-line approaches, because it takes into account all the data points. In on-line modeling, only the actual data point and the defined cluster centers are used to update the model, introducing or replacing rules. That is why the on-line approach performs not as good as the off-line approach. However we are convicted that the differences in the models, will not compromise the use of the model to control the robot.

As future work, the proposed evolving inverse fuzzy model approach will be extended to the tri-dimensional robot workspace and also will be tested to control the PUMA Robotic Manipulator for the straight line trajectory presented in Fig. 2, in order to verify the results already achieved for the inverse model obtained off-line. The influence of using the eTS or xTS models will also be studied, i.e. the trade-off between accuracy (VAF) and computational effort (number of rules), when controlling the robot.

Acknowledgements

This work was funded by FCT trough "Programa POCI2010, Unidade 46", subsidized by FEDER.

References

[1] S. Hutchinson, G. Hager, and P. Corke. A tutorial on visual servo control. *IEEE Transactions on Robotics and Automation*, 12(5):651–670, 1996.

[2] P.J. Sequeira Gonçalves, L.F. Mendonça, J.M. Sousa, and J.R. Caldas Pinto. Uncalibrated eye-to-hand visual servoing using inverse fuzzy models. *IEEE Transactions on Fuzzy Systems*, 16(2):341–353, 2008.

[3] P.J. Sequeira Gonçalves, A. Paris, C. Christo, J.M.C. Sousa, and J.R. Caldas Pinto. Uncalibrated visual servoing in 3d workspace. *Lecture Notes in Computer Science*, 4142:225–236, 2006.

[4] P. Angelov and D. Filev. An approach to online identification of takagi-sugeno fuzzy models. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 34:484–498, 2004.

[5] P. Angelov and Xiaowei Zhou. Evolving fuzzy systems from data streams in real-time. In *Proceedings of the International Symposium on Evolving Fuzzy Systems*, pages 29–35, Ambelside, Lake District, UK, 2006.

[6] B. Espiau, F. Chaumette, and P. Rives. A new approach to visual servoing in robotics. *IEEE Transactions on Robotics and Automation*, 8(3):313–326, 1992.

ISBN: 978-989-95079-6-8

[7] E. Malis, F. Chaumette, and S. Boudet. 2 $\frac{1}{2}$ d visual servoing. *IEEE Transactions on Robotics and Automation*, 15(2):238–250, 1999.

[8] F. Chaumette. Potential problems of stability and convergence in image-based and position-based visual servoing. *The Confluence of Vision and Control, D. Kriegman, G . Hager, A.S. Morse (eds.), LNCIS Series*, (1998):66–78, 2000.

[9] E. Cervera and P. Martinet. Combining pixel and depth information in image-based visual servoing. In *Proceedings of the Ninth International Conference on Advanced Robotics*, pages 445–450, Tokyo, Japan, 1999.

[10] J. Peipmeier, G. McMurray, and H. Lipkin. Uncalibrated dynamic visual servoing. *IEEE Trans. on Robotics and Automation*, 20(1):143–147, February 2004.

[11] I.H. Suh and T.W. Kim. Fuzzy membership function based neural networks with applications to the visual servoing of robot manipulators. *IEEE Transactions on Fuzzy Systems*, 2(3):203–220, 1994.

[12] R. Babuška. *Fuzzy Modeling for Control*. Kluwer Academic Publishers, Boston, 1998.

[13] J.M. Sousa and U. Kaymak. *Fuzzy Decision Making in Modeling and Control*. World Scientific Pub. Co., Singapore, 2002.

[14] D. E. Gustafson and W. C. Kessel. Fuzzy clustering with a fuzzy covariance matrix. In *Proceedings IEEE CDC*, pages 761–766, San Diego, USA, 1979.

[15] S. L. Chiu. Fuzzy model identification based on cluster estimation. *Journal of Intelligent Fuzzy Systems*, 2:267–278, 1994.

[16] J.M. Sousa, C. Silva, and J. Sá da Costa. Fuzzy active noise modeling and control. *International Journal of Approximate Reasoning*, 33:51–70, April 2003.

ISBN: 978-989-95079-6-8

On the reliability of the color gradient vector argument approach.

Belén Prados-Suarez[1] Pilar Sobrevilla[2] Eduard Montseny[3] Santiago Romaní[4]

1.Dept. of Software Engineering, University of Granada
Granada, Spain.
2.MAII Department, Technical University of Catalonia
Barcelona, Spain.
3.ESAII Department, Technical University of Catalonia
Barcelona, Spain.
4.Computer Engineering and Mathematics Department, Rovira I Virgili University
Tarragona, Spain.
Email: belenps@ugr.es, {pilar.sobrevilla,eduard.montseny}@upc.edu, santiago.romani@urv.net

Abstract— *Today many algorithms for monochrome images have been developed based on the information provided by the gradient vector. However, these algorithms can not be applied to color images without an accurate approximation of the color gradient vector argument. Moreover, there is a great number of applications that require to know with the maximum accuracy as possible the direction of the color variation in color images. This paper presents a new breakthrough in our work on the approximation of the argument of the color gradient vector. Taking into consideration that Hue and Saturation provide very useful basis for judging color uniformity, and that in the Smith's HSI color space there is a close relationship between chromaticity and how humans perceive colors, in our proposal the argument of the Chromaticity gradient vector is approached based on the Hue and Saturation information. Moreover, we present a fuzzy measure of the quality or reliability of the proposed approach that, considering the inherent vagueness of the chromaticity values of the image pixels, provides us with a credibility-accuracy degree of the obtained values.*

Keywords— Color gradient, Color image processing, Gradient argument, Gradient reliability, Gradient vector.

1 Introduction

Edge detection is one of the most important tasks in digital image processing and machine vision systems. Generally speaking, an edge implies the existence of abrupt changes or discontinuities in some visual property as light intensity, texture or color.

The advantage of color edge detection schemes over grayscale approaches is easily demonstrated by considering the fact that those edges that exist at the boundary between regions of different colors cannot be detected in grayscale images if there is no change in intensity. Among the existing approaches to detect color borders we can find active contour based methods [1], which use different approaches to perform the deformation process to dynamically adapt the active contour; multi-scale techniques [2], based on comparing the evolution of the borders through different scales; or morphological algorithms [3], that locate borders by computing the difference between a dilation and an erosion. However these methods just localize the maximum color variations or approach their magnitude, but not their direction.

A class of methods which not only localize the maximum color variations (edge localization) but also consider the direction of this variation (edge orientation) is the class of gradient based techniques.

The first methods proposed to compute the color gradient vector were mostly direct applications of the early intensity gradient methods. According to Lucchesse et al [4] these methods can be divided in those techniques that embed the variations of all color channels in a single measure [5, 6], and those that compute the gradient in each channel and then combine them according to certain criteria [7, 8]. A comprehensive analysis of color edge detectors can be found in [9].

Despite the importance and usefulness of the information provided by the argument of the gradient vector for many applications (texture characterization and analysis, 3D reconstruction, vehicle tracking and guidance,...), few techniques explicitly present and approach edge orientation. Among them we can find: proposals that approach the argument as the arc-tangent of the quotient between the vertical and horizontal partial derivatives [10]; Tensor based proposals [11] wherein the direction of the global color gradient is obtained from the tensor gradient components of multi-images regarded as vector fields; and Matrix based proposals [12] that approach the color gradient modulus as the highest eigenvalue in a given matrix, and its direction is given by the eigenvector associated to that eigenvalue.

The majority of the methods presented are focused on the RGB space. However, its lack of distinction between chromatic and achromatic information makes it unsuitable for obtaining edges in an image. A comparison of color edge detectors across several color spaces is presented in [13].

In this paper we present our ongoing work on the argument of the color gradient vector, focused on the Smith's HSI color space [14], which we began in [15]. In our proposal the argument of the *Chromaticity* gradient vector is approached based on the Hue and Saturation information, and we provide a fuzzy reliability degree of the obtained values for each edge pixel. This is quite useful, not only to evaluate the accuracy of our approach, but also to be used in applications that require to know the precision degree with which they are working. This is the case of astronomical images processing [16], the selection of optimal imaging planes in MRI [17] , or even the design of dichroic mirrors for LCDs [18]. In these applications

ISBN: 978-989-95079-6-8

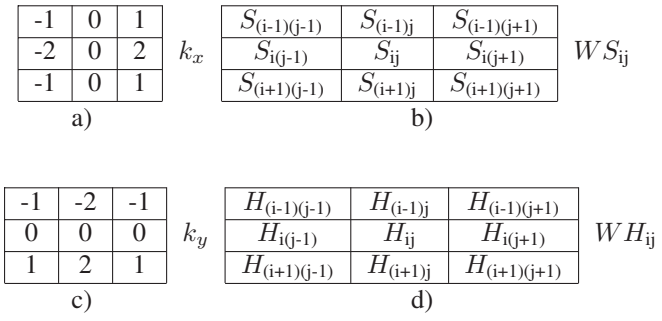


Figure 1: a) and c) Sobel Operators for the x and y directions, respectively. b) and d) 3x3 Saturation and Hue components windows of the source image, respectively.

an accurate contour extraction is necessary, and this is not possible without a good approach to the argument of the gradient vector, which also deals with the problems of imprecise contours.

Considering it, the paper is organized as follows. First, in section 2, we generalize the Sobel operators to compute the Hue and Saturation partial derivatives along the x and y directions. Then we present, in section 3, the way we combine the chromatic information to approach the partial derivatives; it will be used as directional components of the chromaticity gradient vector for approaching its argument. In section 4, we present the process and experiments carried out for obtaining a measure to evaluate the reliability of the color gradient vector argument approached. Finally we present some conclusions.

2 Proposed approximation of the Hue and Saturation partial derivatives

From the existing transformations to convert RGB components into perceptual ones, we have chosen Smith's model [14] based on its high independence between the three components, and its easier and faster calculation compared to other nonlinear models. This way we can consider color as a 1D+2D signal by decomposing it into its achromatic (I) and chromatic (H and S) components, what will provide us with a most useful basis for judging color uniformity.

In [15], having into consideration that Sobel masks (see Figure 1-a and 1-c) have slightly superior noise-suppression characteristics [13] than other operators, as Prewitt one, and that Sobel operator allows convolving separately the Hue and Saturation channels of the image, we presented an approximation to the partial derivatives of the chromatic color components extending this operator to approximate the variation of the Hue and Saturation components along the x and y directions.

2.1 Saturation First Order Derivatives

Given a pixel p_{ij} of the source image I, we will note by S_{ij} the magnitude of its Saturation, and WS_{ij} the window containing the Saturation values of the pixels falling into the 3x3 raster window centered at the pixel (see Figure 1-b).

Due to the fact that the areas of highest saturation gradient are where the saturation of the image changes rapidly over a few pixels, and are thus likely to represent edges, we can compute the Saturation partial derivatives, $\partial Sij/\partial x$ and $\partial Sij/\partial y$, by convolving kernels k_x and k_y with the WS_{ij} window , what provides us with the values given at equations (1) and (2).

2.2 Hue First Order Derivatives

Similarly to the Saturation case, we will note by H_{ij} the Hue magnitude of pixel p_{ij}, and WH_{ij} the window containing the Hue values of the pixels falling into the 3x3 raster window centered at that pixel (see Figure 1-d). However, while Euclidean distance is appropriate to compare the saturation values of two given colors, it is not the case of the Hue component. This is due to the ownership of Hue circularity, that is observed when this component is represented in a two-dimensional space. To overcome this problem, and consider the direction (the sign) of the difference between two Hue values we defined a Hue directed distance as follows:

Definition 2.1 *Let H_1 and H_2 be the hue values of pixels p_{i_1,j_1} and p_{i_2,j_2}, respectively. Then the* directed Hue distance *is defined as:*

$$\widehat{d}(H_1,H_2)=\begin{cases} H_2-H_1 & if\ \ |H_2-H_1|\leq 128 \\ H_2-H_1-255 & if\ \ H_2-H_1>128 \\ H_2-H_1+255 & if\ \ H_2-H_1<-128 \end{cases} \tag{3}$$

Then, convolution of kernels k_x and k_y are applied on the directed distances (3) obtained from the Hue values of WH_{ij}, instead of applying directly the Sobel operator over it.

As a result the Hue partial derivatives in both directions, $\partial Hij/\partial x$ and $\partial Hij/\partial y$, are given by equations (4) and (5).

3 Chromatic Gradient Vector Argument Approach

Once the Hue and Saturation partial derivatives have been approximated, next step consists in combining these values for obtaining the directional components of the *Chromatic Gradient Vector -CGV-*, or *Chromatic Partial Derivatives -CPD-*, which will be proportional to the directional components of the chromatic gradient vector, according to a given proportionality factor. In our case this problem has been figured out by mixing and merging previously obtained partial derivatives approximations according to next process.

1. First we consider two reference systems, one for the *x* direction, RS_x, and other for the *y* direction, RS_y, whose axes are the partial derivatives of *H* and *S* in the corresponding directions, ie: $RS_x=\{\partial H/\partial x, \partial S/\partial x\}$ and $RS_y=\{\partial H/\partial y, \partial S/\partial y\}$. On these systems we represent the values of the approaches obtained at the previous step, what provides us with two chromaticity vectors, one in the *x* direction, $\overrightarrow{Chr_{ij}x}=(\partial H_{ij}/\partial x, \partial S_{ij}/\partial x)$, and other in the *y* direction, $\overrightarrow{Chr_{ij}y}=(\partial H_{ij}/\partial y, \partial S_{ij}/\partial y)$ (line-dot vectors of Figure 2-a and 2-b). We also obtain de modules of these vectors, $\|\overrightarrow{Chr_{ij}x}\|$ and $\|\overrightarrow{Chr_{ij}y}\|$, which we call the *Directional Chromaticity Variations*.

2. Because the vectors obtained are represented into two different reference systems, to be able to mix and merge the information they contain is necessary to get a reference system wherein both vectors can be represented.

ISBN: 978-989-95079-6-8

$$\partial S_{ij}/\partial x = \frac{k_x * WS_{ij}}{8} = \frac{(S_{(i-1)(j+1)} - S_{(i-1)(j-1)}) + 2(S_{(i)(j+1)} - S_{(i)(j-1)}) + (S_{(i+1)(j+1)} - S_{(i+1)(j-1)})}{8} \quad (1)$$

$$\partial S_{ij}/\partial y = \frac{k_y * WS_{ij}}{8} = \frac{(S_{(i+1)(j-1)} - S_{(i-1)(j-1)}) + 2(S_{(i+1)(j)} - S_{(i-1)(j)}) + (S_{(i+1)(j+1)} - S_{(i-1)(j+1)})}{8} \quad (2)$$

$$\partial H_{ij}/\partial x = \frac{\widehat{d}(H_{(i-1)(j+1)}, H_{(i-1)(j-1)}) + 2\widehat{d}(H_{(i)(j+1)}, H_{(i)(j-1)}) + \widehat{d}(H_{(i+1)(j+1)}H_{(i+1)(j-1)})}{8} \quad (4)$$

$$\partial H_{ij}/\partial y = \frac{\widehat{d}(H_{(i+1)(j-1)}, H_{(i-1)(j-1)}) + 2\widehat{d}(H_{(i+1)(j)}, H_{(i-1)(j)}) + \widehat{d}(H_{(i+1)(j+1)}H_{(i-1)(j+1)})}{8} \quad (5)$$

To do it, due to axes $\partial H/\partial x = \partial S/\partial x$ and $\partial H/\partial y = \partial S/\partial y$ point to the higher potentials of *H* and *S*, and that both components grow equally within these axes, our proposal is to project the chromaticity vectors into them. So, our approach to the above mentioned *CPD* vectors, $\overrightarrow{\partial C_{ij}/\partial x}$ and $\overrightarrow{\partial C_{ij}/\partial y}$, is given by equations (6) and (7), where $\overrightarrow{u} = (\cos(45^o), \sin(45^o))$. These new vectors, that appear as dotted arrows in Figures 2-a and 2-b, will be used at next section to approach of the *CGVA*.

$$\overrightarrow{\partial C_{ij}/\partial x} = proj_{\overrightarrow{u}}\overrightarrow{Chr_{ij}x} \quad (6)$$

$$\overrightarrow{\partial C_{ij}/\partial y} = proj_{\overrightarrow{u}}\overrightarrow{Chr_{ij}y} \quad (7)$$

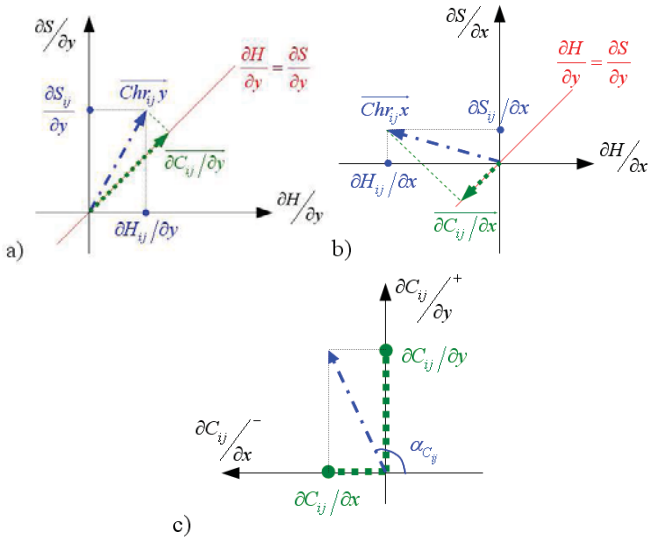


Figure 2: a, b: Calculation of the *CPD* vectors in each direction. c: Approach to the *CGVA*.

3. Applying the usual calculation of a projection, as we are interested in considering the argument of the *CPD* vectors, but considering their direction, the components of the *CGVA* approach are given using the scalar product, as depicted by equations (8) and (9).

$$\partial C_{ij}/\partial x = (\partial H_{ij}/\partial x, \partial S_{ij}/\partial x) \cdot (\sqrt{2}/2, \sqrt{2}/2) \quad (8)$$

$$\partial C_{ij}/\partial y = (\partial H_{ij}/\partial y, \partial S_{ij}/\partial y) \cdot (\sqrt{2}/2, \sqrt{2}/2) \quad (9)$$

Then, translating these values to the coordinate system of axes $\partial C_{ij}/\partial x$ and $\partial C_{ij}/\partial y$ we get the approach to the *Chromatic Gradient Vector*, and the argument of this vector, $\alpha_{C_{ij}} = \arctan(\frac{\partial C_{ij}/\partial y}{\partial C_{ij}/\partial x})$, provides us with the *CGVA* approach, as represented by Figure 2-c. This argument provides us with a direction that is perpendicular to the contour and so it is 0 for a vertical edge and increases for edges moving anti-clockwise of it.

4 Reliability degree of the color gradient argument approach

To assess the performance (goodness) of the proposed approach, we have defined a measure of reliability that must consider the sources of uncertainty of the factors having influence on the results obtained when the approach is applied on an image. So, for obtaining the reliability measure we have developed a set of experiments that allow us to study the relationship among: the error given by the approach with regard to the actual value of the argument, the *Chromatic Partial Derivatives*, and the *Directional Chromaticity Variations*. To carry out our experiments we have designed a set of images as explained in the next section.

4.1 *Obtaining the set of synthetic images*

To perform our experiments we built a set of images, consisting of a square with a circle in the center. The process for obtaining the images has been carried out as follows:

1. First, after dividing the RGB cube into 512 boxels, we select a point of each one of the boxes, so obtaining a set of 512 colors distributed almost homogeneously.

2. Then we generated a set of images using all the possible combinations of colors for background and foreground. For each image if a pixel is completely included in the background or in the circle, it is assigned the corresponding color. If the pixel belongs to background and foreground, after evaluating the area of the pixel belonging to each region, we assign to the pixel a color that is the weighted average of the two colors.

3. Finally, after smoothing the images using a 3x3 window to create a gradient in the boundary of the circles, the transformation from RGB to HSI is carried out.

4.2 *Experimental design*

For all the pixels in the borders of the circumferences, we have computed the argument of the *Chromaticity Gradient Vector* using the proposed approach. Since we know the position of the point and the center of the circle, we know the real value of the argument for each border pixel ($\arg_{ij}$). Having both

ISBN: 978-989-95079-6-8

values we can accurately evaluate the error at pixel (i, j) as $E^{ij} = |\arg_{ij} - \arg \alpha_{C_{ij}}|$.

In our experiments for obtaining the reliability degree we have studied the relationship among the aforementioned error, the *Directional Chromaticity Variations* ($\|\overrightarrow{Chr_{ij}x}\|$, and $\|\overrightarrow{Chr_{ij}y}\|$), and the smaller angles of the *Chromaticity Vectors* ($\overrightarrow{Chr_{ij}x}$, and $\overrightarrow{Chr_{ij}y}$) with the lines $\partial S/\partial x = -\partial H/\partial x$ and $\partial S/\partial y = -\partial H/\partial y$. These angles are given by:

$$E(\alpha_x^{ij}) = \begin{cases} |\arg \alpha_x^{ij} - 135| & \text{if } 45 < \arg \alpha_x^{ij} < 225 \\ |\arg \alpha_x^{ij} - 315| & else \end{cases} \quad (10)$$

$$E(\alpha_y^{ij}) = \begin{cases} |\arg \alpha_y^{ij} - 135| & \text{if } 45 < \arg \alpha_y^{ij} < 225 \\ |\arg \alpha_y^{ij} - 315| & else \end{cases} \quad (11)$$

Where α_x^{ij} and α_y^{ij} are the angles of the vectors $\overrightarrow{Chr_{ij}x}$ and $\overrightarrow{Chr_{ij}y}$ with the lines $\partial S/\partial x = \partial H/\partial x$ and $\partial S/\partial y = \partial H/\partial y$, respectively.

For each pixel (i, j) in the border of the circles we have obtained the values of E^{ij}, $\|\overrightarrow{Chr_{ij}x}\|$, $\|\overrightarrow{Chr_{ij}y}\|$, $E(\alpha_x^{ij})$, and $E(\alpha_y^{ij})$. Then we have analyzed and discussed the relationship between the average error and the values obtained for the four parameters considered by pairs.

4.3 Experimental analysis

For analyzing the obtained results we have represented in $3-D$ graphics the error values E^{ij} for each point (i, j) of the planes: $(E(\alpha_x^{ij}), E(\alpha_y^{ij}))$, $(\|\overrightarrow{Chr_{ij}x}\|, \|\overrightarrow{Chr_{ij}y}\|)$, and the planes $(\|\overrightarrow{Chr_{ij}x}\|, E(\alpha_x^{ij}))$, $(\|\overrightarrow{Chr_{ij}x}\|, E(\alpha_y^{ij}))$, $(\|\overrightarrow{Chr_{ij}y}\|, E(\alpha_y^{ij}))$, and $(\|\overrightarrow{Chr_{ij}y}\|, E(\alpha_x^{ij}))$. In these representations the gray level is proportional to the value of the average error for each point. So, the higher the gray level the lowest the error is.

As can be observed at the left image of Figure 3, when we consider the differences of the arguments, the values of the errors are small or large, and errors distribution indicates a great dependency among the error and the values of $E(\alpha_x^{ij})$, $E(\alpha_y^{ij})$.

When we consider the values of the *Directional Chromaticity Variations* (right image of Figure 3) the values of the errors are small and medium. However, the relationship among the error values and those of the parameters is very small, what is due to for almost each pair $(\|\overrightarrow{Chr_{ij}x}\|, \|\overrightarrow{Chr_{ij}y}\|)$ the errors are small.

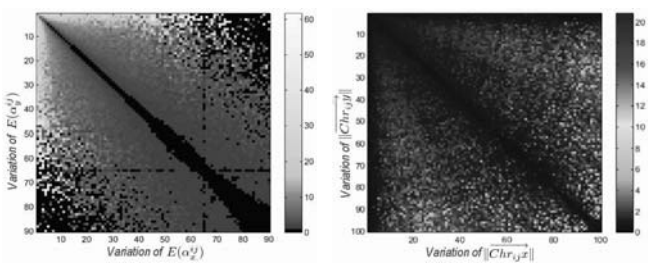


Figure 3: Representation of error magnitude for $E(\alpha_x^{ij})$ vs $E(\alpha_y^{ij})$ in the left image, and for $\|\overrightarrow{Chr_{ij}x}\|$ vs $\|\overrightarrow{Chr_{ij}y}\|$ in the right image.

In the case of the plane $(\|\overrightarrow{Chr_{ij}y}\|, E(\alpha_x^{ij}))$, and similarly for $(\|\overrightarrow{Chr_{ij}x}\|, E(\alpha_y^{ij}))$, the errors are mainly small. The medium and large errors that occur are highly concentrated, leaving almost the entire plane to the small errors (see left image of Figure 4), what implies a correlation among the error and the values of the parameters.

Finally, having a look at right image of Figure 4 it can be observed that in the case of the plane $(\|\overrightarrow{Chr_{ij}y}\|, E(\alpha_y^{ij}))$ appear small and big error values. In this case, while small errors are highly concentrated large errors are widely distributed, which indicates a certain relation between the error and the parameter values. A similar behavior is observed in the case of the plane $(\|\overrightarrow{Chr_{ij}x}\|, E(\alpha_x^{ij}))$

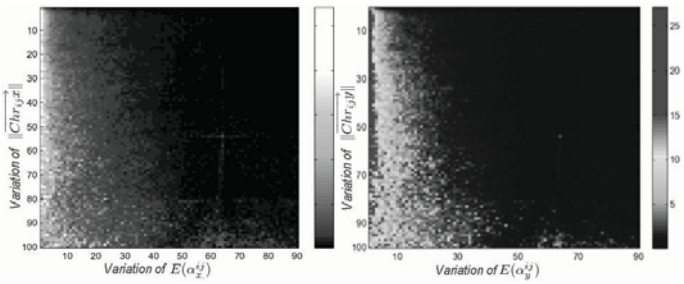


Figure 4: Left image represents the error magnitude for $\|\overrightarrow{Chr_{ij}^{x}}\|$ vs $E(\alpha_x^{ij})$.Right image represents the error magnitude for $\|\overrightarrow{Chr_{ij}^{y}}\|$ vs $E(\alpha_y^{ij})$.

As a consequence of previous analysis it can be deduced that the most important information for obtaining the reliability of the proposed approach is given by considering the values of $E(\alpha_x^{ij})$ vs $E(\alpha_y^{ij})$, and the less relevant information is provided by the values of $\|\overrightarrow{Chr_{ij}x}\|$ vs $\|\overrightarrow{Chr_{ij}y}\|$. On the other hand some information is provided when we consider the pairs $(E(\alpha_x^{ij}), \|\overrightarrow{Chr_{ij}y}\|)$ and $(E(\alpha_x^{ij}), \|\overrightarrow{Chr_{ij}x}\|)$.

4.4 Experimental results and Reliability degree

Based on the results obtained at previous step, we have considered the *Reliability Labels*: $Great$ (G), $Medium$ (M), $Small$ (S), and *Very Small* (VS). Then a reliability degree has been associated to each point of the considered plains, according to the error at each point, by means of the membership functions depicted at Figure 5.

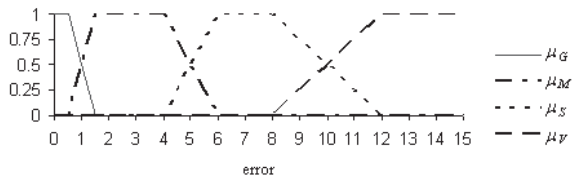


Figure 5: Reliability membership functions.

These membership functions have been obtained through a heuristic analysis of the results. The border values between each of the two central functions have been selected so that the error value of the 75%, the 85% and the 90% of the pixels are lower than such value.

The graphs obtained for each one of the cases considered at the previous step leads to the appearance of some lines. For each case, these lines act as boundaries between the different reliability degrees of the points in the corresponding plain.

ISBN: 978-989-95079-6-8

So, these lines will allow us to obtain the *Reliability Functions* of the proposed approach. Due to lack of space here only present the results obtained for the case of the $E(\alpha_x^{ij})$ vs $E(\alpha_y^{ij})$. For this case Figure 6 depicts the result of using the reliability degrees. Having a look at this image is it clear that appear some lines delimiting the different reliability regions, from $Great$ (dark grey level) to *Very Small* (very clear grey level). The lines obtained in this case are given by equations (12) to (22). Afterwards, these lines are used for obtaining the corresponding *Reliability Degrees* that, in this case are depicted by equations (23) to (26).

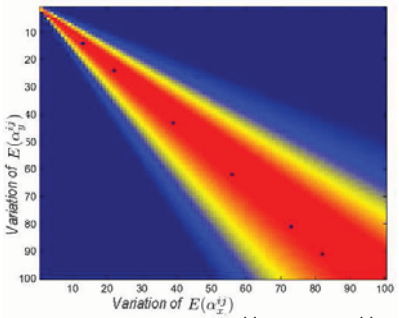


Figure 6: Graph obtained for $E(\alpha_x^{ij})$ vs $E(\alpha_y^{ij})$ using the reliability membership functions.

$$r_0(i,j) : E_x^{ij} = E_y^{ij} \tag{12}$$

$$r_1(i,j) : 25 \cdot E_x^{ij} - 27,5 \cdot E_y^{ij} + 6 = 0 \tag{13}$$

$$r_2(i,j) : 16 \cdot E_x^{ij} - 29 \cdot E_y^{ij} + 13 = 0 \tag{14}$$

$$r_3(i,j) : 10,6 \cdot E_x^{ij} - 29,1 \cdot E_y^{ij} + 16,65 = 0 \tag{15}$$

$$r_4(i,j) : 4,75 \cdot E_x^{ij} - 29,25 \cdot E_y^{ij} + 18,375 = 0 \tag{16}$$

$$r_5(i,j) : 2,5 \cdot E_x^{ij} - 29,5 \cdot E_y^{ij} + 13,5 = 0 \tag{17}$$

$$r_6(i,j) : 27,5 \cdot E_x^{ij} - 25 \cdot E_y^{ij} - 6 = 0 \tag{18}$$

$$r_7(i,j) : 29 \cdot E_x^{ij} - 16 \cdot E_y^{ij} - 13 = 0 \tag{19}$$

$$r_8(i,j) : 29,1 \cdot E_x^{ij} - 10,6 \cdot E_y^{ij} - 16,65 = 0 \tag{20}$$

$$r_9(i,j) : 29,25 \cdot E_x^{ij} - 4,75 \cdot E_y^{ij} - 18,375 = 0 \tag{21}$$

$$r_{10}(i,j) : 29,5 \cdot E_x^{ij} - 2,5 \cdot E_y^{ij} - 13,5 = 0 \tag{22}$$

Finally, for obtaining the *Reliability Degree* of the *Chromatic Gradient Vector Argument Approach* value for each point, the previous reliabilities for each case has to be aggregated. To do it, considering that all the individual values have influence in the final reliability degree, we can consider a *T-Norm* to perform the aggregation. Among the great variety of available *T-Norm* we propose to use the *Minimum* due to: its simplicity, fast computation, and that it is the greatest *T-Norm*.

5 Conclusions

In this paper we have presented an approach of the chromatic component for the argument of the gradient vector. After computing the hue and saturation partial derivatives on each direction, this information has been merged and combined through projections for obtaining the directional components of the chromaticity gradient vector used to approach the argument.

We have also presented a study of the error variability produced by the proposed approach with regard to the four main parameters used within the proposed approach. This will allow us to improve the approximation, as well as obtaining an approach for the module of the gradient vector.

A very important point to highlight is that, as a result of the analysis of the error variability, we obtained a value that provides the credibility of the value provided by the approximation. This is very important because it will allow to use the value of the gradient vector argument for each point knowing its reliability degree.

Acknowledgment

This paper has been partially supported by the projects P07-TIC-03175, P06-TIC-01570 and TIN2007-68063.

References

[1] A. Maalouf, A. Maalouf, P. Carre, B. Augereau, and C. Fernandez-Maloigne. Foveal wavelet-based color active contour. In P. Carre, editor, *Image Processing, 2007. ICIP 2007. IEEE Int. Conf. on*, volume 1, pages I – 245–I – 248, 2007.

[2] I. Vanhamel, I. Vanhamel, I. Pratikakis, and H. Sahli. Multiscale gradient watersheds of color images. *Image Processing, IEEE Transactions on*, 12(6):617–626, 2003.

[3] A.N. Evans and A.N. Evans. Morphological gradient operators for colour images. In *Image Processing, 2004. ICIP '04. 2004 Int. Conf. on*, volume 5, pages 3089–3092 Vol. 5, 2004.

[4] L. Lucchese and S. K. Mitra. Color image segmentation: A state-of-art survey. In *Proc. Indian Nat. Sci. Acad. (INSA-A)*, volume 67-A, pages 207–221, 2001.

[5] M. Chapron. A color edge detector based on statistical rupture tests. In *Image Processing, 2000. Proc. 2000 Int. Conf. on*, volume 2, pages 820–823 vol.2, 2000.

[6] M. A. Ruzon and C. Tomasi. Color edge detection with the compass operator. *Proc. IEEE Conf. Computer Vision and Pattern Recognition.*, 2:160–166, 1999.

[7] M. Heddley. Segmentation of color images using spatial and color space information. *J. Electron. Imag.*, 1(4):374–380, 1992.

[8] Jianping Fan, Jianping Fan, D.K.Y. Yau, A.K. Elmagarmid, and W.G. Aref. Automatic image segmentation by integrating color-edge extraction and seeded region growing. *Image Processing, IEEE Transactions on*, 10(10):1454–1466, 2001.

[9] A. Koschan and M. Abidi. Detection and classification of edges in color images. *Signal Processing Magazine, IEEE*, 22(1):64–73, 2005.

[10] T. Carron and P. Lambert. Color edge detector using jointly hue, saturation and intensity. In *Image Processing, 1994. Proc. ICIP-94., IEEE Int. Conf.*, volume 3, pages 977–981, 1994.

[11] Silvano Di Zenzo. A note on the gradient of a multi-image. *Computer Vision, Graphics, and Image Processing*, 33(1):116–125, 1986.

[12] Anwander-A, Neyran-B, and Baskurt-A. Multiscale color gradient for image segmentation. *Traitement du Signal*, 18(2):129–42, 2001.

[13] S. Wesolkowski, S. Wesolkowski, M.E. Jernigan, and R.D. Dony. Comparison of color image edge detectors in multiple color spaces. In M.E. Jernigan, editor, *Image Processing, 2000. Proc. 2000 Int. Conf. on*, volume 2, pages 796–799 vol.2, 2000.

ISBN: 978-989-95079-6-8

$$\mu_G(E_x^{ij},E_y^{ij})=\begin{cases}0 & \text{if } r_8(i,j)<0 \text{ or } r_2(i,j)>0\\ 1 & \text{if } r_6(i,j)>0 \text{ and } r_1(i,j)<0\\ \dfrac{d((E_x^{ij},E_y^{ij}),r_7(i,j))}{d((E_x^{ij},E_y^{ij}),r_7^{ij})+d((E_x^{ij},E_y^{ij}),r_6(i,j))} & \text{if } r_6(i,j)<0 \text{ and } r_7(i,j)>0\\ \dfrac{d((E_x^{ij},E_y^{ij}),r_2(i,j))}{d((E_x^{ij},E_y^{ij}),r_2^{ij})+d((E_x^{ij},E_y^{ij}),r_1(i,j))} & \text{else}\end{cases} \tag{23}$$

$$\mu_M(E_x^{ij},E_y^{ij})=\begin{cases}0 & \text{if } r_8(i,j)<0 \text{ or } r_3(i,j)>0\\ 1 & \text{if } (r_6(i,j)<0 \text{ and } r_7(i,j)>0)\\ & \textit{or } (r_1(i,j)>0 \text{ and } r_2(i,j)<0)\\ \dfrac{d((E_x^{ij},E_y^{ij}),r_0(i,j))}{d((E_x^{ij},E_y^{ij}),r_0^{ij})+d((E_x^{ij},E_y^{ij}),r_6(i,j))} & \text{if } r_0(i,j)>0 \text{ and } r_6(i,j)<0\\ \dfrac{d((E_x^{ij},E_y^{ij}),r_0(i,j))}{d((E_x^{ij},E_y^{ij}),r_0^{ij})+d((E_x^{ij},E_y^{ij}),r_1(i,j))} & \text{if } r_0(i,j)<0 \text{ and } r_1(i,j)>0\\ \dfrac{d((E_x^{ij},E_y^{ij}),r_8(i,j))}{d((E_x^{ij},E_y^{ij}),r_8^{ij})+d((E_x^{ij},E_y^{ij}),r_7(i,j))} & \text{if } r_7(i,j)<0 \text{ and } r_8(i,j)>0\\ \dfrac{d((E_x^{ij},E_y^{ij}),r_3(i,j))}{d((E_x^{ij},E_y^{ij}),r_3^{ij})+d((E_x^{ij},E_y^{ij}),r_2(i,j))} & \text{else}\end{cases} \tag{24}$$

$$\mu_S(E_x^{ij},E_y^{ij})=\begin{cases}0 & \text{if } (r_{10}(i,j)<0 \text{ or } r_5(i,j)>0)\\ & \textit{or } (r_7(i,j)>0 \text{ and } r_2(i,j)<0)\\ 1 & \text{if } (r_8(i,j)<0 \text{ and } r_9(i,j)>0)\\ & \textit{or } (r_3(i,j)>0 \text{ and } r_4(i,j)<0)\\ \dfrac{d((E_x^{ij},E_y^{ij}),r_7(i,j))}{d((E_x^{ij},E_y^{ij}),r_7^{ij})+d((E_x^{ij},E_y^{ij}),r_8(i,j))} & \text{if } r_7(i,j)<0 \text{ and } r_8(i,j)>0\\ \dfrac{d((E_x^{ij},E_y^{ij}),r_2(i,j))}{d((E_x^{ij},E_y^{ij}),r_2^{ij})+d((E_x^{ij},E_y^{ij}),r_3(i,j))} & \text{if } r_2(i,j)>0 \text{ and } r_3(i,j)<0\\ \dfrac{d((E_x^{ij},E_y^{ij}),r_{10}(i,j))}{d((E_x^{ij},E_y^{ij}),r_{10}^{ij})+d((E_x^{ij},E_y^{ij}),r_9(i,j))} & \text{if } r_9(i,j)<0 \text{ and } r_{10}(i,j)>0\\ \dfrac{d((E_x^{ij},E_y^{ij}),r_5(i,j))}{d((E_x^{ij},E_y^{ij}),r_5^{ij})+d((E_x^{ij},E_y^{ij}),r_4(i,j))} & \text{else}\end{cases} \tag{25}$$

$$\mu_{VS}(E_x^{ij},E_y^{ij})=\begin{cases}0 & \text{if } r_9(i,j)>0 \text{ and } r_4(i,j)><0\\ 1 & \text{if } r_{10}(i,j)>0 \text{ and } r_5(i,j)<0\\ \dfrac{d((E_x^{ij},E_y^{ij}),r_9(i,j))}{d((E_x^{ij},E_y^{ij}),r_9^{ij})+d((E_x^{ij},E_y^{ij}),r_{810}(i,j))} & \text{if } r_9(i,j)<0 \text{ and } r_{10}(i,j)>0\\ \dfrac{d((E_x^{ij},E_y^{ij}),r_4(i,j))}{d((E_x^{ij},E_y^{ij}),r_4^{ij})+d((E_x^{ij},E_y^{ij}),r_5(i,j))} & \text{else}\end{cases} \tag{26}$$

[14] A.R. Smith. Color gamut transform pairs. *Computer Graphics*, 12:12–19, 1978.

[15] P. Sobrevilla B. Prados-Suarez, E. Montseny and S. Romanís. An approach to color gradient vector argument considering data uncertainty. *in Procc Joint Eusflat-LFA Conf.*, 2:705–708, 2000.

[16] Changbom-Park and Yun-Young-Choi. Morphology segregation of galaxies in color-color gradient space. *Astrophysical Journal, Letters*, 635(1):L29–32, 2005.

[17] Levine RA Okada RD Brady TJ. Dinsmore RE, Wismer GL. Magnetic resonance imaging of the heart: positioning and gradient angle selection for optimal imaging planes. *American journal of roentgenology, radium therapy and nuclear medicine.*, 143(6):1135–42, 1984.

[18] GU P. GUAN A. LU W., ZHENG Z. Gradient angle design of dichroic mirrors used in lcd projection display. In *Proc Int Disp Workshops*, volume 11, pages 1697–1699, 2004.

ISBN: 978-989-95079-6-8

Evolutionary Robot Vision and Fuzzy Evaluation for Natural Communication of Partner Robots

Akihiro Yorita Naoyuki Kubota

Graduate School of System Design, Tokyo Metropolitan University,
Hino, Tokyo 191-0065, Japan
e-mail:yorita-akihiro@sd.tmu.ac.jp, kubota@tmu.ac.jp

Abstract— *This paper proposes a method of evolutionary robot vision based on a steady-state genetic algorithm and fuzzy evaluation. In order to improve the communication capability of human-friendly partner robots, the perception of human face should be performed as correctly as possible. First, we discuss the concept of evolutionary robot vision in dynamic environments. Next, we propose growing neural gas for preprocessing as a bottom-up processing, and steady-state genetic algorithm for template matching in human face recognition as a top-down processing. In order to improve the performance of the human face recognition, we use fuzzy evaluation for evaluating the degree of human face. Finally, we show several experimental results and discuss the effectiveness of the proposed method.*

Keywords— Face Recognition, Evolutionary Computation, Robot Vision, Fuzzy Theory, Partner Robots

1 Introduction

Natural communication and interactions among people and robots have been widely researched [1,2]. Furthermore, various types of human-friendly robots such as pet robots, amusement robots, and partner robots have been developed to communicate with people [3,4]. However, there are still many problems to realize natural communication between people and robots. For example, it is difficult for a robot to identify the speaking person and the place interacting with the person. As a result, the content of utterances from the robot might not be suitable for the person and place.

Relevance theory is helpful to discuss the natural communication between a human and robot [21]. Relevance theory is based on a definition of relevance and two principles of the relevance. One is a cognitive principle that human cognition is geared to the maximization of relevance. The other is a communicative principle that utterances create expectations of optimal relevance. The central claim of relevance theory is that the expectations of relevance raised by an utterance are precise enough, and predictable enough to guide the hearer towards the speaker's meaning. Furthermore, in relevance theory, human thought is shared between two people rather than transmitted. Each person has his or her own cognitive environment. An important role of utterances, facial direction, pointing behaviors, and gestures is to make the hearer pay attention to a specific target object or person. As a result, the cognitive environment of the hearer can be enlarged by the utterances or gestures. The cognitive environment shared between two people is called a mutual cognitive environment. Based on the above discussion, a robot also should have such a cognitive environment, and the robot should keep updating the cognitive environment according to the current perception through the interaction with a human in order to realize the natural communication.

Furthermore, the utterance capability of a robot can be applied for preventing dementia of elderly people. Robotic conversation can activate the brain of such elderly people and can improve their concentration and memory abilities. Nursing care for the elderly people can be expected to keep their health by having conversations with robots. However, it is very difficult to continue the meaningful and attractive conversations with robots. Therefore, such a robot requires adaptive perceptual systems to communicate with a human flexibly, and adaptive action systems to learn human behaviors. To realize the learning through interaction with people, we must consider a total architecture of the cognitive development. The cognitive development for robots has been discussed in the fields such as cognitive robotics and embodied cognitive science [8-10]. In the previous research of cognitive robotics, many researchers have proposed the learning methods for the achievement of joint attention, imitative learning, linguistic acquisition from the viewpoints of babies and infants [6,10]. On the other hand, we focus on the refinement of associative memory by using symbolic information used for utterances and patterns based on visual information through interaction with people as cognitive development of robots. We proposed the concept of structured learning and discussed the importance of total architecture of the learning mechanism [30]. However, we did not discuss the performance of the human detection so much. In the previous works, we proposed a simple method of people tracking based on the combination of skin color and hair color, and we have a problem of misdetection of people by objects with similar color combination in the background image. In this paper, we propose a method for detecting a human face based on evolutionary computation and fuzzy evaluation in order to improve the performance of people tracking. The both of evolutionary computation and fuzzy theory are useful and practical in the search under the environment including noise. We proposed the concept of evolutionary robot vision based on the analogy between visual perception and evolutionary search [32]. We apply the concept of evolutionary robot vision and fuzzy evaluation for people tracking.

The paper is organized as follow. In the section 2, we explain the concept of evolutionary robot vision. Next, in the section 3 we propose a growing neural gas for color extraction and a steady-state genetic algorithm with fuzzy inference for face recognition. Section 4 shows several experimental results and discuss the effectiveness of the proposed method.

2 Evolutionary Robot Vision

2.1 Partner robots

We developed two types of partner robots; a mobile PC called MOBiMac and a human-like robot called Hubot in order to realize the social communication with a human [25,26] (Fig.1). Each robot has two CPUs and many sensors such as CCD camera, microphone, and ultrasonic

ISBN: 978-989-95079-6-8

sensors. Therefore, the robots can conduct image processing, voice recognition, target tracing, collision avoidance, map building, and imitative learning.

We have applied steady-state genetic algorithm (SSGA), spiking neural networks (SNN), self-organizing map (SOM), and others for human detection, motion extraction, gesture recognition, and shape recognition based on image processings [25-28]. However, the image processing takes much computational time and cost. Therefore, we discuss the applicability of fuzzy theory and evolutionary computation in robot vision.

2.2. Active Vision for Robots

Computer vision is a research stream on image processing, image understanding, image recovery and others on a computer [9]. The quality of image depends strongly on the lighting condition related with a camera system. Active vision is often used to improve the robustness and flexibility and to eliminate the ill-posed conditions based on the control of camera systems [9]. Robot vision is deeply related with active vision, because a problem on the perception and action of a robot must be solved at the same time. A robot takes actions to perceive the environment when the robot does not know the environment much. Therefore, the robot vision is based on the time-series of image processing, not the processings on a single image. Various technologies for image processings are required for realizing the robot vision, *e.g.*, color processing, target detection, template matching, shape recognition, motion extraction, and optical flow. Recently, evolutionary computation has been applied to improve the performance of image processing. We also have discussed the applicability of the evolutionary computation in robot vision [32-34]. In fact, we proposed a method of people detection, people tracking and gesture recognition.

2.3 Evolutionary Computation for Robot Vision

Evolutionary computation (EC) is a field of simulating evolution on a computer. Evolutionary optimization methods are fundamentally iterative generation and alternation processes of multiple candidate solutions. The optimization is done by the multi-point search operating on a set of individuals, which is called a population. First, we discuss the role of evolutionary search in dynamic environments. Figure 2 shows the temporal patterns of spatial changes in dynamic environments where the vertical axis indicates the state of environmental conditions represented as a value. If the search speed of EC is faster than the changing speed of the environmental conditions, EC can obtain feasible solutions in the facing environmental conditions. However, EC should be adaptive to the facing environmental conditions if the environmental changes can be observed. If the environmental change is very slow and the change is small (Fig.2 (a)), the mutation range should be large according to the amount of the environmental change. Basically, this kind of change can be considered as some noise in a stationary environment. The environment of Fig.2 (b) includes big changes, but the environmental condition is stationary after a big change. If the big change can be observed, most of candidate solutions should be replaced with randomly generated candidate solutions. The environment of Fig.2 (c) is changing non-stationarily with both features of Figs.2 (a) and (b). In general, the change of visual images corresponds to the environment of Fig.2 (c). The visual image of a mobile robot changes according to both the dynamics of environmental changes and the robotic motion. Therefore,

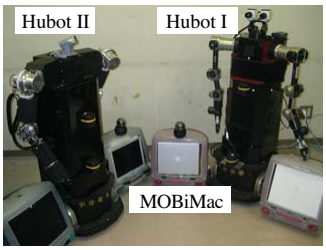


Figure 1: Partner robots; MOBiMac and Hubot

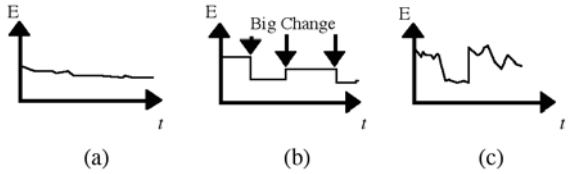
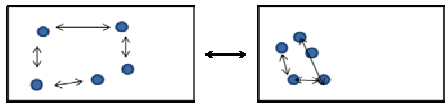
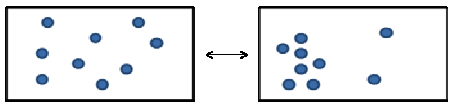


Figure 2: Patterns of changes in dynamic environments

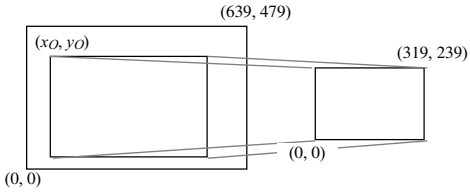


(a) Movement of attentive points in the visual search



(b) Distribution of candidate solutions in evolutionary search

Figure 3: Comparison of visual search and evolutionary search



(a) Original Image (b) Attention range

Figure 4: Preprocessing for human face detection

we discuss visual perception based on active robot vision and EC from the viewpoint of human visual perception.

Figure 3 shows the comparison between the visual search and evolutionary search. The left and right figures show the distribution of search points in the focused search and distributed search, respectively. The visual search controls the searching area based on the fast movement of attentive point. The region of interest (ROI) for the information extraction is deeply related with the search of geometrical features included in the visual target. The search results are reflected to the next visual search. On the other hand, the evolutionary search controls the searching area based on the selection pressure to candidate solutions. If the selection pressure is high, the candidate solutions are centralized toward the better candidate solutions. Otherwise, candidate solutions are globally distributed in the search space. The next search points in the evolutionary search are generated by crossover and mutation. The search in ROI is mainly performed by mutation and local search, while a new search point for ROI is generated by crossover. The degree of interest is calculated by the fitness value. If the fitness value is high, the focused search should be performed. We

ISBN: 978-989-95079-6-8

apply a steady-state genetic algorithm (SSGA) to realize the continuous and real-time search for the robot vision like human visual perception in a dynamic environment, because SSGA can easily obtain feasible solutions through environmental changes with low computational cost.

2.4 Image Processing for Pattern Recognition

Various types of pattern matching methods such as template matching, cellular neural network, neocognitron, and dynamic programming (DP) matching, have been applied for the human detection in image processing. In general, pattern matching is composed of two steps of target detection and target recognition. The aim of target detection is to extract a target candidate from an image (segmentation), and the aim of the target recognition is to identify the target from classification templates. Furthermore, we can discuss the pattern matching from the viewpoints of the bottom-up processing and top-down processing. The candidate detection is considered as the bottom-up processing based on the color distribution in segments of an image, while the target recognition is considered as top-down processing based on the similarity between the extracted target candidate and classification templates. The synthetic combination of bottom-up processing and top-down processing realizes the efficient and effective search. In this paper, we apply growing neural gas and SSGA for bottom-up processing and top-down processing, respectively.

3. Human Tracking based on Evolutionary Robot Vision

3.1 Growing Neural Gas for Bottom-up Preprocessing

An image is a set of pixels with color information. The segmentation into target candidates based on color distribution is very important to reduce the computational cost of direct template matching. Therefore, we apply an unsupervised clustering method for the segmentation based on color distribution in an image. Growing neural gas (GNG) is a competitive learning network as one of variants of SOM [22] used as a clustering method.

The learning algorithm of GNG is shown as follows.

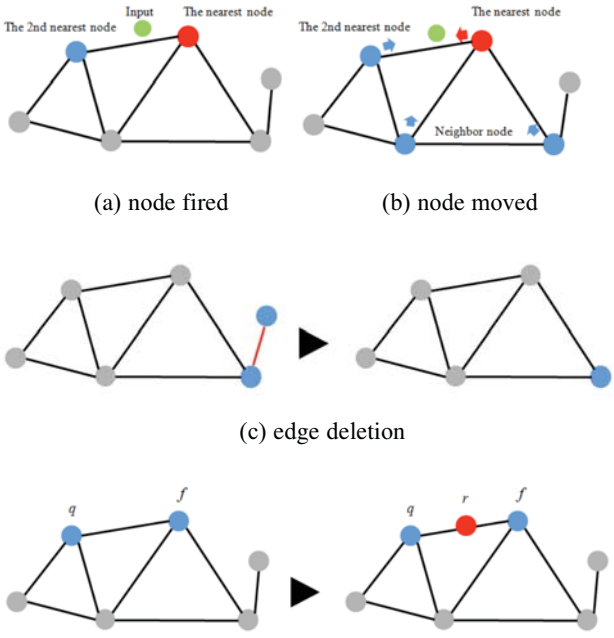
w_i : the nth dimensional vector of a node ($w_i \in \mathbf{R}^n$)
A : a set of nodes
N_i : a set of nodes connected to the ith node
c : a set of edges
$a_{i,j}$: the age of the edge between ith and jth node

Step 0. Generate two units at random position, w_{C1}, w_{C2} in **Rn**. Initialize the connection set.
Step 1. Generate at random an input data v according to $p(v)$.
Step 2. Determine the nearest unit s_1 and the second-nearest unit s_2 by

$$s_1 = \arg\min_{i \in A} \|v - w_i\| \quad (1)$$

$$s_2 = \arg\min_{i \in A \setminus \{s_1\}} \|v - w_i\| \quad (2)$$

where v is composed of the position (x, y) and color information (R, G, B) of color pixel on the image(see Fig.5(a)).
Step 3. If a connection between s_1 and s_2 does not yet exist , create it. Set the age of the connection between s_1 and s_2 to zero.



(a) node fired (b) node moved

(c) edge deletion

(d) node addition

Figure 5: How to learn GNG nodes and edge



Original Image GNG for Human shape recognition

Figure 6: An example of GNG

$$a_{s_1,s_2} = 0 \quad (3)$$

Step 4. Add the squared distance between the input signal and the winner to a local error variable E_{s1}(see Fig.5(b)).

$$E_{s_1} \leftarrow E_{s_1} + \|v - w_{s_1}\|^2 \quad (4)$$

Step 5. Adapt the reference vectors of the winner and its direct topological neighbors by the learning rate η_1^G and η_2^G, respectively.

$$w_{s_1} \leftarrow w_{s_1} + \eta_1^G \cdot (v - w_{s_1}) \quad (5)$$

$$w_j \leftarrow w_j + \eta_2^G \cdot (v - w_j) \quad if\ c_{s_1,j} = 1 \quad (6)$$

Step 6. Increment the age of all edges emanating from s_1.

$$a_{s_1,j} \leftarrow a_{s_1,j} + 1 \quad if\ c_{s_1,j} = 1 \quad (7)$$

Step 7. Remove edges with the age larger than $a_{\max}$. If units have no more emanating edges after this, remove those units(see Fig.5(c)).
Step 8. If the number of input signals generated so far is an integer multiple of a parameter λ, insert a new unit as follows(see Fig.5(d)).
Step 8-1. Determine the unit q with the maximum accumulated error.

$$q = \arg\max_{i \in A} E_i \quad (8)$$

Step 8-2. Determine the unit f with the maximum accumulated error among the neighbors of q.

$$f = \arg\max_{c \in N_q} E_c \quad (9)$$

Step 8-3. Add a new unit r to the network and interpolate its reference vector from q and f.

ISBN: 978-989-95079-6-8

$$w_r = 0.5 \cdot \left(w_q + w_f \right) \quad (10)$$

Step 8-4. Insert edges connecting the new unit r with units q and f, and remove the original edge between q and f.

Step 8-5. Decrease the error variables of q and f by the discount rate α.

$$E_q \leftarrow (1-\alpha)E_q \text{ and } E_f \leftarrow (1-\alpha)E_f \quad (11)$$

Step 8-6. Interpolate the error variable of r from q and f

$$E_r = 0.5 \cdot \left(E_q + E_f \right) \quad (12)$$

Step 9 Decrease the error variables of all units

$$E_c \leftarrow (1-\beta)E_c, \ c \in A \quad (13)$$

Step 10 If a termination condition (e.g., some performance measure) is not yet fulfilled continue with step 2.

Figure 6 shows an example of the learning of GNG. In this way, the color distribution can be extracted from the image by using GNG.

3.2 People Detection and Tracking as Top-down Processing

SSGA simulates a continuous model of the generation, which eliminates and generates a few individuals in a generation (iteration) [13]. The genotype is represented by $g_{i,j}$ (i=1,2,..., G, j=1,2,..., M) and fitness value is represented by f_i. One iteration is composed of selection, crossover, and mutation. The worst candidate solution is eliminated ("Delete least fitness" selection strategy), and is replaced with the candidate solution generated by the crossover and the mutation.

We use the elitist crossover and adaptive mutation [15]. The elitist crossover randomly selects one individual and generates an individual by combining genetic information from the selected individual and the best individual with the crossover probability. If the crossover probability is satisfied, the elitist crossover is performed. Otherwise, a simple crossover is performed between two randomly selected individuals. Next, the following adaptive mutation is performed to the generated individual,

$$g_{i,j} \leftarrow g_{i,j} + \left(\alpha_j \cdot \frac{f_{\max} - f_i}{f_{\max} - f_{\min}} + \beta_j \right) \cdot N(0,1) \quad (14)$$

where f_i is the fitness value of the ith individual, f_{max} and f_{min} are the maximum and minimum of fitness values in the population; $N(0,1)$ indicates a normal random variable with a mean of zero and a variance of one; α_j and β_j are the coefficients ($0<\alpha_j<1.0$) and offset ($\beta_j>0$), respectively. In the adaptive mutation, the variance of the normal random number is relatively changed according to the fitness values of the population in case of maximization problems.

The robot must recognize a human face from complex background speedily. Therefore, we use SSGA for human detection as one of search methods. The human face candidate positions based on human skin and hair colors are extracted by SSGA with template matching. Figure 7 shows a candidate solution of a template used for detecting a human face. A template is composed of numerical parameters of $g_{i,1}^H$, $g_{i,2}^H$, $g_{i,3}^H$, and $g_{i,4}^H$. The number of individuals is G^H. The initial population of SSGA for human detection at the discrete time step t is updated by using the reference vector of GNG in addition to the candidate solutions obtained at the previous time step t-1. The fitness value of the ith individual is calculated by the following equation,

$$f_i^H = C_{Skin}^H + C_{Hair}^H + \eta_1^H \cdot C_{Skin}^H \cdot C_{Hair}^H - \eta_2^H \cdot C_{Other}^H \quad (15)$$

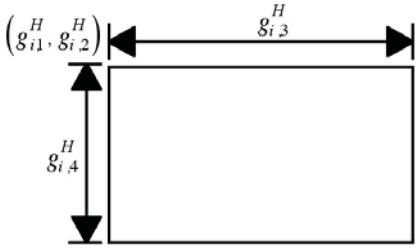
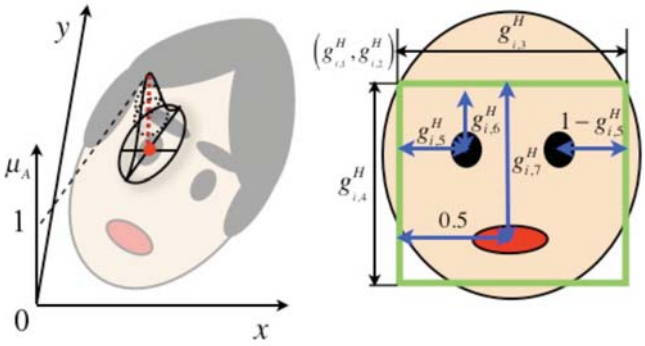


Figure 7: A template used for human detection in SSGA



(a) Fuzzy Evaluation based on Gaussian membership (b) The position of facial landmarks

Figure 8: A template used for face recognition

where C_{Skin}^H, C_{Hair}^H and C_{Other}^H indicate the numbers of pixels of the colors corresponding to human skin, human hair, and other colors, respectively; η_1^H and η_2^H are the coefficients (η_1^H, $\eta_2^H>0$). Therefore, this problem results in the maximization problem. The iteration of SSGA is repeated until the termination condition is satisfied. Here SSGA for the human detection is called SSGA-H.

Since SSGA-H extracts the area of skin colors and hair colors in the human detection, various objects except humans might be detected. Therefore, the human tracking is performed according to the time series position of the ith human candidate ($g_{i,1}^H$, $g_{i,2}^H$) obtained by SSGA-H. The position of the jth human candidate in the human tracking ($X_{k,1}$, $X_{k,2}$) is updated by the nearest human candidate position within the tracking range. In addition, the width and height of the human candidate for the human tracking ($X_{k,3}$, $X_{k,4}$) are updated by the size of the detected human ($g_{i,3}^H$, $g_{i,4}^H$). The update is performed as follows (j=1,2,3,4);

$$X_{k,j} = (1-\lambda)X_{k,j} + \lambda \cdot x_{i,j} \quad (16)$$

Furthermore, the time counter for the reliability of human tracking is used. If the human candidate position in the human tracking is performed, the time counter is incremented. Otherwise, the time counter is decremented. If the time counter is larger than the threshold (HT), the human count is started. Sometimes, several human candidates are close each other, because several human candidates in a single human can be generated by the human detection. Therefore, the removal processing is performed when human candidates are coexisting within the tracking range.

3.3 Fuzzy Evaluation for Face Recognition

The human detection based on color distribution sometime extracts something with similar color distribution. In order to improve the performance of human face recognition, we

ISBN: 978-989-95079-6-8

Table 1: Initial values used in Gaussian membership function

	$a_{h,1}$	$a_{h,2}$	$b_{h,1}$	$b_{h,2}$
Right eye	0.2	0.2	0.4	0.2
Left eye	0.8	0.2	0.4	0.2
Mouth	0.5	0.6	0.4	0.37

Table 2: Search range of $g_{i,5}^H, g_{i,6}^H, g_{i,7}^H$

	$g_{i,5}^H$	$g_{i,6}^H$	$g_{i,7}^H$
min	0.1	0.15	0.7
max	0.35	0.35	0.9

can use the color information of facial landmarks. The human detection by SSGA can be considered as a coarse search, and the extraction of facial landmarks can be considered as a fine search. We apply fuzzy evaluation for face recognition based on the position of facial landmarks (Fig.8 (a)). We use the position of eyes and mouth. We use

$$\mu_{A(h,i,j)} = \begin{cases} \exp\left(-\frac{(x-a_{h,1})^2}{2b_{h,1}^{\ 2}} - \frac{(y-a_{h,2})^2}{2b_{h,2}^{\ 2}} \right) & if\ p(i,j) = c_h \\ 0 & otherwise \end{cases} \tag{17}$$

where (x,y) is the normalized position of the pixel (i,j) on the image; $p(i,j)$ is the color ID of a the pixel (i,j) on the image; c_h is the color ID of the hth facial landmark is the center of a facial landmark; $(a_{h,1},a_{h,2})$ and $(b_{h,1},b_{h,2})$ are the normalized position and size of the hth facial landmark in the template candidate extracted by SSGA. Therefore, this value is high if the color pixel corresponding to the facial landmark is near with the center of the facial landmark. Therefore, we can evaluate the degree of existing each facial landmark as follows;

$$f_{Land,h} = \sum_{(i,j)\in g_k} \mu_{A(h,i,j)} \tag{18}$$

where g_k is the template of the kth candidate solution in SSGA; h is facial the landmark ID. Furthermore, we can evaluate the degree of face as follows;

$$f_{Face} = \prod_{h=1}^{H} f_{Land,h} \tag{19}$$

where H is the number of facial landmarks. We use right eye (h=1), left eye (h=2) and mouth (h=3) for the evaluation (H=3). Figure 8 (b) shows the positions of facial landmarks where $(g_{i,5}^H, g_{i,6}^H)$ is the position of the right eye and $(0.5, g_{i,7}^H)$ is the position of the mouth. Therefore, $(a_{1,1},a_{1,2}) = (g_{i,5}^H, g_{i,6}^H)$, $(a_{2,1},a_{2,2}) = (1-g_{i,5}^H, g_{i,6}^H)$, and $(a_{3,1},a_{3,2}) = (0.5, g_{i,7}^H)$. However, the position of each facial landmark is peculiar to a person. Therefore, the position of membership function corresponding to each facial landmark should be updated according to the detected person. As a result, the number of parameters for human detection is 7.

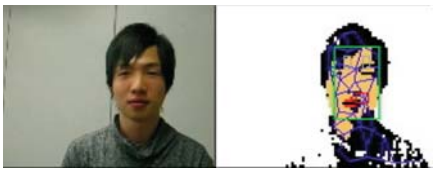
4 Experimental results

This section shows experimental results of human recognition of a partner robot. The maximal number of nodes in GNG is 50. The population size of SSGA is 50. The number of generations (iterations) SSGA in each frame is 200 including the initial evaluations after the frame of image is updated. This value is relatively small comparing with that of the search by standard GA, but the search by SSGA is a time series of continuous search in a dynamic environment including a small change. Table 1 shows the initial values of shape information used in fuzzy evaluation, and Table 2 shows the search range of face parameters used in SSGA.

Figure 9 shows experimental results of the proposed method. First of all, we conducted experiments on face recognition of two different people (Case 1 and Case 2). Figure 9.(a) and (b) show snapshots of original image (left), and human-object detection results by SSGA (right) where a



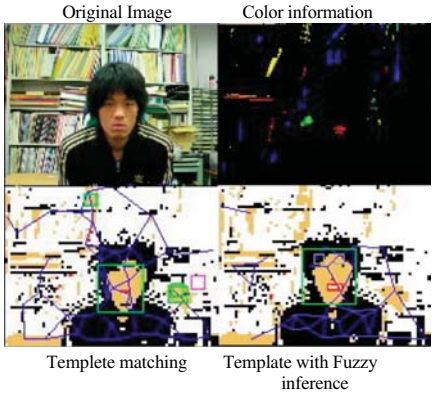
(a) Case 1



(b) Case 2



(c) Case 3



(d) Case 4

Figure 9: Experimental results of face recognition

Table 3: The values of human face obtained in experimental results of Cases 1 and 2

	$g_{i,5}^H$	$g_{i,6}^H$	$g_{i,7}^H$
case 1	0.31	0.19	0.77
case 2	0.26	0.27	0.78

ISBN: 978-989-95079-6-8

green box indicates the extracted human face. Table 3 shows the values of human face obtained in experimental results of Cases 1 and 2. In these results, the proposed method can extract human face with identifying the position of facial landmarks correctly.

Next, we conducted experiments on face recognition where the image includes skin color of object (Case 3) and the image includes the complicated background (Case 4). Figure 9 (c) shows experimental results of the proposed method in Case 3. The proposed method can extract a human face, although the person wears a helmet with having a skin-color book. Figure 9 (d) shows experimental results of the proposed method in Case 4; original image (upper-left), color information (upper-right), template matching without fuzzy evaluation (lower-left), and template matching with fuzzy evaluation (lower-right). The proposed method succeeds to extract a human face correctly. On the other hand, the template matching without fuzzy evaluation fails to extract human face, because some books located in the left of the person on the background image are recognized as a person. In this way, the proposed method can extract a human face in various environmental conditions.

5 Conclusions

This paper proposed a method of human face detection based on the template matching with a steady-state genetic algorithm and fuzzy evaluation. The experimental results show the effectiveness of the proposed method. Membership functions are very useful to evaluate candidate solutions including noisy data. The essence of the proposed method in the flexibility of the search by combining steady-state genetic algorithm and fuzzy evaluation.

As a future work, we will develop a method of human face detection in case of rotation of human face. Furthermore, we apply the proposed method to the associative learning between the perceptual information and symbolic information peculiar with the interacting person.

REFERENCES

[1] J.Cassell, Embodied Conversational Agents: Representation and Intelligence in User Interface, AI Magazine, Winter Vol.22, No.3, pp.67-83, 2001.

[2] Y.Nakauchi, R.Simmons, A Social Robot that Stands in Line, Journal of Autonomous Robots, Vol. 12, No. 3, pp.313-324, 2002.

[3] M.Imai, T.Ono, H.Ishiguro: Physical Relation and Expression: Joint Attention for Human-Robot Interaction, IEEE Transactions on Industrial Electronics, Vol.50, No.4, pp.636-643, 2003.

[4] H.Ishiguro, M.Shiomi, T.Kanda, D.Eaton, and N.Hagita, Field Experiment in a Science Museum with communication robots and a ubiquitous sensor network, Proc. of Workshop on Network Robot System at ICRA2005, 2005.

[5] J.A.Anderson, E.Rosenfeld,. *Neurocomputing*, The MIT Press, Cambridge, Massachusetts, US , 1988.

[6]. David Marr, Vision, W. H. Freeman, San Francisco, 1982.

[7] C.-H. Chen, L.-F. Pau, P.-S-P. Wang, *Handbook of Pattern Recognition & Computer Vision*, World Scientific Publication Co. Pte. Ltd, 1993.

[8] K.Fukushima Neural network model restoring partly occluded patterns, *Knowledge-Based Intelligent Information and Engineering Systems, (7th International Conference, KES 2003)*, Part II, eds: V. Palade, R. J. Howlett, L. Jain, Berlin - Heidelberg: Springer-Verlag, pp. 131-138, 2003.

[9] H.H.Bulthoff, S..W. Lee, T.A. Poggio, C. Wallraven , *Biologically Motivated Computer Vision*, Springer-Verlag, 2002.

[10] D.Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning, Massachusetts: Addison Wesley Publishing Company Inc, 1989.

[11] J.H.Holland, Adaptation in Natural and Artificial Systems , First edition, Massachusetts, The MIT Press, 1992.

[12] D.Fogel, Evolutionary Computation, New York: IEEE Press, 1995.

[13] G.Syswerda, G. A study of reproduction in generational and steady-state genetic algorithms, *Genetic Algorithms*, Morgan Kaufmann Publishers Inc., San Mateo, 1991.

[14] S.Watanabe, T.Hiroyasu, M.Miki, LCGA : Local Cultivation Genetic Algorithm for Multi-Objective Optimization Problems, Proceedings of the Genetic and Evolutionary Computation Conference (GECCO2002), pp.702, 2002.

[15] M.Lozano, F.Herrera, N.Krasnogor, D.Molina, Real-Coded Memetic Algorithms with Crossover Hill-Climbing, Evolutionary Computation, Vol. 12, No. 3, Pages 273-302, MIT Press, Cambridge, Massachusetts, US, 2004.

[16] J.Branke, Evolutionary Optimization in Dynamic Environments, Volume 3 of the Book Series on Genetic Algorithms and Evolutionary Computation, Kluwer Academic Pub Published, 2001.

[17] R. C. Eberhart, J. Kennedy, Y. Shi, Swarm Intelligence, Morgan Kaufmann Publisher, San Francisco, 2001.

[18] E. Bonabeau, M. Dorigo, G. Theraulaz, Swarm Intelligence: From Natural to Artificial Systems, Oxford University Press, New York, 1999.

[19] M. Dorigo and T. Stutzle, Ant Colony Optimization. Bradford Books, The MIT Press, 2004.

[20] Indra A.Sulistijono and N.Kubota, Human Head Tracking Based on Particle Swarm Optimization and Genetic Algorithm, Journal of Advanced Computational Intelligence and intelligent Informatics, Vol.11 No.6, Jul. 2007, pp.681-687, 2007

[21] D.Sperber, D.Wilson, Relevance - Communication and Cognition, Blackwell Publishing Ltd., 1995.

[22] B. Fritzke. A growing neural gas network learns topologies. In G. Tesauro, D. S. Touretzky, and T. K. Leen, editors, Advances in Neural Information Processing Systems 7, pages 625-632. MIT Press, Cambridge MA, 1995.

[23] M.Hynar, M.Burda, J.Sarmanova: Unsupervised clustering with growing self-organizing neural network -- a comparison with non-neural approach. DATESO pp.58-68, 2005

[24] B. Fritzke. Growing cell structures - a self-organizing network for unsupervised and supervised learning. Technical Report ICSTR-93-026, International Computer Science Institute, Berkeley, May 93.

[25] N.Kubota, K.Nishida, Cooperative Perceptual Systems for Partner Robots Based on Sensor Network, International Journal of Computer Science and Network Security, Volume 6, Number 11, pp.19-28, 2006.

[26] N.Kubota, Visual Perception and Reproduction for Imitative Learning of A Partner Robot, WSEAS Transaction on Signal Procesing, Issue 5, Vol. 2, pp. 726-731, 2006.

[27] N.Kubota, Computational Intelligence for Structured Learning of A Partner Robot Based on Imitation, *Information Science*, No. 171, pp.403-429, 2005.

[28] N.Kubota; Y.Nojima, F.Kojima, T.Fukuda,. Multiple Fuzzy State-Value Functions for Human Evaluation through Interactive Trajectory Planning of a Partner Robot, *Soft Computing,* Vol.10, No.10 pp.891-901, 2006.

[29] W. Zhao, R. Chellappa, J. Phillips, and A. Rosenfeld,"Face Recognition: A Literature Survey", ACM Computing Survey, December Issue, pp.399-458, 2003.

[30] N.Kubota, A.Yorita, Structured Learning for Partner Robots based on Natural Communication, Proc. (CD-ROM) of 2008 IEEE Conference on Soft Computing in Industrial Applications (SMCia), pp. 303-308, 2008.

[31] N.Kubota, A.Yorita, M.Satomi, Visual Tracking of A Partner Robot through Interaction with People, Proc. of Joint 4th International Conference on Soft Computing and Intelligent Systems and 9th International Symposium on advanced Intelligent Systems (SCIS & ISIS 2008), pp. 1032-1037, 2008.

[32] N.Kubota, Y.Tomioka, Evolutionary Robot Vision for Human Tracking of Partner Robots in Ambient Intelligence, Proc (CD-ROM) of Congress of Evolutionary Computation, pp. 1491-1496, 2007.

[33] N.Kubota, T.Ohinata, S.Wakisaka, H.Liu, Extraction of Facial Landmarks for a Partner Robot Based on Evolutionary Computation, Proc. of Computation Intelligence, Robotics and Autonomous Systems (CIRAS 2007), pp. 44-49, 2007.

[34] Indra A.Sulistijono, and N.Kubota, Evolutionary Robot Vision for Multiple Human Heads Tracking of A Partner Robot, Proceeding of 16th IEEE International Conference on Robot & Human Interactive Communication (Ro-Man2007), pp. 463-468, 2007.

ISBN: 978-989-95079-6-8

Combining Wavelets and Computational Intelligence Methods with applications on Multi-class Classification datasets.

Carlos C. Bracho[1]

1 Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada
Email: carloscb@ece.ualberta.ca

Abstract—*In this paper, we propose a novel algorithm for wavelet feature extraction as input to a supervised Multi-Class Classifier to improve classification performance. In particular, to select the best wavelets coefficient features, we first compute the energy-based variance distribution from wavelets coefficients at different subbands as well as the entropy-based fuzzy measures associated with the training instances. Once we get these entropy-based fuzzy measures associated with the different subsets of wavelets subbands, we apply the Möbius Transform to these entropy-based fuzzy measures to extract the Multivariate Mutual Information associated with the different subsets of wavelets subbands. The goal of these measures is twofold: assign weights (based on the wavelets information content) to all subsets of wavelets subbands and extract the independent (in terms of the multivariate mutual information) subsets of wavelets subbands. In our case, the optimal subsets of wavelets subbands as wavelets features vectors to train a Bayesian Network Model are those which provide a multivariate mutual information equal to zero. Experimental results with the multi-class SRBCT cancer dataset, show that our proposed approach achieves lower classification error in comparison with other methods proposed in the literature.*

Keywords— fuzzy measures, mutual information, wavelets.

1 Introduction

In recent years, Wavelets-based classification have gained a great relevance in different kinds of applications such as image processing [6][7], time series[8][2], bioinformatics[9][12], and pattern recognition[10][11]. The early research in wavelets-based classification was focused mainly on extracting energy values from the wavelet-subbands decomposition and using them for classification. However, in pattern recognition tasks such as classification, it is well known that proper feature selection help to improve the classification performance. Hence, a process which consists on removing irrelevant wavelets features as well as selecting a subset of wavelets from the wavelet-subbands decomposition has had a paramount importance in Wavelets-based Classification tasks.

In this paper, we propose a novel algorithm that uses entropy-based fuzzy measures and multivariate mutual information–based tests to identify the suitable subsets of wavelets features as the correct input to a supervised K2-based Bayesian Network to improve classification performance. At present time, Bayesian Networks [13[14] have became one of the most important technologies in the area of applied artificial intelligence. Roughly speaking, Bayesian Networks are graphical structures that model the probabilistic cause-effect relationships among several related variables. Particularly, the K2 Algorithm[15] is a greedy algorithm that learns the structure of Bayesian Networks from data in an efficient way given a prior order of the variables (in improper order of data will provide poor classification performance). In our case, we are proposing entropy-based fuzzy measures, its Möbius representation and multivariate mutual information to provide not only the correct order but also the correct wavelets features to a K2-based Multi-Class Bayesian Network for classifying a multi-class cancer dataset.

The remaining of this paper is organized as follows: in Section 2, we provide background on Wavelets, Bayesian Networks, Fuzzy Measures, Entropy-based Fuzzy Measures and Fuzzy Measures-based Multivariate Mutual Information. In Section 3, we detail our multi-class classification approach. In Section 4 we present the experimental results obtained using a cancer dataset with our proposed approach. Finally, in Section 5, we offer a summary and discussion of future work.

2 Background

2.1 The Wavelet Transform

The Wavelet Transform (WT)[2] is a natural context for analyzing the time-varying properties of most real-world signals. In a narrow sense, WT decomposes a signal into different time scales (wavelet- subbands decomposition) to detect features of its short-term as well as its long-term dynamics. Essentially, this transform is defined in two spaces[2]: continuous and discrete. In the continuous space, the Continuous Wavelet Transform CWT is defined by means of two basic functions (known as the mother and the father wavelet function, respectively). The father wavelet function $\varphi(t)$ depending of time t is used to capture the smooth part of the data. By contrast, the mother wavelet function $\psi(t)$ is stretched and translated to detect the rough components of the data. These two functions, to be considered as wavelet functions, must satisfy certain properties [2].

Complementary to the CWT described above, the Discrete Wavelet Transform (DWT) can be defined from an engineering point of view by means of cascading a set of couple of low and high filters [2]. These couples of filters are known in the literature as the wavelet and scaling filters. Together, this set of filters must fulfill the Quadrature Mirror Filter (QMF) property to obtain a perfect reconstruction pass-band filter.

Of particular interest in this paper, is the application of a modified version of DWT known as the Maximal Overlap

ISBN: 978-989-95079-6-8

Wavelet Transform (MODWT)[2], to provide wavelets features to a K2-based Bayesian Network. Roughly speaking, MODWT is a rescaled version of the wavelet and scaling filter used in the common DWT but with the following advantages:

- Wavelet coefficients are aligned with the original signal values.
- MODWT provides a suitable format for the traditional tabular framework used in machine learning approaches.
- MODWT provides more redundancy at each wavelet-subband.

Definition 1
Let *j*, *t* be the decomposition level (wavelet subband), and time index, respectively, and X_t a real-valued uniformly sampled signal whose size is *N*. The j^{th} level MODWT wavelet and scaling coefficient $W_{j,t}$ and $V_{j,t}$, respectively, are defined by convolving the impulse response of the wavelet and scaling filters with the signal X_t in the following way [2]:

$$W_{j,t} = \sum_l h_{j,l} * X_{t-l \bmod N} \quad (1)$$

$$V_{j,t} = \sum_l g_{j,l} * X_{t-l \bmod N} \quad (2)$$

Where *l* is the length of both the wavelet filter impulse response *h* and the scaling impulse response *g*. In this paper, MODWT was implemented by means of a Pass-Band Wavelet Filter (using the WMTSA matlab toolbox [2]). In particular, we use a Daubechies Extremal Phase Filter with *l=6* (DB6). One fundamental property of this filter is that it transforms a signal in terms of difference of several order of averages at different wavelet-subbands.

2.2.1 MODWT Energy-based variance distribution
Another important property of WT is its ability to preserve the energy of a signal. Roughly speaking, the WT decomposes the total energy of a signal at different wavelet-subbands defining an Energy Density Distribution (EDD). In particular, this EDD is computed by means of wavelets and scaling coefficients of MODWT expressed as follows[2]:

$$\sum_t X_t^2 = \sum_j \sum_{n=0}^{N_j-1} W^2{}_{j,t} + \sum_{n=0}^{j-1} V^2{}_n \quad (3)$$

A closely related concept to the wavelet-based energy decomposition is the wavelet variance which can be associated with a probability distribution and used in this paper for purposes of computing the entropy-based fuzzy measures. Mathematically, this wavelet variance is defined as follows [2]:

$$\sigma_X^2 = \frac{1}{N} \sum_{j=1}^{MaxLevel} \left\| W_j \right\|^2 + \frac{1}{N} \left\| V_{MaxLevel} \right\|^2 - \overline{X}^2 \quad (4)$$

Where:

N: Number of observations in the signal.
$MaxLevel$: Maximum level of decomposition .

W_j : Vector of wavelet coefficients defined at subband j .

$V_{MaxLevel}$: Vector of scaling coefficients defined at subband j .

$\overline{X}$: Sample mean of signal X_t.

2.2 Bayesian Networks
Bayesian Networks [16][13] have become one of the most important technologies in the area of applied artificial intelligence. They have shown to offer reliable methods for prediction, decision making, classification, and data mining in different areas such as medicine, image processing, marketing, banking, finance, etc. Roughly speaking, Bayesian Networks are graphical structures that model the probabilistic cause-effect relationships among several related variables. Mathematically, a Bayesian Network is modelled by means of a Directed Acyclic Graph (representing the above cause-effect relationships) and a set of Conditional Probabilities (one conditional probability for each node given its parent set in the graph).

Definition 2
A Bayesian Network $B = (\chi, G, P)$ consists of :

- A Direct Acyclic Graph $G = (V, E)$ with nodes $V = (V_1, V_2, .., V_n)$.
- A set of discrete random variables χ modeling the nodes of G.
- A set of conditional probability distributions $P(X_v \mid X_{parent(v)})$ for each random variable $X_v \in \chi$ where :

$$P(\chi) = {}_{v \in V} P(X_v \mid X_{pa(v)}) \quad (5)$$

The last property specifies a joint probability distribution over X known as the Bayesian Network chain rule.

ISBN: 978-989-95079-6-8

2.2.1 K2-based Bayesian Network learning

Generally speaking, the structure learning process under the K2 algorithm is made by means of an optimization process in which a quality measure Q of a Bayesian Network structure given a training dataset X is maximized; that is:

$$\max Q(B_S \mid X) \tag{6}$$

K2 is a greedy search algorithm that learns from data presenting a good performance when a prior order of the data exists. This order refers to have first in a sequence the parents and then the children. Assuming that the attributes have an order, the algorithm starts setting the parent of node x_1 to the empty set $\varnothing$. Then, the algorithm visists each subsequent node following the order in the sequence order adding $Parents_i$ to the set of parents of node attribute x_i only if the inclusion of the $Parents_i$ to the node attribute x_i maximize the network-structure posterior probability distribution. At the same time, it includes the relationship between the parent and the children in question by adding arcs between them.

2.2.2 Fuzzy Measures

Fuzzy measures are mainly known in the context of Fuzzy Integrals[3][5]. Grabisch[4], presents the usage of fuzzy measures and fuzzy integral in supervised classification and feature extraction problems. In these papers the author proposes an optimization approach which is based on a heuristic least mean squares (HLMS) algorithm, to compute the fuzzy measures as well as the Choquet Integral in classification tasks obtaining remarkable results.

Definition 3

Let the triple $K = (\Omega, A, g)$ a fuzzy measure space containing a family A of subsets defined on Ω. A set-valued function $g : A \to 0,1]$ is called a fuzzy measure if:

$$g(\varnothing) = 0 \tag{7}$$

$$g(S) = 1 \tag{8}$$

$$A_i \subset A_j \to g(A_i) \le g(A_j). \forall A_i, A_j \in S \tag{9}$$

Any set-valued function $g : A \to [0,1]$ can be uniquely expressed in terms of its MöbiusTransform represented by:

$$g(T) = \sum_{S \in T} m^g(S), \ \forall T \subseteq A, \tag{10}$$

where the set function $m^g : A \to [0,1]$ is called the Möbius Inversion Transform of g and is given by[17]:

$$m^g(S) = \sum_{T \subseteq S} (-1)^{|S|-|T|} g(T), \ \forall S \subseteq A, \tag{11}$$

2.2.2.1Entropy-based fuzzy measures

In comparison with the Grabish approach, in this paper we are interested in applying and extending a novel unsupervised approach to compute entropy-based fuzzy measures proposed by Kojadinovic[1]. The main characteristic of this approach is the replacement of the notion of attributes importance with the notion of information content in attributes by means of the entropy and the mutual information concepts.

Definition 5

Let 2^S denote the power set of $S = \{W_1, W_2, .., W_m\}$. An entropy-based fuzzy measure is a set-valued function $g_H : 2^S \to [0,1]$ defined as follows[1]:

$$g_H(S) = \begin{cases} 0 & S = \varnothing \\ (-1)^{|S|+1} H(P_{(\overleftarrow{W}_1,..)}) & S = \{\overleftarrow{W}_1,.\} \end{cases} \tag{12}$$

where :

$H(P_{(\overleftarrow{W}_1,..)})$: Entropy associated with the joint probability distribution $P_{(\overleftarrow{W}_1,..)}$). (In our case W_m means the maximum wavelet-subband, $P_{(\overleftarrow{W}_1,..)}$ refers to the joint probability distribution among the different subset of wavelet-subbands, and $H(P_{(\overleftarrow{W}_1,..)})$ to their associated entropy).

Definition 6

Any entropy-based fuzzy measure $g^H : 2^S \to [0,1]$ can be estimated by:

$$\tilde{g}^H = \frac{H(P_{(W_{i1},..,W_{is})})}{H(P_{(W_{i1},..,W_{iM})})} \tag{13}$$

Definition 7

Any entropy-based fuzzy measure $g^H : 2^S \to [0,1]$ can be uniquely expressed in terms of its Möbius Transform represented by:

$$g^H(S) = \begin{cases} 0 & S = \varnothing \\ (-1)^{|S|+1} I(\overleftarrow{W}_1;..) & S = \{\overleftarrow{W}_1;..\} \end{cases} \tag{14}$$

where :

I : *Multivariate Mutual Information*

ISBN: 978-989-95079-6-8

Proposition 1
A set of wavelets subbands is called independent if its Möbius Transform associated with its entropy-based fuzzy measure is equal to zero.

Proof. From equation 7, clearly $(-1)^{|S|+1} \neq 0$ for all S. Hence, $I(\overleftarrow{X}_1;..\overleftarrow{X}_r) = 0$ implies $g^H(S) = 0$ ■.

3 Proposed Approach

The proposed approach aims to identify the suitable wavelet information as the correct input for the the Bayesian Network. By contrast with other wavelet feature extraction approaches reported in the literature (in which most of them deal at the granularity of wavelets coefficients), this approach presents a novel approach which deals with sets of wavelets-subbands using entropy-based fuzzy measures, its Möbius Transform and multivariate mutual information. To compute the entropy-based fuzzy measures associated with each set of wavelets-subbands, we first consider the energy-based variance distribution at different wavelet-subbands (this Energy Distribution is discretized to fulfill the monotonicity property of fuzzy measure) as random vectors. This set of energy-based variance distributions are used to estimate the entropy-based fuzzy measures associated with the different sets of wavelets subbands. Once we have defined the set of entropy-based fuzzy measures, we apply the Möbius Inversion Transform to these entropybased fuzzy measures and then we select those sets of wavelets subbands whose Möbius Inversion Transform is equal to zero (i.e. set of independent wavelets-subbands in all cancer classes whose multivariate mutual information is equal to zero). This set of wavelets subbands are the wavelet-training and wavelet-testing to our supervised Bayesian Network Model (see Algorithm I and II).

Algorithm I: Training Phase

1.- Data Preprocessing withWavelet Transform

- Apply Daubechies (DB6) Extremal Phase Filter to the training dataset.

2.- Energy-based Variance Distribution Estimation and discretization

- Compute equations (3) and (4).

3.- Estimate entropy-based Fuzzy Measures

- Compute the Wavelet-subbands Join Probability Distributions with the maximum likelihood estimator[1].
- Compute the entropy associated with the above Wavelet-subbands Joint Probability Distributions.
- Compute entropy-based fuzzy measures using equation (13).

4.- Apply the Möbius Transform to the entropy-based Fuzzy Measures obtained in step 3 to obtain their associated Multivariate Mutual Information

- Compute equations (14).

5.- Wavelet Feature Vector Extraction to train Bayesian Network.

- Select sets of wavelet-subbands whose Möbius Transform is equal (or aprox.) to zero (Proposition 1).
- Concatenate wavelets coefficients of different wavelet-subbands that fulfill the above condition. These wavelets features are the input to train a Bayesian Network. (If exists several sets of wavelet-subbands that fulfill the above condition, you might select the subset with less wavelet coefficients)

End Algorithm I

Algorithm II: Testing Phase

1.- Data Preprocessing withWavelet Transform

- Apply Daubechies (DB6) Extremal Phase Filter to the testing dataset.

2.- Select the same set of wavelet-subbands obtained in step 5 of algorithm I.

- Apply 10-fold cross validation to verify performance of the trained Bayesian Network with the selected wavelet coefficients.

End Algorithm II

4 Experimental results

Given the great interest in the scientific community in exploring DNA microarray which in general consist of thousands of genes, to evaluate the performance of our proposed approach, we use a highly and noisy multidimensional multi-class microarray dataset. In particular, we use a cancer microarray dataset known as SRBCT[9] consisting of 2,308 genes and 63 samples from four types of cancers: neuroblastoma (NB), rhabdomyosacoma (RMS), Burkitt lymphomas (BL) and Ewing family of tumors(EWS). In our case, we consider a total of 32 instances for the training phase partitioned uniformly with respect to each kind of cancer (8 samples for each type of cancer) and 31 instances as the testing dataset (eight samples from each type of cancer except one with seven instances as testing corresponding to the Burkitt lymphomas).

ISBN: 978-989-95079-6-8

In the training phase after applying the first four steps, we found that the wavelet-subband 5 (containing 2308 wavelet coefficients) provide us a multivariate mutual information equal to zero fulfilling the condition established in proposition 1. Once we apply step 5 of algorithm 1, we use a 10-fold cross-validation method and consider an input SRBCT-training dataset to the Multi- Class Bayesian Network (which is implemented. in the data mining environment known as weka [19]) in the following way:

$$T = \left\{ \left(\overleftarrow{W}_i, y_i \right) : \overleftarrow{W}_i \in \Re^{2308} \right\} \tag{14}$$

where:

$$\overleftarrow{W}_i \in \Re^{2308} \tag{15}$$

$$y_i \in \{NB, RMS, BL, EWS\} \tag{16}$$

Once trained, we test our Bayesian Network classifier obtaining the following results reported in table I (in comparison with other results reported in the literature, in this table we can see that there are two methods providing 100% of accuracy partitioning the multi-class problem as a collection of N binary sub-problems. In our case, we solve the multi-class problem without using this artefact):

Table I. Classification Performance

Method	Accuracy %
Our Method	**92%**
KNN	30%
Naive Bayes	60%
lvs1 SVM/Wavelet[9]	100%
SVM-OVA	80%
Decision Tree	75%

5 Conclusions

We have reported experimental results using a novel wavelet-computational intelligence approach for improving classification of multi-class datasets. In particular, we propose a novel approach for extracting entropy-based fuzzy measures from the wavelet-subband decomposition as well as their multivariate mutual information (by means of the Möbius Inversion Transform). With this information, we select the most suitable wavelets (based on the concept of independence given in terms of multivariate mutual information) to train and evaluate a multiclass Bayesian Network to classify cancer types obtaining satisfactory results. In conclusion, the propose approach provides a general framework to deal with data which are inherently noisy and high dimensional such as microarray gene expression data. Future work in this research will proceed in refining and validating the proposed approach in other applications.

Acknowledgment

This work was partially supported by CONACYT Mexico. The author also likes to thank all the reviewers of this paper for their valuable comments.

References

[1] I. Kojadinovic, Unsupervised aggregation of commensurate correlated attributes by means of the Choquet Integral and Entropy Functionals. *International Journal of Intelligent Systems*. Vol 23, pp. 128-154. 2008.

[2] Donald B. Percival, A. T.Walden,*Wavelet Method for Time Series Analysis*. Cambridge Series in Statistical and Probabilistics Mathematics. 2000.

[3] G. Choquet, Theory of Capacities. *Annales de l'Institut Fourier*, Vol 5, pp. 131-295. 1953.

[4] M. Grabisch*, Fuzzy integral for classification and feature extraction.* Fuzzy measures and integrals Theory and applications. Physica Verlag, pp. 135-148. 2000.

[5] G. Beliakov, *Aggregation Functions: A guide for Practitioners*. Springer Studies in Fuzziness and Soft Computing. 2007.

[6] K. Huang and S. Aviyente. Information-theoretic wavelet packet subband selection for texture classification.*Signal Processing* 86, 1410-1420, 2006.

[7] M. Shoshany.Wavelet decomposition for reducing flux density effects on hyperspectral classification. *IEEE Geoscience and Remote Sensing Letters*, v6, No. 1, 38-41, January 2009.

[8] T. Qian. *Wavelet Analysis and Applications*. Birkhauser, First Edition, January 2007.

[9] S. Prabakaran, R. Sahu, S. Verma. Classification of multi class dataset using wavelet power spectrum. *Data Mining Knowledge Discovery*, v15, 297-319, 2007.

[10] D. Li,W. Pedrycz, N. Pizzi. FuzzyWavelet packet based feature extraction method and its application to biomedical signal classification. *IEEE Transactions on Biomedical Eng*, v52-6, 1132- 1139, June 2005.

[11] T. Celik, and T. Tjahjadi. Multiscale texture classification using dual-tree complex wavelet transform. *Pattern Recognition Letters*, v30-3, 331-339, February 2009.

[12] C. Aggarwal, and P. Bradley. On the use of Wavelet Decomposition for String Decomposition. *Data Mining Knowledge Discovery*, v10-2, 117-139, 2005.

[13] U. Kjaerulff, and A. Madsen. *Bayesian Networks and Influence Diagrams*. Information Science and Statistics, Springer, 2008.

[14] X. Chen, G. Anantha, X. Lin. Improving Bayesian Network Structure Learning with Mutual Information-Based Node Ordering in the K2 Algorithm. *IEEE Transactions on Knowledge and Data Eng*, v20-5, 1-13, May, 2008.

[15] G. Cooper and E. Herskovits. A Bayesian Method for the induction of Probabilistic Networks from Data. *Machine Learning*, v9, 309-347, 1992.

[16] R. Neapolitan and X. Jian. *Probabilistics methods for financial and marketing informatics*. Morgan and Kauffman publishers, 2007.

[17] H. Nguyen and E. Walker. *A first course in Fuzzy logic*. Chapman & Hall/CRC , 2006.

[18] T. Cover and J. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2006.

[19] I. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. San Francisco, Morgan Kauffman, 2005.

ISBN: 978-989-95079-6-8

Poincaré recurrence theorem in MV-algebras

Beloslav Riečan
Faculty of Natural Sciences, Matej Bel University
Tajovského 40, 974 01 Banská Bystrica, Slovakia
Mathematical Institute of Slovak Acad. of Sciences
Štefánikova 49, 814 73 Bratislava, Slovakia
Email: riecan@fpv.umb.sk

Abstract— *The classical Poincaré weak recurrence theorem states that for any probability space $(\Omega, \mathcal{S}, P)$, any P-measure preserving transformation T, and any $A \in \mathcal{S}$, almost all points of A return to A. In the present paper the Poincaré theorem is proved when the σ-algebra $\mathcal{S}$ is replaced by any σ-complete MV-algebra.*

Keywords— Measure preserving transformation, Poincaré recurrence theorem, σ-complete MV-algebra.

1 Introduction

Let $(\Omega, \mathcal{S}, P)$ be a probability space, i.e. $\Omega \neq \emptyset$, $\mathcal{S}$ is a σ-algebra of subsets of Ω (i.e. $\Omega \in \mathcal{S}; A \in \mathcal{S} \Longrightarrow \Omega \setminus A \in \mathcal{S}$, and $A_n \in \mathcal{S}$ $(n = 1, 2, ...) \Longrightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{S}$), and $P : \mathcal{S} \to [0, 1]$ is such that $P(\Omega) = 1$, and $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$, whenever $A_n \in \mathcal{S}$ $(n = 1, 2, ...)$ and $A_n \cap A_m = \emptyset$ $(n \neq m)$. Let $T : \Omega \to \Omega$ be such that $A \in \mathcal{S}$ implies $T^{-1}(A) \in \mathcal{S}$, and

$$P(T^{-1}(A)) = P(A), \tag{1}$$

whenever $A \in \mathcal{S}$; such transformations T are called measure-preserving. The Poincaré recurrence theorem states that almost every point $x \in A$ will return to A, i.e.

$$P(A \setminus \bigcup_{n=1}^{\infty} T^{-n}(A)) = 0, \tag{2}$$

whenever $A \in \mathcal{S}$. There is also the stronger variant of the Poincaré theorem: almost all points of A will return to A infinitely many times. It means that for any $x \in A \setminus B$ (where $P(B) = 0$) and any $k \in N$ there exists $n \geq k$ such that $T^n(x) \in A$:

$$P(A \setminus \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} T^{-1}(A)) = 0. \tag{3}$$

The strong recurrence theorem has been proved in various connections (see [7] for Boolean algebras or [3] for topological spaces). In [8] the theorem has been proved for a special class of MV-algebras (σ-complete weakly σ-distributive MV algebras with product). In the present paper we prove the weak version for arbitrary MV-algebras.

2 MV-algebras

The notion of an MV-algebra has been introduced by Chang [1] as an algebraic system

$$(M, \oplus, \odot, \neg, u, 0)$$

where $\oplus, \odot$ are binary operations, $\neg$ is a unary oparation and $0, u$ are fixed elements. By the Mundici theorem ([4]) every MV-algebra corresponds to a unique unital l-group G (lattice ordered group with a distinguished order unit u),

$$\begin{aligned} M = [0, u] &= \{x \in G; 0 \leq x \leq u\}, \\ a \oplus b &= (a + b) \wedge u, \\ a \odot b &= (a + b - u) \vee 0, \\ \neg a &= u - a. \end{aligned} \tag{4}$$

Analogously as in quantum structures (see [2]), instead of a probability masure, a state $m : M \to [0, 1]$ can be defined.

2.1. Definition. A *state* on an MV-algebra $(M, \oplus, \odot, \neg, u, 0)$ is a mapping $m : M \to [0, 1]$ satisfying the following conditions:

(i) $m(u) = 1$;

(ii) m is additive,
i.e. $a \odot b = 0 \Longrightarrow m(a \oplus b) = m(a) + m(b)$;

(iii) m is continuous, i.e. $a_n \nearrow a \Longrightarrow m(a_n) \nearrow m(a)$.

2.2. Proposition. Any state $m : M \to [0, 1]$ is strongly additive, i.e. it satisfies the following implication $a_1 + a_2 + ... + a_n \leq u \Longrightarrow m(a_1 + a_2 + ... + a_n) = \sum_{i=1}^{n} m(a_i)$.

Proof. It can be proved by induction.

The following proposition is evident.

2.3. Proposition. A mapping $m : M \to [0, 1]$ is a state if and only if the following conditions are satisfied:

(i) $m(u) = 1$;

(ii) if (a_n) is a sequence of elements of M such that $a_1 + a_2 + ... + a_n \leq u$ for any $n \in N$, then

$$m(\bigvee_{n=1}^{\infty} \sum_{i=1}^{n} a_i) = \sum_{i=1}^{\infty} m(a_i). \tag{5}$$

3 Main result

3.1. Definition. Let M be a σ-complete MV-algebra, and $m : M \to [0, 1]$ be a state. By an *m-preserving transformation* of M we understand a mapping $\tau : M \to M$ satisfying the following conditions:

(i) $\tau(u) = u, \tau(0) = 0$;

ISBN: 978-989-95079-6-81

(ii) $\tau(a \odot b) = \tau(a) \odot \tau(b)$;

(iii) $a \leq b \Longrightarrow \tau(b - a) = \tau(b) - \tau(a)$;

(iv) $\tau(a \wedge b) = \tau(a) \wedge \tau(b)$;

(v) $\tau(\bigvee_{n=1}^{\infty} a_n) = \bigvee_{n=1}^{\infty} \tau(a_n)$;

(vi) $m(\tau(a)) = m(a)$.

3.2. Definition. $a \setminus b = a \odot (\neg b)$.

3.3. Theorem. Let M be a σ-complete MV-algebra, $m : M \to [0, 1]$ be a state. Then for any $a \in M$

$$m(a \setminus \bigvee_{i=1}^{\infty} \tau^i(a)) = 0. \tag{6}$$

Proof. Put $b_j = \bigvee_{i=j}^{\infty} \tau^i(a), b = a \setminus \bigvee_{i=1}^{\infty} \tau^i(a) = (a - \bigvee_{i=1}^{\infty} \tau^i(a)) \vee 0 = (a - b_1) \vee 0$.

By induction it can be proved that

$$b + \tau(b) + ... + \tau^n(b) \leq \bigvee_{j=0}^{n+1} \bigvee_{i=j+1}^{n+1} (\tau^j(a) - b_i) \vee 0.$$

Since

$$(\tau^i(a) - b_i) \vee 0 \leq \tau^i(a) \leq u$$

we obtain

$$b + \tau(b) + ... + \tau^n(b) \leq u.$$

By the strong additivity property

$$m(\bigvee_{i=0}^{\infty} \tau^i(b)) = \sum_{i=0}^{\infty} m(\tau^i(b)) = \sum_{i=0}^{\infty} m(b),$$

whence

$$m(b) = 0.$$

3.4. Example.

Consider a probability space $(\Omega, \mathcal{S}, \mu)$ and the MV-algebra $\mathcal{M} = \{f : \Omega \to [0, 1]; f \text{ is } \mathcal{S} \text{ - measurable } \}$,

where

$$f \oplus g = \min(f + g, 1),$$
$$f \odot g = \max(f + g - 1, 0),$$
$$\neg f = 1 - f,$$

$u = 1_\Omega, 0 = 0_\Omega$. Moreover, define $m : \mathcal{M} \to [0, 1]$ by the formula

$$m(f) = \int_\Omega f d\mu,$$

and

$$\tau(f) = f \circ T$$

where T is measure preserving map. By Theorem 3.3 we obtain that

$$\int_\Omega (f - \bigvee_{i=1}^{\infty} f \circ T^i) \vee 0) d\mu = 0,$$

hence μ-almost everywhere

$$(f - \bigvee_{i=1}^{\infty} f \circ T^i) \vee 0 = 0,$$

or equivalently

$$f \leq \bigvee_{i=1}^{\infty} f \circ T^i$$

μ-almost everywhere.

Acknowledgment

The paper was supported by grant VEGA 1/2002/05 and grant APVV LPP-0046-06.

References

[1] C.C. Chang. Algebraic analysis of many valued logics. *Trans. Amer. Math. Soc.* 88: 467 - 490, 1958.

[2] A. Dvurečenskij and S. Pulmannová. *New Trends in Quantum Structures*. Kluwer, Dordrecht, 2000.

[3] P. Maličký P. Category version of the Poincaré recurrence theorem. *Topology and its Applications* 154:2709 - 2713, 2007.

[4] D. Mundici. Interpretation of AFC^*-algebras in Łukasiewicz sentential calculus. *J. Funct. Anal.* 65: 15 - 63, 1986.

[5] M.G. Nadkarni. *Basic Ergodic Theory*. Birkhäuser Verlag, Basel, 1998.

[6] H. Poincaré. *Les méthodes nouvelles de la mécanique classique céléste*, vol. 3. Gauthiers - Villars, Paris, 1899.

[7] B. Riečan. A note on the Poincaré recurrence theorem on Boolean rings. (In Russian.) *Mat. - fyz. čas.* 15: 13 - 22, 1965.

[8] B. Riečan. Poincaré recurrence theorem in MV-algebras. *Math. Slovaca*, submitted.

ISBN: 978-989-95079-6-82

Notes on Q-probabilities on intuitionistic fuzzy events

Magdaléna Renčová
Faculty of Natural Sciences, Matej Bel University
Tajovského 40, 074 01 Banská Bystrica, Slovakia
E-mail: rencova@fpv.umb.sk

Abstract— *Following [9] some properties of Q-probability and Q-states are studied. Representation theorem of Q-probabilities and Q-states, the existence of the joint observable and The central limit theorem are proved.*

Keywords— Q-probability, Q-state, representation theorem, intuitionistic fuzzy events, joint Q-observable, Central limit theorem.

1 Introduction

Although there are different opinions about intuitionistic fuzzy events, the following definitions are accepted generally ([1], [5]). Let $(\Omega, \mathcal{S})$ be a measurable space. By an intuitionistic fuzzy event ([5]) we mean any pair

$$\mathbf{A} = (\mu_A, \nu_A)$$

of $\mathcal{S}-$measurable functions, such that $\mu_A, \nu_A : \Omega \to [0,1]$ and $\mu_A + \nu_A \leq 1$.

The function μ_A is the membership function and the function ν_A is the non-membership function. The family $\mathcal{F}$ of all intuitionistic fuzzy events is ordered in the following way:

$$\mathbf{A} \leq \mathbf{B} \Leftrightarrow \mu_A \leq \mu_B, \nu_A \geq \nu_B.$$

Evidently

$$\mathbf{A} \wedge \mathbf{B} = (\mu_A \wedge \mu_B, \nu_A \vee \nu_B),$$
$$\mathbf{A} \vee \mathbf{B} = (\mu_A \vee \mu_B, \nu_A \wedge \nu_B).$$

It is easy to see that $\mathbf{A_n} \nearrow \mathbf{A}$ if and only if $\mu_{A_n} \nearrow \mu_A$ and $\nu_{A_n} \searrow \nu_A$.

The notion of intuitionistic fuzzy event is a natural generalization of the notion of a fuzzy event. Given a fuzzy event μ_A, the pair $(\mu_A, 1-\mu_A)$ is an intuitionistic fuzzy event, so intuitionistic fuzzy events can be seen as generalizations of fuzzy events. Hence we want to define probability on intuitionistic fuzzy events generalizing probability on fuzzy events. And actually, two constructions were proposed independently by Gregorzewski [5] and Gerstenkorn [4], both based on the Łukasiewicz operations

$$a \oplus b = \min(a+b, 1),$$

$$a \odot b = \max(a+b-1, 0).$$

Operations $\oplus, \odot$ on $[0,1]^2$ (not necessarily Łukasiewicz operations) can be naturally extended to intuitionistic fuzzy events in the following way

$$\mathbf{A} \oplus \mathbf{B} = (\mu_A \oplus \mu_B, \nu_A \odot \nu_B),$$

$$\mathbf{A} \odot \mathbf{B} = (\mu_A \odot \mu_B, \nu_A \oplus \nu_B),$$

where $\mathbf{A} = (\mu_A, \nu_A)$ and $\mathbf{B} = (\mu_B, \nu_B)$.

If $\mu : \Omega \to [0,1]$ is a fuzzy set, then $(\mu, 1-\mu)$ is an IF set corresponding to this fuzzy set. Similarly as in the classical case, in the fuzzy case and in the quantum case, a probability (or state) has been introduced as a mapping $m : \mathcal{F} \to [0,1]$ being continuous, additive and satisfying some boundary conditions. Here the main difference is the additivity which is now of the following form

$$m(\mathbf{A}) + m(\mathbf{B}) = m(\mathbf{A} \oplus \mathbf{B}) + m(\mathbf{A} \odot \mathbf{B}).$$

There exists a general representation theorem for IF-probability. If $(\Omega, \mathcal{S}, P)$ is a probability space, then to any Łukasiewicz state $m : \mathcal{F} \to [0,1]$ there exists $\alpha \in [0,1]$ such that

$$m(\mathbf{A}) = (1-\alpha) \int_\Omega \mu_A dP + \alpha(1 - \int_\Omega \nu_A dP)$$

for any $\mathbf{A} \in \mathcal{F}$ (see [2]). Of course, the constructions (see [4] [5]) can be obtained as a very special case.

Generally, there are infinitely many possibilities how to define additivity

$$m(\mathbf{A}) + m(\mathbf{B}) = m(\mathcal{S}(\mathbf{A}, \mathbf{B})) + m(\mathcal{T}(\mathbf{A}, \mathbf{B}))$$

where

$$\mathcal{S}(\mathbf{A}, \mathbf{B}) = (S(\mu_A, \mu_B), T(\nu_A, \nu_B)),$$
$$\mathcal{T}(\mathbf{A}, \mathbf{B}) = (T(\mu_A, \mu_B), S(\nu_A, \nu_B))$$
$$S, T : [0,1]^2 \to [0,1]$$

being such binary operations (T is a t-norm and S is dual t-conorm [6]), that

$$S(u,v) + T(1-u, 1-v) \leq 1.$$

The Kolmogorov probability theory has 3 fundamental notions: probability, random variable and expectation. In our fuzzy case, an analogous situation occurs.

Throughout this paper we consider the following operations with intuitionistic fuzzy events

$$A \oplus_Q B = ((\mu_A{}^n + \mu_B{}^n)^{\frac{1}{n}} \wedge 1; 1 - ((1-\nu_A)^n + (1-\nu_B)^n)^{\frac{1}{n}} \wedge 1),$$

$$A \odot B = (\mu_A, \nu_A) \odot (\mu_B, \nu_B) = ((\mu_A + \mu_B - 1) \vee 0; (\nu_A + \nu_B) \wedge 1).$$

Remark 1.1 *The operation $\oplus_Q$ was introduced by Yager [6], the operation $\odot$ is Łukasiewicz operation. This is a special case of operations studied in [9], where $\varphi(u) = u^n$, $n \in N$ is fixed for each $u \in [0,1]$. Special case $n = 2$ is studied in [2].*

We are not able to embed the family $\mathcal{F}$ with these operations into an MV-algebra. Of course, we are able to prove probability representation theorems, to construct the joint observable and prove such fundamental theorems as central limit theorem or laws of large numbers.

ISBN: 978-989-95079-6-83

2 Q-probability and Q-observables

Definition 2.1 *Let $\mathcal{F}$ be the family of all intuitionistic fuzzy events, $\mathcal{J}$ be the family of all compact subintervals of the unit interval $[0,1]$. Q-probability is any mapping $\mathcal{P}: \mathcal{F} \to \mathcal{J}$ satisfying the following conditions:*

(i) $\mathcal{P}((\mathbf{1},\mathbf{0})) = [1,1], \mathcal{P}((\mathbf{0},\mathbf{1})) = [0,0]$;

(ii) $\mathbf{A} \odot \mathbf{B} = (0,1) \Rightarrow \mathcal{P}(\mathbf{A} \oplus_Q \mathbf{B}) = \mathcal{P}(\mathbf{A}) + \mathcal{P}(\mathbf{B})$;

(iii) $\mathbf{A_n} \nearrow \mathbf{A} \Rightarrow \mathcal{P}(\mathbf{A_n}) \nearrow \mathcal{P}(\mathbf{A})$.
(Here $[a_n, b_n] \nearrow [a,b]$, if $a_n \nearrow a, b_n \nearrow b$.)

Remark 2.2 *If A is a crisp set, then $\mu_A = \xi_A, \nu_A = \xi_B$, Theorem 2.7 implies that*

$$\mathcal{P}^{sharp}(A) = (1-\alpha)p(A) + \alpha r(A)$$

$$\mathcal{P}^{sharp}(A') = (1-\alpha)p(A') + \alpha r(A')$$

hence $\mathcal{P}^{sharp}(A) + \mathcal{P}^{sharp}(A') = (1-\alpha)p(\Omega) + \alpha r(\Omega) = 1$. It follows that there is a large class of examples extending the classical definition. It makes possible to construct different models describing some real processes.

Definition 2.3 *A mapping $m: \mathcal{F} \to [0,1]$ is called a Q-state, if the following conditions are satisfied:*

(i) m((1,0))=1, m((0,1))=0;

(ii) $\mathbf{A} \odot \mathbf{B} = (0,1) \Rightarrow m(\mathbf{A} \oplus_Q \mathbf{B}) = m(\mathbf{A}) + m(\mathbf{B})$;

(iii) $\mathbf{A_n} \nearrow \mathbf{A} \Rightarrow m(\mathbf{A_n}) \nearrow m(\mathbf{A})$.

Example 2.4 *Let $(\Omega, \mathcal{S}, p)$ be a probability space, then a natural example of Q-state is a function $m: \mathcal{F} \to [0,1]$ defined by the following*

$$m((\mu_A, \nu_A)) = \int_\Omega {\mu_A}^n dp,$$

where $n \in N$ is fixed natural number.

Let us suppose, that $\mathcal{P}$ maps $\mathcal{F}$ to $\mathcal{J}$. We will present this mapping with functions $\mathcal{P}^\flat, \mathcal{P}^\sharp: \mathcal{F} \to [0,1]$ in the following manner $\mathcal{P}(\mathbf{A}) = [\mathcal{P}^\flat(\mathbf{A}), \mathcal{P}^\sharp(\mathbf{A})], \mathbf{A} \in \mathcal{F}$. Shorter notation is used further on is $\mathcal{P} = [\mathcal{P}^\flat, \mathcal{P}^\sharp]$.

Theorem 2.5 *$\mathcal{P}: \mathcal{F} \to \mathcal{J}$, is a Q-probability if and only if $\mathcal{P}^\flat, \mathcal{P}^\sharp: \mathcal{F} \to [0,1]$ are Q-states.*

Proof Let us suppose that $\mathcal{P}$ is an Q- probability, then since

$$[1,1] = \mathcal{P}((\mathbf{1},\mathbf{0})) = [\mathcal{P}^\flat((\mathbf{1},\mathbf{0})), \mathcal{P}^\sharp((\mathbf{1},\mathbf{0}))],$$

we have $1 = \mathcal{P}^\flat((\mathbf{1},\mathbf{0}))$ and $1 = \mathcal{P}^\sharp((\mathbf{1},\mathbf{0}))$.
Further let $\mathbf{A} \odot \mathbf{B} = (\mathbf{0},\mathbf{1})$. Then

$$[\mathcal{P}^\flat(\mathbf{A}) + \mathcal{P}^\flat(\mathbf{B}), \mathcal{P}^\sharp(\mathbf{A}) + \mathcal{P}^\sharp(\mathbf{B})] = [\mathcal{P}^\flat(\mathbf{A}), \mathcal{P}^\sharp(\mathbf{A})] + [\mathcal{P}^\flat(\mathbf{B}), \mathcal{P}^\sharp(\mathbf{B})] = \mathcal{P}(\mathbf{A}) + \mathcal{P}(\mathbf{B}) = \mathcal{P}(\mathbf{A} \oplus_Q \mathbf{B}) = [\mathcal{P}^\flat(\mathbf{A} \oplus_Q \mathbf{B}), \mathcal{P}^\sharp(\mathbf{A} \oplus_Q \mathbf{B})],$$

hence

$$\mathcal{P}^\flat(\mathbf{A}) + \mathcal{P}^\flat(\mathbf{B}) = \mathcal{P}^\flat(\mathbf{A} \oplus_Q \mathbf{B})$$

and

$$\mathcal{P}^\sharp(\mathbf{A}) + \mathcal{P}^\sharp(\mathbf{B}) = \mathcal{P}^\sharp(\mathbf{A} \oplus_Q \mathbf{B}).$$

Finally

$$\mathbf{A_n} \nearrow \mathbf{A} \text{ implies } [\mathcal{P}^\flat(\mathbf{A_n}), \mathcal{P}^\sharp(\mathbf{A_n})] = \mathcal{P}(\mathbf{A_n}) \nearrow \mathcal{P}(\mathbf{A}),$$

hence

$$\mathcal{P}^\flat(\mathbf{A_n}) \nearrow \mathcal{P}^\flat(\mathbf{A}) \text{ and } \mathcal{P}^\sharp(\mathbf{A_n}) \nearrow \mathcal{P}^\sharp(\mathbf{A}).$$

The opposite implication can be proved similarly. □

Let us find the representation theorems for Q-states and Q-probabilities. We are able to find this representation only for representable Q-probabilities.

Definition 2.6 *Q-probability $\mathcal{P}_0 = [\mathcal{P}_0^\flat, \mathcal{P}_0^\sharp]: \mathcal{F} \to \mathcal{J}$ is representable, if there exist functions $f, g: R^2 \to R$ and probabilities $p, r: \mathcal{S} \to [0,1]$, such that*

$$\mathcal{P}_0((\mu_A, \nu_A)) =$$

$$= [f(\int_\Omega (\mu_A)^n dp, \int_\Omega (1-\nu_A)^n dr), g(\int_\Omega (\mu_A)^n dp, \int_\Omega (1-\nu_A)^n dr)].$$

Definition 2.7 *Q-state $m_0: \mathcal{F} \to [0,1]$ is representable, if there exist a function $f: R^2 \to R$ and the probabilities $p, r: \mathcal{S} \to [0,1]$ such that*

$$m_0((\mu_A, \nu_A)) = f\left(\int_\Omega (\mu_A)^n dp, \int_\Omega (1-\nu_A)^n dr\right).$$

Theorem 2.8 Representation Theorem
Let $m_0: \mathcal{F} \to [0,1]$ be a representable Q-state. Then there exist $\alpha \in [0,1]$ and probabilities $p, r: \mathcal{F} \to [0,1]$ such that for each $A \in \mathcal{F}$

$$m_0((\mu_A, \nu_A)) = (1-\alpha) \int_\Omega {\mu_A}^n dp + \alpha(\int_\Omega (1-\nu_A)^n dr),$$

where $n \in N$ is fixed.

Firstly, let us prove the following lemma.

Lemma 2.9 *Let $f: [0,1]^2 \to R$ be an additive and continuous function, then f is linear.*

Proof Let $f: [0,1]^2 \to R$, we show, that for each $A \in [0,1]^2$ and each $\alpha \in R$ such that

$$f(\alpha A) = \alpha f(A).$$

Consider cases:

(I) $\alpha \in N$, then $f(\alpha A) = f(\underbrace{A + ... + A}_{\alpha}) = \alpha f(A)$.

(II) $\alpha \in Q^+$, so $\exists p, q \in Z, (p,q) = 1, \frac{p}{q} > 0$.
We have $f(A) = \underbrace{f\left(\frac{1}{q}A\right) + ... + f\left(\frac{1}{q}A\right)}_{q}$,
where $A \in [0,1]^2$, so $\frac{1}{q}A \in [0,1]^2$.
Then $f(A) = qf\left(\frac{1}{q}A\right)$ and so $f\left(\frac{1}{q}A\right) = \frac{1}{q}f(A)$.
Let us take $A \in [0,1]^2$ such that $\frac{p}{q}A \in [0,1]^2$,
then $\frac{1}{q}A \in [0,1]^2$ and

ISBN: 978-989-95079-6-84

$$f(\tfrac{p}{q}A) = \underbrace{f\left(\frac{1}{q}A\right) + ... + f\left(\frac{1}{q}A\right)}_{p} =$$
$$= pf\left(\tfrac{1}{q}A\right) = \tfrac{p}{q}f(A).$$

(III) Since f is continuous and $f(rA) = rf(A)$ for $r \in Q^+$, such that $A, rA \in [0,1]^2$, by approximating any real $x \in R$ by rational $(x_n)_{n=1}^{\infty}$ we get

$$f(xA) = f(\lim_{n\to\infty} x_n A) = \lim_{n\to\infty} f(x_n A) =$$
$$= \lim_{n\to\infty} x_n f(A) = xf(A).$$

Thus we proved that f is linear on [0,1]. □

Proof of Theorem 2.7
Let m_0 be a representable Q-state, we are looking for a formula for function f from Definition 2.6. From property (i) of Q-state (Definition 2.2) we get

$$m_0((\mathbf{1},\mathbf{0})) = f(\int_\Omega 1^n dp, \int_\Omega (1-0)^n dr) = f(1,1),$$

because of that

$$\int_\Omega 1 dp = \int_\Omega 1 dr = 1$$

we get

$$m_0((\mathbf{1},\mathbf{0})) = f(1,1) = 1.$$

Analogously $m_0((\mathbf{0},\mathbf{1})) = f(0,0) = 0$.
What about additivity? Assume that $A \odot B = (\mathbf{0},\mathbf{1})$, so

$$\mu_A + \mu_B \leq 1, \nu_A + \nu_B \geq 1.$$

We get

$$m_0(A \oplus_Q B) =$$
$$= f(\int_\Omega (\sqrt[n]{((\mu_A)^n + (\mu_B)^n)})^n dp,$$
$$\int_\Omega (1 - (1 - \sqrt[n]{((1-\nu_A)^n + (1-\nu_B)^n)}))^n dr) =$$
$$= f(\int_\Omega (\mu_A)^n + (\mu_B)^n dp, \int_\Omega (1-\nu_A)^n + (1-\nu_B)^n dr).$$

Analogously

$$m_0(A) + m_0(B) =$$
$$f(\int_\Omega (\mu_A)^n dp, \int_\Omega (1-\nu_A)^n dr) + f(\int_\Omega (\mu_B)^n dp, \int_\Omega (1-\nu_B)^n dr).$$

Let us denote by

$$\int_\Omega \mu_A dp = u_1, \int_\Omega \mu_B dp = v_1,$$

$$\int_\Omega (1-\nu_A) dr = u_2, \int_\Omega (1-\nu_B) dr = v_2,$$

we get

$$m_0(A \oplus_Q B) = f(u_1 + v_1, u_2 + v_2),$$
$$m_0(A) + m_0(B) = f(u_1, u_2) + f(v_1, v_2).$$

Since the property (ii) holds, we have

$$f(u_1 + v_1, u_2 + v_2) = f(u_1, u_2) + f(v_1, v_2),$$

and that is why the function f is linear ($f(x+y) = f(x) + f(y)$ holds). This equality holds by the previous Lemma. Finaly f is continuous by the (iii) property of Definition 2.2. □

Theorem 2.10 Representation Theorem of Q-probabilities *If $\mathcal{P}_0$ is a representable Q-probability, then there exist real numbers $\alpha, \beta \in [0,1]$ and probability mesures $p, r_1, r_2 : \mathcal{S} \to [0,1]$ such that $\alpha r_1 \leq \beta r_2$ that for each $A = (\mu_A, \nu_A) \in \mathcal{F}$ there holds*

$$\mathcal{P}_0((\mu_A,\nu_A)) = [(1-\alpha)\int_\Omega (\mu_A)^n dp + \alpha \int_\Omega (1 -$$
$$\nu_A)^n dr_1, (1-\beta)\int_\Omega (\mu_A)^n dp + \beta \int_\Omega (1-\nu_A)^n dr_2].$$

Proof Let $\mathcal{P}_0(A) = [\mathcal{P}_0^\flat(A), \mathcal{P}_0^\sharp(A)]$ be representable Q-probability. Following Q-states $\mathcal{P}_0^\flat, \mathcal{P}_0^\sharp$ could be written by previous formulas. □

3 Q-observables and p-joint Q-observables

First, let us denote Borelian sets by $\mathcal{B}(R)$.

Definition 3.1 *A mapping $x : \mathcal{B}(R) \to \mathcal{F}$ is called a Q-observable, if the following conditions are satisfied:*

(i) $x(R) = (1,0), x(\emptyset) = (0,1)$;

(ii) *if* $A \cap B = \emptyset$ *then* $x(A) \odot x(B) = (\mathbf{0},\mathbf{1})$, *and* $x(A \cup B) = x(A) \oplus_Q x(B)$;

(iii) $A_n \nearrow A \Rightarrow x(A_n) \nearrow x(A)$.

Theorem 3.2 *Let $x : \mathcal{B}(R) \to \mathcal{F}$ be an Q-observable, $\mathcal{P} = [\mathcal{P}^\flat, \mathcal{P}^\sharp] : \mathcal{F} \to \mathcal{J}$ be an Q-probability. Then the functions $\mathcal{P}^\flat \circ x : \mathcal{B}(R) \to [0,1], \mathcal{P}^\sharp \circ x : \mathcal{B}(R) \to [0,1]$, are probability measures.*

Proof The proof is straightforward. □

Theorem 3.3 *Let $x : \mathcal{B}(R) \to \mathcal{F}$ be an Q-observable, $x(A) = (x^\flat(A), 1 - x^\sharp(A)); \omega \in \Omega$. Then the functions $p_\omega^\flat, p_\omega^\sharp : \mathcal{B}(R) \to [0,1]$ defined by*

$$p_\omega^\flat(A) = (x^\flat(A)(\omega))^n;$$

$$p_\omega^\sharp(A) = (x^\sharp(A)(\omega))^n$$

are probability measures.

Proof Use instead of $\varphi(u) = u^n$ in Theorem 2.7 in [9]. □

ISBN: 978-989-95079-6-85

Definition 3.4 *Let $x, y : \mathcal{B}(R) \to \mathcal{F}$ be Q-observables. By the p-joint Q-observable h of x, y we understand a mapping $h : \mathcal{B}(R^2) \to \mathcal{F}$ satisfying the following conditions*

(i) $h(R^2) = (\mathbf{1}, \mathbf{0}); h(\emptyset) = (\mathbf{0}, \mathbf{1})$;

(ii) *if* $A \cap B = \emptyset$ *then* $h(A) \odot h(B) = (\mathbf{0}, \mathbf{1})$ *and* $h(A \cup B) = h(A) \oplus_Q h(B)$;

(iii) $A_n \nearrow A \Rightarrow h(A_n) \nearrow h(A)$;

(iv) $h(C \times D) = x(C).y(D)$ *for any* $C, D \in \mathcal{B}(R)$. *here* $(\mu_C, \nu_C).(\mu_D, \nu_D) = (\mu_C.\mu_D, 1 - (1 - \nu_C).(1 - \nu_D))$

Remark 3.5 *Analogously we can extend Definition 3.4 for finite collection of Q-observables.*

Theorem 3.6 *To any Q-observables $x, y : \mathcal{B}(R) \to \mathcal{F}$ there exists their p-joint Q-observable $h : \mathcal{B}(R^2) \to \mathcal{F}$.*

Proof Use instead of $\varphi(u) = u^n$ in Theorem 2.9 in [9]. □

4 Application of Q-observables

Let us mention one version of Central limit theorem: let $(\xi_i)_{i=1}^{\infty}$ be a sequence of independent, equally distributed, square integrable random variables,

$$E(\xi_i) = a, \sigma^2(\xi_i) = \sigma^2 \text{ for all } i \in N.$$

Then for any $t \in R$ there holds

$$\lim_{n\to\infty} p\left(\left\{\omega; \frac{\overline{\zeta_n}(\omega) - a}{\sigma}\sqrt{n} < t\right\}\right) = \Phi(t).$$

Here $\overline{\zeta_n} = \frac{1}{n}\sum_{i=1}^{n} \xi_i$ and $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{u^2}{2}} du$.
Now we are going to formulate an analogous assertion for Q-observables.
First, we shall mention some useful definitions:

Definition 4.1 *For any Q-probability $\mathcal{P} = [\mathcal{P}^\flat, \mathcal{P}^\sharp] : \mathcal{F} \to \mathcal{J}$ and any Q-observable $x : \mathcal{B}(R) \to \mathcal{F}$ we define the expected values by*

$$E_\flat(x) = \int_R t d\mathcal{P}_x^\flat(t);$$

$$E_\sharp(x) = \int_R t d\mathcal{P}_x^\sharp(t)$$

and the variances by

$$\sigma_\flat^2(x) = \int_R (t - E_\flat(x))^2 d\mathcal{P}_x^\flat(t);$$

$$\sigma_\sharp^2(x) = \int_R (t - E_\sharp(x))^2 d\mathcal{P}_x^\flat(t),$$

where $\mathcal{P}_x^\flat = \mathcal{P}^\flat \circ x, \mathcal{P}_x^\sharp = \mathcal{P}^\sharp \circ x$, assuming that the integrals exist.

Assume $T = (\xi_1, ..., \xi_n) : \Omega^n \to R^n$ is a random vector and $g : R^n \to R$ is a Borel measurable function (e.g. $g(u_1, ..., u_n) = \frac{1}{n}\sum_{i=1}^{n} u_i$). Then

$$g(\xi_1, ..., \xi_n) = g \circ T : \Omega^n \to R^n,$$

is a transformation of T. Hence we get the following formula

$$(g \circ T)^{-1}(A) = T^{-1}(g^{-1}(A))$$

for any $A \in \mathcal{B}(R)$. The formula justifies the following definition.

Definition 4.2 *Let $g_n : R^n \to R$ be a Borel function, $x_1, ..., x_n : \mathcal{B}(R) \to \mathcal{F}$ be Q-observables, $h_n : \mathcal{B}(R^n) \to \mathcal{F}$ their joint observable. Then the g_n-transformation of h_n is a Q-observable $y_n : \mathcal{B}(R) \to \mathcal{F}$ given by $y_n(A) = h_n(g_n^{-1}(A))$ for any $A \in \mathcal{B}(R)$.*

Definition 4.3 *Let $(x_n)_{n=1}^{\infty}$ be a sequence of Q-observables, $(h_n)_{n=1}^{\infty}$ be a sequence of the joint Q-observables $h_n : \mathcal{B}(R^n) \to \mathcal{F}$ of $x_1, x_2, ..., x_n$ (for $n \in N$), $m : \mathcal{F} \to [0, 1]$ be a Q-state. The sequence $(x_n)_{n=1}^{\infty}$ is independent (with respect to m), if for any $n \in N$ and any $C_1, C_2, ..., C_n \in \mathcal{B}(R)$ there holds*

$$m(h_n(C_1 \times C_2 \times ... \times C_n)) = m(x_1(C_1))...m(x_n(C_n)).$$

Definition 4.4 *A sequence $(x_n)_{n=1}^{\infty}$ of Q-observables is equally distributed, if $m(x_n(A)) = m(x_1(A))$ for any $n \in N$ and $A \in \mathcal{B}(R)$.*

Theorem 4.5 (Central limit theorem)
Let $(x_n)_{n=1}^{\infty}$ be a sequence of independent, equally distributed, square integrable Q-observables, where $E_\flat(x_n) = a^\flat, (E_\sharp(x_n) = a^\sharp)$ $\sigma_\flat^2(x_n) = \sigma_\flat^2$, $(\sigma_\sharp^2(x_n) = \sigma_\sharp^2)$ for each $n \in N$. Then for any $t \in R$ there the following holds

$$\lim_{n\to\infty} \mathcal{P}^\flat\left(\tfrac{x_1+...+x_n-na^\flat}{\sigma_\flat\sqrt{n}}((-\infty, t))\right) = \tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{u^2}{2}} du,$$

$$(\lim_{n\to\infty} \mathcal{P}^\sharp\left(\tfrac{x_1+...+x_n-na^\sharp}{\sigma_\sharp\sqrt{n}}((-\infty, t))\right) = \tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{u^2}{2}} du).$$

Proof Use Theorem 4.1 in [9]. □

5 Conclusion

Generalizing some notions proposed in [2] we constructed a $Q-$probability theory. The theory includes some known results ($n = 1, n = 2$), and also it opens the door for some other applications. We have proved some representation theorems. As an open question and an inspiration for the future research remains the problem of conditional probabilities for this framework.

Acknowledgement

The paper was supported by grant VEGA 1/0539/08.

ISBN: 978-989-95079-6-86

References

[1] Atanassov, K.: *Intuitionistic Fuzzy Sets: Theory and Applications.* Physica Verlag, New York, 1999.

[2] Atanassov, K., Riečan, B.: *On two new types of probability on IF-events.*In: Proc. of the First International Workshop on Intuitionistic Fuzzy Sets, Generalized Nets and Knowledge Engineering, Warsawa, 2008, pp. 11-22.

[3] Čunderlíková - Lendelová, K., Riečan, B.: *The probability theory on B-structures.* In: Proc.of the First International Workshop on Intuitionistic Fuzzy Sets, Generalized Nets and Knowledge Engineering, Warsawa, 2008, pp. 33-60.

[4] Gerstenkorn, T., Manko, J.: *Probabilities of intuitionistic fuzzy events.* In: Issues in Inteligent Systems: Paradigms (O.Hrzniewicz et al. eds.). EXIT, Warsawa, 63-58.

[5] Grzegorzewski, P., Mrowka, E.: *Probability of intuitionistic fuzzy events.* In: Soft Methods in Probablity, Statistics and Data Analysis. (P.Grzegorzewski et al. eds.). Physica Verlag, New York 2002, 105-115.

[6] Klement, E.P., Mesiar, R., Pap, E.:*Triangular Norms.* Trends in Logic, Vol. 8, Kluwer Academic Publishers, Boston/London/Dordrecht, 2000, 406 pp.

[7] Lendelová, K.: *Strong law of large numbers for IF-events.* Proc. Eleventh Int. Conf. IPMU, Paris 2006, 2363-2366.

[8] Renčová, M., Riečan, B.: *Probability on IF - sets: an elementary approach.* In: First International Workshop on Intuitionistic Fuzzy Sets, Generalized Nets and Knowledge Engineering. London : University of Westminster, 8 - 17, 2006.

[9] Renčová, M.: On the $\varphi-$probability and $\varphi-$observables. Submitted to Fuzzy Sets and Systems. 2008.

[10] Riečan, B.: *General form of M-probabilities on IF-events.* In: Proc. of IPMU, Spain, 1675-1677, 2008.

[11] Riečan, B.: *On a problem of Radko Mesiar: general form of IF probabilities.* Fuzzy Sets and Systems 152, 2006, 1485 - 1490.

[12] Riečan, B.: *Probability theory on intuitionistic fuzzy sets.* In: A volume in honor of Daniele Mundici´s 60^{th} birthday. Lecture Notes in Computer Science, 2007.

ISBN: 978-989-95079-6-87

A new bridge between fuzzy sets and statistics

Pedro Terán

Departamento de Estadística e I.O. y D.M.
Universidad de Oviedo
E-33071 Gijón (Spain)
Email: teranpedro@uniovi.es

Abstract— *Our basic question is as follows: given a, maybe multivariate, probability distribution, how to formalize the idea that some points are more central than others? We present an account of the notion of centrality which is based on fuzzy events and is valid for single distributions and for families of distributions (including imprecise probability models).*

This unifying framework is natural and subsumes many known concepts like the following (not an exhaustive list): (a) univariate location estimators like the mean, the median and the mode, (b) the interquartile interval and the Lorenz curve of a random variable, (c) several generalized medians, trimmed regions and statistical depth functions from multivariate analysis, (d) most known location estimators for random sets, (e) the probability mass function of a discrete random variable and the coverage function of a random closed set, (f) the Choquet integral with respect to an infinitely alternating or infinitely monotone capacity.

Keywords— Centrality, Fuzzy event, Multivariate analysis, Random set, Statistical depth.

1 Introduction

All points are central, but some points are more central than others.[1]

Our aim is to show how a simple idea with a natural interpretation within fuzzy set theory allows to define, in a unified way, many notions from statistics. These include well-known location estimators for random variables; some multivariate generalizations which grew into the theory of statistical depth functions in the last fifteen years; trimming methods based on them; means and medians of random sets; and even Choquet integrals. In fact, the notion applies as soon as we have a model to which we are able to associate a family of probability distributions.

The main tool is the notion of a fuzzy event and its probability [25]. A *fuzzy event* in a measurable space Ω is a measurable mapping $A : \Omega \to [0, 1]$. The membership $A(\omega)$ denotes the degree to which event A occurs when $\omega \in \Omega$ occurs. A probability measure on Ω extends to all fuzzy events by the natural formula

$$P(A) = \int A \mathrm{d}P.$$

The structure of this communication is as follows. Section 2 presents the fuzzy notion of centrality of a point in a probability distribution. Section 3 extends it to families of distributions. Sections 4, 5, 6 and 7 exhibit specific examples covered by our framework. Finally, Section 8 closes the paper with some final remarks.

[1] Paraphrasing a famous sentence from Orwell's 'Animal farm'.

2 Centrality and location estimators based on fuzzy events

What does it mean to speak about the 'centrality' of a point in a data set or a probability distribution? Although it is informally clear that the mean, the median and the mode are three traditional answers to the problem of pinpointing where the 'center' of a distribution is, a formal definition of centrality is far more elusive. In this section, an intuitively appealing view of centrality based on fuzzy set theory is presented.

We begin by choosing a family $\mathcal{A}$ of fuzzy events, called *reference events*. Roughly speaking, a point x is considered *central* if the probability of a reference event is high whenever it contains x. That is, *x is central if the probable reference events are implied by the occurrence of x*, whereas x is not central if it belongs to unlikely reference events. Note that this definition of centrality is relative to the chosen family of reference events.

This notion is reminiscent of the mode as the most likely point, and actually the mode is what appears when the reference events are chosen to be the (crisp) points.

It is possible to derive a crisp definition of centrality from the ideas above: we could say that a point is central if

$$P(A) \geq A(x) \quad \forall A \in \mathcal{A}.$$

But although this might capture the notion that the value of $P(A)$ is large insofar as the degree to which x belongs to A is large, a gradual restriction seems even more faithful to the idea. Thus, we say that a point is *central to degree at least* α (with respect to the family $\mathcal{A}$) if

$$P(A) \geq \alpha \cdot A(x) \quad \forall A \in \mathcal{A}.$$

The set of all the central points to degree at least α is the α-cut of a fuzzy set. That allows us to define the fuzzy set C given by

$$C(x) = \sup\{\alpha \in (0, 1] \mid \forall A \in \mathcal{A}, P(A) \geq \alpha A(x)\},$$

whose membership function indicates the degree of centrality of each point in the probability distribution P.

This definition can be rewritten in the language of fuzzy logic, as follows.

Proposition 1. *Denote Goguen's fuzzy implication by I. Then,*

$$C(x) = \inf_{A \in \mathcal{A}} I(A(x), P(A)).$$

Since $P(A)$ can be seen as the degree of truth of the proposition 'A is probable' in an adequate fuzzy logic (see e.g. [7]), the fuzzy set C captures the intuitive idea that its elements are those such that *'for any reference event A, if $x \in A$ then A is probable'*.

In principle, the Goguen implication can be replaced by a different implication. In this paper, we content ourselves with showing that many statistical objects arise from this choice (whose reason is historical, as explained in the final section), without analyzing other possible choices. However, note that changing the implication may lead to obtaining the same examples with a mere modification of the shape of the reference events.

From a statistical point of view, those points with the maximal centrality can serve as location estimators which might be called *maximal centrality estimators*. In general, there may be more than one maximally central point (like e.g. several modes), so all the maximally central points form the *central region*. Points with centrality at least α form the *α-central region*.

3 Extension to families of distributions

As shown in Section 4, the notion of centrality presented above is directly related to some traditional location estimators, as well as to several multivariate generalizations which have been introduced and studied mostly in the last fifteen years. However, to account for some of those cases and specially to cover more general types of models like random sets, it is necessary to extend it to study the centrality of a point in a *family* of distributions.

When the model consists of more than one single distribution, the underlying distribution generating the data is known or assumed to belong to a specified family. Some situations covered by this more general setting are the following.

(a) A parametric or non-parametric model.

(b) Upper and lower probabilities.

(c) A Choquet capacity.

(d) A random set.

(e) A credal set.

(f) A neighbourhood of a probability.

(g) Other families defined in terms of the distribution of a random variable.

Given a random variable ξ, instances of (g) include the family $\{P_{g \cdot \xi} \mid g \text{ is } [0,1]\text{-valued}\}$ used to define the zonoid [11] of ξ; and the family $\{Q \leq \alpha^{-1} P_\xi\}$ used to define trimmed regions in [4]). In such examples, a family of distributions is constructed whose central region contains interesting information about ξ, although maybe no longer related to centrality itself. For instance, the lift zonoid of ξ, a multivariate generalization of the Lorenz curve, is the central region of a family of distributions defined from the distribution of ξ.

In such a that situation, for statistical purposes a point should be considered central in the family if it is potentially a central point of the true distribution, which could be any member of the family. The rationale is that a model defined by a family of distributions is more imprecise than pinpointing a single distribution, so the compatibility of a point being central with the information available should be larger [2].

That suggests the following definition: given a family $\mathcal{P}$ of probability distributions, the fuzzy set C of central points is defined by

$$C(x) = \sup\{\alpha \in (0,1] \mid \forall A \in \mathcal{A}, \sup_{P \in \mathcal{P}} P(A) \geq \alpha A(x)\}.$$

The rest of the definitions (α-central region, etc.) are modified analogously.

Observe that the location information in a family of distributions is summarized using the same object as for a single distribution. Thus, the complexity of that summary is the same as for a single distribution (although calculating C may be computationally more expensive). That implies that statistical procedures based on C, devised for single distributions, will extend immediately to the situations above, particularly imprecise probability models (Choquet capacities, random sets).

As will be shown later, in the cases when C coincides with a statistical depth function such procedures already exist.

Let us finally note that some of the results below may require, for technical reasons, that the family $\mathcal{P}$ considered is not an arbitrary one but compact with respect to the topology of weak convergence (convergence in distribution).

4 Univariate examples

As already mentioned, the three basic examples of location estimators arise as special cases of maximally central points of a distribution.

Theorem 2. *Let ξ be a random variable. For appropriate choices of $\mathcal{A}$ and $\alpha \in (0,1]$, the α-central region C_α is:*

(1) The expectation of ξ, if it exists (otherwise, the corresponding α-central region is empty).

(2) The median of ξ.

(3) The mode of ξ, if ξ is discrete.

The necessary choices of α in Theorem 2 are: $\alpha = 1$ in (1); $\alpha = .5$ in (2); α the height of C in (3). In fact, in all the three cases α coincides with the height of the fuzzy set C. In subsequent examples, for ease of presentation we omit the choice of α in each case.

[2] In any case, it is also interesting to consider the dual approach: to which degree a point is guaranteedly central in *all* the distributions of the model. That was pointed out to me by Didier Dubois.

As to the other ingredient, $\mathcal{A}$, one may ask whether the families of reference events needed to retrieve these classical estimators may be very complicated. In fact, they are quite simple:

(1) $\{A_y(x) = |x-y|/(1+|x-y|)\}_{y\in\mathbf{R}}$ for the expectation.

(2) $\{A_y^+(x) = I_{[y,\infty)}(x), A_y^-(x) = I_{(-\infty,y]}(x)\}_{y\in\mathbf{R}}$ for the median.

(3) $\{A_y(x) = I_{\{y\}}(x)\}_{y\in\mathbf{R}}$ for the mode.

In (2) and (3), I_B denotes the indicator function of a crisp set B.

Subsequent examples use the same choices of $\mathcal{A}$ or suitable multivariate generalizations. Again to avoid burdensome presentation, we omit the specific choice of $\mathcal{A}$ in each case.

Interestingly, (1) is associated to a new characterization of the expectation of a random variable in terms of the bounded metric $d(x, y) = |x - y|/(1 + |x - y|)$.

Proposition 3. *Let ξ be an integrable random variable. Then, its expectation is the unique real number x such that, for any number y,*

$$d(x, y) \leq Ed(\xi, y);$$

namely $E\xi$ is identified by the property that every number is closer in d-distance to $E\xi$ than it is (in mean) to ξ.

Proof (outline of the main ideas). There exists a characterization of the mean, due to Shafik Doss [6], as the unique real number x such that, for any y,

$$\|x - y\| \leq E\|\xi - y\|.$$

It suffices to show that both properties are equivalent. The easy part, with the help of the Jensen inequality, is to show that a number satisfying the condition in the statement must also fulfil Doss's property (and so our condition is formally stronger and only the mean could satisfy it).

To prove that the mean has indeed the property in the statement we need to go against Jensen's inequality, so things get more involved. The key is to take a divergent sequence of values for y and show that, when y is very far from x, Jensen's inequality is almost an equality, so in the limit we can 'reverse' it. □

Actually, this geometric characterization is valid in any separable Banach space, if we just replace the Euclidean distance by the norm.

A couple more univariate examples are as follows. These are no longer related to location estimation.

Proposition 4. *Let ξ be a random variable. For appropriate choices of $\mathcal{A}$ and $\alpha \in (0, 1]$, the α-central region C_α is:*

(1) The interquartile interval of ξ.

(2) The Lorenz curve of ξ (more exactly, the curve together with the region below it, its hypograph).

Both the interquartile interval and the median are examples of the interval between the p-th quantile and the $(1 - p)$-th quantile (for $p = .25$ and $p = .5$). Modulating the value of α (precisely, $\alpha = .5 - p$), we can obtain any such interval for $0 \leq p \leq .5$. For $p = 0$, that is the interval between the minimum and maximum values of ξ.

In this connection, let us remark that two known tools in exploratory data analysis, the box-plot and its multivariate generalization (the bag-plot), are plotted using only information subsumed by our framework.

5 Multivariate examples

In the multivariate case, the absence of a natural order makes it impossible for a point to have all the usual properties of the univariate median. That has led to the definition of many *depth functions* [14]; often the point maximizing the depth function is called a *generalized median* or a *maximal depth estimator*.

Depth functions aim at ordering, in a center-outward sense, the points in a data set or, more generally, those of $\mathbf{R}^d$ with respect to a given distribution. A typical application of depth (among many others) is outlier detection: sample points with very small depth are trimmed out as outliers.

The notion of depth is still somewhat vague, and its semantics is specially unclear for multimodal or very asymmetric distributions. Zuo and Serfling [26] underlined four desirable properties of depth functions, which had already been studied in particular cases:

(a) Depth is maximal at a center of symmetry of the distribution, if the latter exists.

(b) Depth decreases along any ray departing from the center.

(c) Depth vanishes as the distance to the center goes to infinity.

(d) The depth function should be affinely equivariant (so that conclusions do not depend on the chosen coordinate system).

It must be emphasized that, often, accepted notions of statistical depth fail some of those properties (e.g. L^p-depth).

From any depth function, an *α-trimmed region* can be constructed which is formed by those points having depth at least α. Plotting the contours of the α-trimmed regions constitutes an easy device for multivariate exploratory analysis.

To give the reader a grasp of how depth functions work, let us just mention some ways to calculate depth.

Convex hull peeling depth of a point in a data sample is calculated as follows. Take the convex hull of the sample. The points outside have depth 0. Now the sample points in the boundary of the convex hull are 'peeled off'; points left outside the cloud by doing so have depth

proportional to 1. The sample points in the boundary of the convex hull of the remaining points are peeled off, and points left outside the cloud are given depth proportional to 2. The procedure goes on until the innermost points have been peeled off.

Simplicial depth of a point is calculated as the probability that it belongs to the random simplex whose vertices are independent and identically distributed copies of the variable.

Similarly, expected convex hull depth is inversely proportional to the number k of copies of the variable which are needed to ensure that the point is in the Aumann expectation of the sample $\{\xi_1, \cdots, \xi_k\}$.

Finally, for halfspace depth one considers the probability that the variable lies in a halfspace, and calculates the infimum of such values over those halfspaces whose boundary contains the point.

Some statistical depth functions in the literature are fuzzy sets of central points, as the following result shows.

Theorem 5. *Let ξ be a random vector. For appropriate choices of $\mathcal{A}$, $\mathcal{P}$ and $\alpha \in (0, 1]$, the centrality $C(x)$ is:*

(1) The convex hull peeling depth of x (Huber [9], 1972; Barnett [2], 1976).

(2) The simplicial depth of x (Liu [15], 1990).

(3) The majority depth of x (Liu–Singh [16], 1993).

(4) The probability of the event $\{\xi = x\}$, if ξ is discrete.

The univariate examples presented above extend to the multivariate setting as follows (including the central regions associated to further notions of depth).

Theorem 6. *Let ξ be a multivariate random vector. For appropriate choices of $\mathcal{A}$, $\mathcal{P}$ and $\alpha \in (0, 1]$, the α-central region C_α is:*

(1) The expectation of ξ, if it exists.

(2) The generalized median of ξ according to any of the depth functions (1-3) above.

(3) The α-trimmed region of ξ according to any of the depth functions (1-3) above.

(4) The generalized median and the α-trimmed region of ξ according to the halfspace depth (Tukey [23], 1975).

(5) The generalized median and the α-trimmed region of ξ according to the zonoid depth (Mosler [17], 2002).

(6) The generalized median and the α-trimmed region of ξ according to the symmetric order depth (Cascos–López-Díaz [5], 2004).

(7) The generalized median and the α-trimmed region of ξ according to the expected convex hull depth (Cascos [3], 2007).

(8) The mode of ξ, if ξ is discrete.

(9) The zonoid of ξ (see e.g. [11, 17]).

(10) The lift zonoid of ξ (Koshevoy–Mosler [12], 1998).

The lift zonoid generalizes the Lorenz curve and so serves to study multivariate concentration and inequality. It is also remarkable that it characterizes the distribution of ξ.

6 Multivalued examples

When a random set is considered, the statistical model takes the following form. The underlying random variable ξ is not directly observed; only a larger set of values X containing ξ is observed. Then, we know that ξ is such that $\xi \in X$ almost surely, namely ξ is a *selection* of X. In that case, $\mathcal{P}$ is the family of all selections of X.

Most location estimators for random sets are unified as special cases of central or α-central regions.

Theorem 7. *Let X be a random compact set and $\mathcal{P}$ the family of all its selections. Then, for appropriate choices of $\mathcal{A}$ and $\alpha \in (0, 1]$, the α-central region C_α is:*

(1) The Aumann expectation of X (Kudō [13], 1954; Aumann [1], 1965) if Ω is non-atomic or X is convex (in general, the convex hull of the Aumann expectation).

(2) The Vorob'ev expectation of X (Vorob'ev [24], 1984).

(3) The Herer expectation of X (Herer [8], 1990).

(4) The radius-vector expectation of X (Stoyan–Stoyan [18], 1994).

(5) The Vorob'ev median of X (Stoyan–Stoyan [18], 1994).

Another notion covered by our framework is the *coverage function* of a random set X, given by $p_X(x) = P(x \in X)$.

Proposition 8. *Let X be a random closed set. Then, p_X is both an integral depth function in the sense of [22] and a fuzzy set of central points, corresponding to taking $\mathcal{A}$ to be the family of all crisp points and $\mathcal{P}$ the family of all selections of X.*

The corresponding central region is formed by all the *fixed points* of the random set.

7 Choquet capacities

Let ν be a *Choquet capacity* in the sense of [10], namely a function from the Borel σ-algebra of $\mathbf{R}^d$ to $[0, 1]$ with the following properties:

a) $\nu(\emptyset) = 0$, $\nu(\Omega) = 1$, $\nu(A) \leq \nu(B)$ if $A \subset B$,

b) $\nu(C_n) \searrow \nu(C)$ if $C_n \searrow C$ and C_n, C are closed,

c) $\nu(A_n) \nearrow \nu(A)$ if $A_n \nearrow A$.

A more general definition, in which closed sets are replaced by compact sets, is possible. A capacity is called *2-alternating* if

$$\nu(A \cup B) + \nu(A \cap B) \leq \nu(A) + \nu(B),$$

and *infinitely alternating* if

$$\nu(\bigcap_{i=1}^{n} A_i) \leq \sum_{\emptyset \neq I \subset \{1,\ldots,n\}} (-1)^{\mathrm{card}(I)+1} \nu(\bigcup_{i \in I} A_i)$$

whenever $A_1, \ldots, A_n$ are Borel sets.

The dual capacity to ν is given by $\overline{\nu}(A) = 1 - \nu(A^c)$. The dual to a 2-alternating or infinitely alternating capacity has the property called 2-monotonicity or infinite monotonicity, respectively. The *Choquet integral* of a random variable ξ with respect to ν is

$$\int_{(C)} \xi \mathrm{d}\nu = \int_0^\infty \nu(\{\xi \geq t\})\mathrm{d}t + \int_{-\infty}^0 [\nu(\{\xi \geq t\}) - 1]\mathrm{d}t,$$

which exists provided at least one of the two improper Riemann integrals is finite. In that case, ξ is called ν-integrable.

Our notion of centrality can be applied to capacities by taking $\mathcal{P}$ to be the core of ν, namely the family of all probability distributions $P \leq \nu$ dominated by ν. Observe that Zadeh's definition of the probability of a fuzzy event extends immediately to capacities by setting $\nu(A) = \int_{(C)} A \mathrm{d}\nu$.

Proposition 9. *Let ν be a 2-alternating Choquet capacity and let I be Goguen's fuzzy implication. Then, taking $\mathcal{P}$ to be the core of ν results in*

$$C(x) = \inf_{A \in \mathcal{A}} I(A(x), \nu(A)).$$

The Choquet integrals with respect to ν and $\overline{\nu}$ are sometimes called the upper and lower Choquet integrals. They can be retrieved in our framework.

Theorem 10. *Let ν be an infinitely alternating Choquet capacity and $\mathcal{P}$ its core. Let ξ be a ν-integrable random variable. Then, for an appropriate choice of $\mathcal{A}$, the central region C_1 is the interval between $\int_{(C)} \xi \mathrm{d}\overline{\nu}$ and $\int_{(C)} \xi \mathrm{d}\nu$.*

As a consequence, we obtain the following geometric characterizations of the Choquet integral.

Proposition 11. *Let ν be a Choquet capacity, and let ξ be a ν-integrable random variable. If ν is infinitely alternating, then the Choquet integral $\int_{(C)} \xi \mathrm{d}\nu$ is the largest value x satisfying either of the following equivalent properties:*

(1) $|a - x| \leq \int_{(C)} |a - \xi| \mathrm{d}\nu$ *for all* $a \in \mathbf{R}$,

(2) $d(a, x) \leq \int_{(C)} d(a, x) \mathrm{d}\nu$ *for all* $a \in \mathbf{R}$.

Analogously, if ν is infinitely monotone, then $\int_{(C)} \xi \mathrm{d}\nu$ is the smallest value x satisfying either property.

8 Concluding remarks

1. The proofs of the results presented in this contribution use pieces of information from several submitted papers, mostly [22, 20] but also [19, 21], which are not explicitly about fuzzy sets. The proofs will be presented in a forthcoming paper. The communication format allows us to present the ideas in a convenient sequence from particular to general, although the logical path of their derivation is more involved.

2. There is an interesting connection between defuzzification and depth-based statistical estimation. Indeed, defuzzification (understood as obtaining a single point, not a crisp set) of a fuzzy set of central points appears very close to location estimation. The two main methods of depth-based estimation in the literature are generalized medians and the depth-weighted estimators given by

$$DWE = \int xD(x)\mathrm{d}x / \int D(x)\mathrm{d}x,$$

see e.g. [27]), which correspond to the maximum and centroid methods of defuzzification.

3. The choice of the Goguen implication, rather than any other fuzzy implication, is rooted in earlier works which inspired our own. The genealogy is as follows. Koshevoy and Mosler [11] defined the *zonoid trimmed regions*

$$\{E[\xi \cdot g(\xi)] \mid g : \mathbf{R}^d \to [0, \alpha^{-1}] \text{ is measurable}\}$$

formed by the expectations of ξ with respect to all probability measures Q such that $\frac{\mathrm{d}Q}{\mathrm{d}P} \leq \alpha^{-1}$. Cascos generalized that idea to an abstract definition of *integral trimmed regions*

$$D^\alpha_{\mathcal{F}}(x) = \{x \mid \exists Q \leq \alpha^{-1} P \ : \ \forall f \in \mathcal{F} \ f(x) \leq \int f \mathrm{d}Q\}$$

where $\mathcal{F}$ is a family of integrable functions) which were studied in his Ph.D. thesis and the paper [4]. One already finds $P \geq \alpha Q$ explicitly there, although as a comparison between probabilities. Integral trimmed regions in the sense of [4] receive a natural interpretation in connection to stochastic orders. They coincide with integral trimmed regions in the sense of [22] (no longer interpretable from stochastic orders) if $\mathcal{F}$ is formed by upper semicontinuous functions. In turn, those are a special case of integral central regions [22], which are very similar to central regions in this paper (although, the notions of integral depth in [22] and fuzzy centrality in this paper seem to be more far apart).

4. It is natural to ask whether fuzzy centrality and/or statistical depth are in turn related to Zadeh's proposal that probability theory should include a notion of the *usual value* of a variable.

Acknowledgment

Research has been partially carried out while the author was with the *Universidad de Zaragoza*, and partially funded by Spain's *Ministerio de Educación y Ciencia* (MTM2005-02254 and TSI2005-02511), *Ministerio de Ciencia e Innovación* (MTM2008-01519) and Aragón's *Gobierno* (PM2007-034).

References

[1] R. J. Aumann (1965). Integrals of set-valued functions. *J. Math. Anal. Appl.* 12:1–12.

[2] V. Barnett (1976). The ordering of multivariate data. *J. Royal Statist. Soc. A* 139:318–352 (with discussion).

[3] I. Cascos (2007). The expected convex hull trimmed regions of a sample. *Comput. Statist.* 22:557–569.

[4] I. Cascos, M. López-Díaz (2005). Integral trimmed regions. *J. Multivar. Anal.* 96:404–424.

[5] I. Cascos-Fernández, M. López-Díaz (2004). The zonoid hull of a multivariate probability distribution and its trimmed regions. In: M. López-Díaz, M. Á. Gil, P. Grzegorzewski, O. Hryniewicz, J. Lawry, editors. *Soft methodology and random information systems*, 423–430. Springer, Berlin.

[6] S. Doss (1949). Sur la moyenne d'un élément aléatoire dans un espace distancié [*On the mean of a random element in a metric space*, in French]. *Bull. Sci. Math.* 73:1–26.

[7] P. Hájek, L. Godo, F. Esteva (1995). Fuzzy logic and probability. *Proc. 11th Annual Conf. Artif. Intelligence*, 237–244.

[8] W. Herer (1990). Mathematical expectation and martingales of random subsets of a metric space. *Probab. Math. Statist.* 11:291–304.

[9] P. J. Huber (1972). Robust statistics: a review. *Ann. Math. Statist.* 43:1041–1067.

[10] P. J. Huber, V. Strassen (1973). Minimax tests and the Neyman-Pearson lemma for capacities. *Ann. Statist.* 1:251–263.

[11] G. Koshevoy, K. Mosler (1997). Zonoid trimming for multivariate distributions. *Ann. Statist.* 25:1998–2017.

[12] G. Koshevoy, K. Mosler (1998). Lift zonoids, random convex hulls and the variability of random vectors. *Bernoulli* 4:377–399.

[13] H. Kudō (1954). Dependent experiments and sufficient statistics. *Nat. Sci. Rep. Ochanomizu Univ.* 4:151–163.

[14] R. Liu, R. Serfling, D. Souvaine, editors (2006). *Data depth: robust multivariate analysis, computational geometry and applications*. American Mathematical Society, Providence.

[15] R. Y. Liu (1990). On a notion of data depth based on random simplices. *Ann. Statist.* 18:405–414.

[16] R. Y. Liu, K. Singh (1993). A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc.* 88:252–260.

[17] K. Mosler (2002). *Multivariate dispersion, central regions and depth. The lift zonoid approach*. Lect. Notes Statist. 165. Springer, Berlin.

[18] D. Stoyan, H. Stoyan (1994). *Fractals, random shapes and point fields*. Wiley, Chichester.

[19] P. Terán. On the equivalence of Aumann and Herer expectations of random sets. *Test* 17:505–514.

[20] P. Terán. Classical location estimators for random sets in the light of statistical depth. *Submitted for publication*.

[21] P. Terán. Herer and Aumann expectations of random sets. *Submitted for publication*.

[22] P. Terán. Integral central regions and integral depth. *Submitted for publication*.

[23] J. W. Tukey (1975). Mathematics and the picturing of data. *Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974)*, 523–531. Canad. Math. Congress, Montreal.

[24] O. Y. Vorob'ev (1984). *Mean-measure modelling* [in Russian]. Nauka, Moscow.

[25] L. A. Zadeh (1968). Probability measures of fuzzy events. *J. Math. Anal. Appl.* 23:421–427.

[26] Y. Zuo, R. Serfling (2000). General notions of statistical depth function. *Ann. Statist.* 28:461–482.

[27] Y. Zuo, H. Cui, X. He (2004). On the Stahel-Donoho estimator and depth-weighted means of multivariate data. *Ann. Statist.* 32:167–188.

L-Prime Spectrum of Modules

Bayram Ali Ersoy[1] and Dan Ralescu[2]
1 Yildiz Technical University, Department of mathematics, Davutpasa Kampusu, Esenler, Istanbul, TURKEY.
2 University of Cincinnati, Cincinnati, OH 45221-0025, USA.
E-mail : ersoya@yildiz.edu.tr, dan.ralescu@uc.edu

Abstract: *We investigate the Zariski Topology on the L-prime spectrum of modules consisting of the collection of all prime L-submodules and prove some useful results.*

Keywords: L- ideal, L-submodule, L-prime submodule, Special L-submodule, L-Irreducible submodule, L-Prime Spectrum, Zariski topology, Generic point.

1 Introduction

Throughout this paper *R* is a commutative ring with identity and *M* is a unitary *R*-module. The prime spectrum *Spec(R)* and the topological space obtained by introducing Zariski topology on the set of prime ideals of a commutative ring with identity play an important role in the fields of commutative algebra, algebraic geometry and lattice theory. Also, recently the notion of prime submodules and Zariski topology on *Spec(M)*, the set of all prime submodules of a module *M* over a commutative ring with identity *R*, are studied by many authors (for example see [1,2], [3]) that is in [3] a module with Zariski topology is called *top module* and it is shown that every multiplication module is a topmodule. As it is well known [4], introduced the notion of a fuzzy subset μ of a nonempty set *X* as a function from *X* to unit real interval $I=[0,1]$. [5] replaced I by a complete lattice *L* in the definition of fuzzy sets and introduced the notion of *L*-fuzzy sets. the notion of fuzzy groups was introduced by Rosenfeld [6]); and fuzzy sub-modules of *M* over *R* were first introduced by Negoita and Ralescu [7]. İn [8] Pan studied fuzzy finitely generated modules and fuzzy quotient modules (also see [9]). In the recent five years a remarkable amount of work has been done on fuzzy ideals in general and prime fuzzy ideals in particular, and some interesting topological properties of the spectrum of fuzzy prime ideals of a ring are obtained (see [10], [11], [12], [13], [14], [15] and [16,17].

Finally R.Ameri and R. Mahjoob 2007 studied Zariski topology on $L-Spec(M)$ defining all prime submodules of M. They investigated some basic properties of prime L-submodules and characterize the prime L-submodules of M. They established the relationship between primeless and L-primeless for a given module via the role of the lattice L. Finally they investigated the Zariski topology on $L-Spec(M)$. They showed that for L-top modules Zariski topology on $L-Spec(M)$ exists. In this paper we contınue the paper [18] R.Ameri and R. Mahjoob 2007. $L(M)$ defines all $L-$submodules of M. We define L-special submodule of M to construct a topology on L-spec(M). Then we examine the relationship between with the L-special submodule and L-irreducible submodule. L_T_p module is defined to construct a topology. We introduce that If $L(M)$ is L_T_p module then L-Spec(M) is a T_0 space. We show that L-Spec(M) is a T_1 space if and only if for every $\mu\in\text{L-Spec(M)}$ is maximal. The topological space $(\text{L-Spec(M)}=\aleph,T)$ where M is a Noetherian $R-$module is compact. We show that where Y is a closed subset of $\aleph$, If $J(Y)$ is L-prime submodule of M then Y is irreducible. And if $\eta\in Y$ is a generic point of Y then Y is irreducible.

2 Preliminaries.

Throughout of this paper, by *R* we mean a commutative ring with identity, and *M* is a unital *R*-module and *L* denotes a complete lattice. By an *L*-subset μ of a non-empty set *X*, we mean a function μ from *X* to *L* and if *L* = [0, 1], then μ is called a fuzzy subset of *X*. *LX* denotes the set of all *L*-subsets of *X*.

Definition 2.1: A fuzzy ideal of R is a fuzzy subset of R such that;
i- $\mu(x-y)\geq\min(\mu(x),\mu(y)) \quad \forall x,y\in R$
ii- $\mu(xy)\geq\mu(x) \quad \forall x,y\in R$
The set of all *L*-ideals of *R* is denoted by *L I (R)*.

Definition 2.2: If μ is a fuzzy subset of M, then for any $t\in[0,1]$ the set $\mu_t=\{x\in M \mid \mu(x)\geq t\}$ is called a level subset of M with respect to μ.

Definition 2.3([19]): Let μ and σ be a fuzzy ideal of R Then $\forall x\in R$ there exists $y,z\in R$ s.t.
$\mu.\sigma(x,y)=\sup_{x=yz}\{\min(\mu(y),\sigma(z))\}$.

Definition 2.4: Let R be a ring and $\mu\in LI(R)$. A non constant μ is called L-prime ideal for every $\eta,\beta\in LI(R)$, $\eta\beta\subseteq\mu$ implies that $\eta\subseteq\mu$ or $\beta\subseteq\mu$.

By L-Spec(R) we mean the set of all prime *L*-ideals of *R*.

Definition 2.5: A fuzzy submodule of M is a fuzzy subset of M such that
1- $\mu(0)=1$

2- $\mu(rx) \geq \mu(x) \quad \forall r \in R$ and $\forall x \in M$

3- $\mu(x+y) \geq \min(\mu(x), \mu(y)) \quad \forall x, y \in M$

L-(M) denotes the set of all *L*-submodules of M .

Theorem 2.6: μ is a fuzzy submodule of M if and only if μ_t is a submodule of M for all $t \in \operatorname{Im}\mu$.

Definition 2.7 ([20])**:** For $\mu, \nu \in L^M$ and $\eta \in L^R$, define $\mu : \eta \in L^M$ as follows: $\mu : \eta = \bigcup \{ \nu \in L^M \mid \eta\nu \subseteq \mu \}$.

Definition 2.8: Let M be an R module and $\mu \in L(M)$. A non constant μ is called L-prime submodule for every $\eta \in LI(R)$ and $\beta \in L(M)$, $\eta\beta \subseteq \mu$ implies that $\eta \subseteq \mu$ or $\beta \subseteq \mu : 1_M$.

By L-Spec(M) we mean the set of all *L*- prime submodules of M .

3 L-prime Spectrum on Modules

Example3.1:

$$\mu(x,y) = \begin{cases} 1 & \text{if } (x,y) = (0,0) \\ 1/3 & \text{if } (x,y) \in (\mathbb{Z},0) - \{(0,0)\} \\ 0 & \text{otherwise} \end{cases}$$ is L-submodule of $(\mathbb{Z},0)$,

$$\sigma(x,y) = \begin{cases} 1 & \text{if } (x,y) = (0,0) \\ 1/2 & \text{if } (x,y) \in (0,\mathbb{Z}) - \{(0,0)\} \\ 0 & \text{otherwise} \end{cases}$$ is L-submodule of $(0,\mathbb{Z})$ and

$$\beta(x,y) = \begin{cases} 1 & \text{if } (x,y) = (0,0) \\ 0 & \text{otherwise} \end{cases}$$ is L-prime submodule of $(\mathbb{Z},\mathbb{Z})$. $\mu \not\subset \beta$ and $\nu \not\subset \beta$ but $\mu \cap \nu \subseteq \beta$. We show that $\eta \cap \nu \subseteq \mu \Rightarrow \eta \subseteq \mu \text{ or } \nu \subseteq \mu$ is not always true for L-prime submodules in general; Every prime submodule is not L-Special submodule.

Definition 3.2: A submodule μ of M is called L-Special submodule, if for any η, ν L-submodule of M such that $\eta \cap \nu \subseteq \mu \Rightarrow \eta \subseteq \mu \text{ or } \nu \subseteq \mu.$

Example 3.3: Every prime L-ideal of M is L-Special submodule of R module R .

Definition 3.4: Let μ be a L-subset of M . $V(\mu)$ is a subset of all L-prime submodules of M defined by $V(\mu) = \{ \eta \mid \eta \text{ is } L-\text{prime submodule of } M,\ \mu \subseteq \eta \}$

For any L-submodule μ of M , it is easy to show that $V(1_M) = \varnothing$, $V(0_M) = L - Spec(M)$, $\bigcap_{i \in I} V(\mu_i) = V(\sum_{i \in I} \mu_i)$ and $V(\eta) \cup V(\nu) \subseteq V(\eta \cap \nu)$.

Proposition 3.5: Let M be an R module. For any L-submodule η, ν of M , $V(\eta \cap \nu) = V(\eta) \cup V(\nu)$ if and only if every L-prime ideal μ of M is L-Special submodule.

Proof: Suppose that for all $\eta, \nu \in L(M)$, $V(\eta \cap \nu) = V(\eta) \cup V(\nu).$

Let $\eta \cap \nu \subseteq \mu$ for all L-prime ideal μ of M .

$$\eta \cap \nu \subseteq \mu \Rightarrow \mu \in V(\eta \cap \nu) = V(\eta) \cup V(\nu)$$
$$\Rightarrow \mu \in V(\eta) \text{ or } \mu \in V(\nu)$$
$$\Rightarrow \eta \subseteq \mu \text{ or } \nu \subseteq \mu.$$

Conversely, let any every L-prime ideal μ of M be L-Special submodule. Then

$$\mu \in V(\eta \cap \nu) \Leftrightarrow \eta \cap \nu \subseteq \mu$$
$$\Leftrightarrow \eta \subseteq \mu \text{ or } \nu \subseteq \mu$$
$$\Leftrightarrow \mu \in V(\eta) \cup V(\nu).$$

Definition 3.6: A L-submodule μ of M is called L-irreducible submodule, if $\eta \cap \nu = \mu \Rightarrow \eta = \mu$ or $\nu = \mu$ for any η, ν L-submodule of M .

Proposition 3.7: Every L-special submodule μ of M is L-irreducible module.

Proof: Let $\eta \cap \nu = \mu.$ Then $\eta \subseteq \mu \text{ or } \nu \subseteq \mu$. And also $\mu \subseteq \eta \text{ or } \mu \subseteq \nu$ then $\eta = \mu \text{ or } \nu = \mu.$

Proposition 3.8: Let μ be L-submodule of M , ν be prime L-submodule of M . If $[\eta : \mu] \subseteq [\nu : \mu]$ implies $\eta \subseteq \nu$ for each η L-submodule of M , then ν is L-special submodule of M.

Proof: Let $\eta \cap \mu \subseteq \nu.$ Since $[\nu : 1_M]$ is L-prime ideal by R. Ameri ([1]) theorem 3.6

$$[\eta \cap \mu : 1_M] \subseteq [\nu : 1_M] \Rightarrow [\eta : 1_M] \cap [\mu : 1_M] \subseteq [\nu : 1_M]$$
$$\Rightarrow [\eta : 1_M] \subseteq [\nu : 1_M] \text{ or } [\mu : 1_M] \subseteq [\nu : 1_M]$$
$$\Rightarrow \eta \subseteq \nu \text{ or } \mu \subseteq \nu.$$

Definition 3.9: An L-submodule of μ of M is called T' L-submodule, if

i) $(\eta \cap \nu)\mu = \eta\mu \cap \nu\mu \qquad \forall \eta, \nu \in$ LI(R)

ii) $[\eta:\mu]\subseteq[\nu:\mu]\Rightarrow\eta\subseteq\nu \quad \forall\eta,\nu\in L(M)$.

Theorem 3.10: Let μ be $T^{'}$ L-submodule of M. Then η is L-special submodule if and only if for $(\eta\subseteq\mu)$ $[\eta:\mu]$ is a special L-ideal.

Proof: Let $\nu,\beta\in LI(R)$ and $\mu,\eta\in L(M)$ such that $\nu\cap\beta\subseteq[\eta:\mu]$. Then since
$[\eta:\mu]\mu\subseteq\eta$,
$(\nu\cap\beta)\mu\subseteq[\eta:\mu]\mu\Rightarrow\nu\mu\cap\beta\mu\subseteq\eta$. Since η is L-special submodule, $\nu\mu\subseteq\eta \ \ or\ \beta\mu\subseteq\eta$. Then $\nu\subseteq[\eta:\mu]$ or $\beta\subseteq[\eta:\mu]$.

Conversely, let $\eta\cap\nu\subseteq\mu$ for all $\nu,\eta\in LI(R)$ and $\mu,\beta\in L(M)$ Then $[\eta\cap\nu:\beta]\subseteq[\mu:\beta]\Rightarrow[\eta:\beta]\cap[\nu:\beta]\subseteq[\mu:\beta]$. Therefore for $T^{'}$ L-submodule then $[\eta:\beta]\subseteq[\mu:\beta]\ or\ [\nu:\beta]\subseteq[\mu:\beta]$. Then $\eta\subseteq\mu$ or $\nu\subseteq\mu$.

For any L-submodule μ of M, $V(\mu)$ denotes the set of all prime submodules of M containing μ, that is; $V(\mu)=\{\eta\in\aleph|\ \mu\subseteq\eta\}$. It is clear that; $V(1_M)=\{\ \}$ and $V(0_M)=\aleph$. Also we can show that for any family of submodules of $\{\mu_i\}_{i\in I}$ of M and $\mu,\nu\in L(M)$, $\bigcap_{i\in I}V(\mu_i)=V(\sum_{i\in I}\mu_i)$, $V(\mu)\bigcup V(\nu)\subseteq V(\mu\bigcap\nu)$. Therefore if $\zeta(M)$ denotes the collection of all subsets $V(\mu)$ of $L-Spec(M)$ then $\zeta(M)$ contains empty set and itself. Also $\zeta(M)$ closed under arbitrary intersection. But $\zeta(M)$ is not closed under finite union. Therefore on the set of all prime submodules Zariski topology does not exist since third rule of Zariski topology is not satisfied, i.e; $V(\mu\cap\nu)\neq V(\mu)\cup V(\nu)$. If this inequality replaces with equality these modules can be called as L_T_p modules.

Definition 3.11: Let μ,ν be L-submodule of M. $L(M)$ is called L_T_p module if and only if $V(\mu\cap\nu)=V(\mu)\cup V(\nu)$.

Lemma 3.12: $L(M)$ is L_T_p module if and only if every L-prime submodule μ is L-special submodule of M.

Proof: From proposition 2.5.

Lemma 3.13: Let μ,β be an L-submodule of M. If $(\alpha\subseteq\mu)$ such that $\langle\alpha\rangle=\beta$, then $V(\alpha)=V(\beta)$.

Proof:Let μ,β be an L-submodule of M and $\langle\alpha\rangle=\beta$. Since $\alpha\subseteq\beta$, $V(\beta)\subseteq V(\alpha)$.
Conversely Let $\mu\in V(\alpha)$. Then $\alpha\subseteq\mu$. This implies that $\langle\alpha\rangle\subseteq\langle\mu\rangle$ and so $\beta\subseteq\langle\mu\rangle$. Since μ is L-prime submodule of M, then $\beta\subseteq\mu$. Therefore $\mu\in V(\beta)$. This completes the proof.

Definition 3.14: $L(M)$ is distrubutive if $\eta,\nu,\mu\in L(M)$
i) $(\eta+\nu)\cap\mu=(\eta\cap\mu)+(\nu\cap\mu)$
ii) $(\eta\cap\nu)+\mu=(\eta+\mu)\cap(\nu+\mu)$.

Definition 3.15: Let Y be a subset of $L(M)$. $J(Y)$ is the intersection of all prime submodules which belongs to Y.

Definition 3.16: Let μ be a L-subset of M.
$p.rad(\mu)=\cap\{\eta|\ \mu\subseteq\eta,\ \eta \text{ is prime L-submodule}\}$ is called a radical of μ.

Definition 3.17: Let M be an R module.
$\aleph$=L-Spec(M)=$\{\mu|\mu \text{ is prime L-submodule of M}\}$ is called L-prime spectrum of M.

Definition 3.18: $V(x)=\{\eta\in\aleph|\ \eta(x)=1\}$ for all $x\in M$.

If we could not find any L-prime submodule ν which contains L-prime submodule μ, then $p.rad(\mu)=1_M$.

Theorem 3.19: Let $L(M)$ be L_T_p module with $p.rad(\mu)=\mu$. Then $L(M)$ is distrubutive.

Proof: $\eta,\nu,\mu\in L(M)$.

$$
\begin{aligned}
(\eta + \nu) \cap \mu &= p.rad((\eta + \nu) \cap \mu) \\
&= J(V((\eta + \nu) \cap \mu)) \\
&= J(V(\eta + \nu) \cup V(\mu)) \\
&= J(V(\eta \cup \nu) \cup V(\mu)) \\
&= J(V(\eta \cup \nu) \cup V(\mu)) \\
&= J((V(\eta) \cap V(\nu)) \cup V(\mu)) \\
&= J((V(\eta) \cup V(\mu)) \cap (V(\nu) \cup V(\mu))) \\
&= J((V(\eta \cap \mu) \cap (V(\nu \cap \mu))) \\
&= J((V(\eta \cap \mu) \cup (\nu \cap \mu))) \\
&= J((V(\eta \cap \mu) + (\nu \cap \mu))) \\
&= p.rad((\eta \cap \mu) + (\nu \cap \mu)) \\
&= ((\eta \cap \mu) + (\nu \cap \mu))
\end{aligned}
$$

Therefore $L(M)$ is distrubutive.

Theorem 3.20: Let $L(M)$ be L_T_p module, μ be L-submodule of M, and Y be a subset of L-Spec(M). Then

i) $V(\mu)$ is closed in L-Spec(M) and $J(Y)$ is an L-submodule of M equal to $p.rad(J(Y))$.

ii) $V(J(Y))$ is the closure of Y in L-Spec(M).

Proof: i) It is clear that $V(\mu)$ is closed in L-Spec(M) and $J(Y)$ is an L-submodule of M. Finally

$$
\begin{aligned}
p.rad(J(Y)) &= p.rad(\cap\{\eta \,|\, \eta \text{ is prime L-submodule in } Y\} \\
&= \cap\{\eta \,|\, \eta \in Y\} \qquad \textbf{ii} \\
&= J(Y).
\end{aligned}
$$

) Let $V(\mu)$ be closed set containing Y. That is $Y \subseteq V(\mu)$. Consequently, $V(J(Y)) \subseteq V(\mu)$. Since $Y \subseteq V(J(Y))$, then $V(J(Y))$ is the smallest closed subset of L-Spec(M). Thus $\overline{Y} = V(J(Y))$.

Proposition 3.21: Let Y be a subset of $L - Spec(M)$. Then

i) $J(V(J(Y))) = p.rad(J(Y)) = J(Y)$

ii) $V(J(V(Y))) = V(p.rad(Y)) = V(Y)$.

Corollary 3.22: For every family $\{Y_i\}_{i \in \Lambda}$ be a closed subsets of L-Spec(M), $J(\cap_{i \in \Lambda} Y_i) = p.rad(\sum J(Y_i))$.

Proof: Since $\{Y_i\}_{i \in \Lambda}$ be a closed subsets of L-Spec(M), then every $Y_i = V(Y_i)$ for each $i \in \Lambda$. So

$$
\begin{aligned}
J(\cap_{i \in \Lambda} Y_i) &= J(\cap_{i \in \Lambda} V(J(Y_i))) \\
&= J(V(\cup_{i \in \Lambda} J(Y_i))) \\
&= J(V(\sum J(Y_i))) \\
&= p.rad(\sum J(Y_i)) \text{ by proposition 2.20.}
\end{aligned}
$$

Corollary 3.23: Let $\mu \in Y \subseteq$ L-Spec(M). Then $\overline{\{\mu\}}$ the closure of μ is the set $V(\mu)$. We say that $\{\mu\}$ is the closed point of L-Spec(M) if and only if μ is maximal submodule in L-Spec(M).

Proof: Let $Y = \{\mu\}$ then $V(J(Y)) = \overline{\{\mu\}}$ by theorem 2.19 and $V(J(Y)) = V(J(\mu)) = V(\mu)$.

$\{\mu\}$ is the closed, that is $\{\mu\} = \overline{\{\mu\}}$. This implies $V(J(\mu)) = \mu$. So μ is maximal submodule in L-Spec(M).

Conversely now suppose that μ is maximal in L-Spec(M), then $V(\mu) = \{\mu\} = Y$ and so $Y = \overline{Y}$. Hence Y is closed.

Proposition 3.24: If $L(M)$ is L_T_p module and $\mu, \nu \in$ L-Spec(M). Then $\mu \in \overline{\{\nu\}}$ if and only if $\mu \subseteq \nu$.

Proposition 3.25: Let $L(M)$ be L_T_p module and L-Spec(M) is a T_0 space.

Proof: Let $\mu, \nu \in$ L-Spec(M) be two distinct points. We have two cases;

i) $\mu \in \overline{\{\nu\}} \Rightarrow \nu \subseteq \mu$. Since $\mu \neq \nu$, $\nu \notin \overline{\{\mu\}}$, then $\nu \in \overline{\{\mu\}}^c$. Therefore $\overline{\{\mu\}}^c$ is an open set which contains ν but not μ.

ii) $\nu \in \overline{\{\mu\}} \Rightarrow \mu \subseteq \nu$. Since $\mu \neq \nu$, $\mu \notin \overline{\{\nu\}}$, then $\mu \in \overline{\{\nu\}}^c$. Therefore $\overline{\{\nu\}}^c$ is an open set which contains μ but not ν.

Proposition 3.26: L-Spec(M) is a T_1 space if and only if for every $\mu \in$ L-Spec(M) is maximal.

Proof: Let $\forall \mu \in \text{L-Spec(M)}$ be maximal $\Rightarrow$ $\overline{\{\mu\}} = V(J(\mu)) = V(\mu) \Rightarrow$ Since μ is maximal, $\mu = \overline{\{\mu\}}$. This means that $\{\mu\}$ is the closed. Thus, L-Spec(M) is a T_1 space.

Conversely vice versa. (L-Spec(M) is a T_1 space. $\mu = \overline{\{\mu\}}$.)

Definition 3.27: $\aleph(\mu) = \aleph - V(\mu)$ is called the complement of $V(\mu)$ in L-Spec(M).

Proposition 3.28: Let Y be a subset of $\aleph$ and $\overline{Y}$ denote the closure of Y. Then $\overline{Y} \subseteq V(1_N)$ where $N = \bigcap_{\beta \in Y} \beta_*$.

Proof: Clearly $1_N(x) = 1 \Leftrightarrow \beta(x) = 1 \ \ \forall \beta \in Y.$ Therefore if $\eta \in Y,$ then $1_N \subseteq \eta$ and consequently $\eta \in V(1_N).$ Therefore the closed set $V(1_N)$ containing Y, contains $\overline{Y}$.

Definition 3.29: Let x be element of M, $V(x) = \{\mu \in \text{L-Spec(M)} \mid \ \mu(x)=1\}.$

Theorem 3.30: Let $f : M \to M'$ module epimorphism and $\overline{f} : \text{L-Spec(M}') \to \text{L-Spec(M)}$ be a function defined by $\overline{f}(\eta) = f^{-1}(\eta).$ Then $\overline{f}$ is continuous, injective and $\text{L-Spec(M}')$ is homeomorphic to the closed subset $V(1_{\ker f}).$ If f is an isomorphism, then $\overline{f}$ is homeomorphism.

Proof: Let μ and ν be an L-prime submodules of M'. Since f is surjective, $\overline{f}(\mu) = \overline{f}(\nu) \Rightarrow f^{-1}(\mu)(m) = f^{-1}(\nu)(m) \ \ \forall m \in M.$ Note that; R. Ameri([1]) theorem 3.14 $f^{-1}(\nu)$ is prime submodule. Then by definition, $\mu(\overline{f}(m)) = \nu(\overline{f}(m)) \ \forall m \in M.$ This implies that, $\mu = \nu.$ So $\overline{f}$ is injective.

If $V(m)$ is a basic closed set in L-Spec(M), then $\overline{f}^{-1}(V(m))$ is a basic closed set in L-Spec(N). Because

$$\begin{aligned}
\overline{f}^{-1}(V(m)) &= \left\{ \mu \in \text{L-Spec(M}') \,\middle|\, \overline{f}(\mu)(m) = 1 \right\} \\
&= \left\{ \mu \in \text{L-Spec(M}') \,\middle|\, f^{-1}(\mu)(m) = 1 \right\} \\
&= \left\{ \mu \in \text{L-Spec(M}') \,\middle|\, \mu(\overline{f}(m)) = 1 \right\} \\
&= V'(f(m)_1).
\end{aligned}$$

Hence $\overline{f}$ is continuous. Let μ and ν be an L-prime submodules of M'. $f^{-1}(\mu)$ is constant on $Kerf$. Indeed $f^{-1}(\mu)(m) = \mu(f(m)) = \mu(0) = 1$ for all $m \in Kerf$. Then $f^{-1}(\mu) \in V(1_{Kerf})$. If $\nu \in V(1_{Kerf})$ then $\nu \in \text{L-Spec(M)}$ constant on $Kerf$. Since f is an isomorphism it follows from theorem 3.5.11 that $f(\nu) \in \text{L-Spec(M}')$. This defines a function $\overline{g} : V(1_{Kerf}) \to \text{L-Spec(M}')$ where $\overline{g}(\nu) = f(\nu).$ Clearly $\overline{g} = \overline{f}^{-1}$. To prove the continuity of $\overline{g}$, $V'(f(m))$ be a closed set in $\text{L-Spec(M}')$. Then

$$\begin{aligned}
\overline{g}^{-1}(V'(f(m))) &= \overline{f}(V'(f(m)_1) \\
&= \left\{ \overline{f}(\nu) \,\middle|\, \nu(f(m)) = 1 \right\} \\
&= \left\{ f^{-1}(\nu) \,\middle|\, f^{-1}(\nu)(m) = 1 \right\} \\
&= V(m) \cap V(1_{Kerf}).
\end{aligned}$$

which is closed subset of $\text{L-Spec(M)}.$

Finally, suppose that f is an isomorphism. Then $Kerf = \{0\}$ and $V(1_{Kerf}) = V(0_M) = \text{L-Spec(M)}.$

Corollary 3.31: If N is submodule M such that $N \subseteq \bigcap_{\eta \in \aleph} \eta_*$, then L-Spec(M) and $\text{L-Spec}(M/N)$ are homeomorphic.

Proof: Let f be a natural homomorphism of M onto M/N. Then each $\eta \in \aleph$ is constant on $Kerf$.

Theorem 3.32: Let M be a Noetherian $R-$module. The topological space $(\text{L-Spec(M)} = \aleph, T)$ is compact.

Proof: If $L - \{1\}$ has no prime element, then $\aleph = \varnothing$ proof is complete. Suppose that $L - \{1\}$ have prime

elements and a be a prime element of $L-\{1\}$. Let $\{\aleph((m_i)_t)\mid i\in\Lambda, t\in K\subseteq L-\{0\}\}$ be a cover of $\aleph$ by its basic open sets. Let $\vee\{t\mid t\in K\}=b.$ Then $\{\aleph((m_i)_b)\mid i\in\Lambda\}$ also cover $\aleph$. Therefore $\aleph=\cup\{\aleph((m_i)_b)\mid i\in\Lambda\}=\aleph(\bigcup_{i\in\Lambda}(m_i)_b)).$ Hence $V(\bigcup_{i\in\Lambda}(m_i)_b))=\varnothing.$ Let N be a prime submodule of M. Define $\eta: M\to L,$ defined by $\eta(m)=\begin{cases}1 & \text{if } m\in N\\ 0 & \text{otherwise}\end{cases}$. So η is L-prime submodule then $\eta\in\{\aleph((m_i)_b)\mid i\in\Lambda\}$ and so $\bigcup_{i\in\Lambda}(m_i)_b\not\subset\eta.$ This implies that $\exists i\in\Lambda$ s.t. $(m_i)_b\not\subset\eta.$ Hence either $b>\eta(m_i)$ or b and $\eta(m_i)$ are noncomparable. In either case $m_i\notin N$. Therefore $\{m_i\mid i\in\Lambda\}$ is not contained in any proper submodule of M. $\langle\{m_1,m_2,...,m_k\}\rangle=M$ since M is Noetherian. Suppose $V(\cup_{i=1}^k(m_i)_b)\neq\varnothing$ let $\beta\in V(\cup_{i=1}^k(m_i)_b).$ Then $\cup_{i=1}^k(m_i)_b\subseteq\beta\Rightarrow\beta(m_i)\geq b$ for all $i=1,...k$. Suppose $\exists i,$ s.t. $\beta(m_i)\neq 1.$ Now $\beta(m)=\beta(\sum_{i=1}^k r_im_i)\geq\wedge_{i=1}^k\beta(m_i)\geq\wedge_{i=1}^k b=b.$ Now $\beta\in\aleph$ and so $\beta\in\{\aleph((m_i)_b)\mid i\in\Lambda\}.$ Therefore $\exists i\in\Lambda$ s.t. $\beta\in\aleph((m_j)_b).$ Then $(m_j)_b\not\subset\beta.$ Thus either $b>\beta(m_j)$ or b and $\beta(m_j)$ are noncomparable. However, $\beta(m_j)=\beta(\sum_{i=1}^k r_im_i)\geq\wedge_{i=1}^k\beta(m_i)>b.$ This is a contradiction. Hence $\beta(m_j)=1.$ This implies that $M=\beta_*.$ This is a contradiction. Therefore $V(\cup_{i=1}^k(m_i)_b)=\varnothing$. Consequently, $\{\aleph((m_i)_b)\mid i\in 1,...,k\}$ is a subcover of $\aleph$. This completes the proof.

Definition 3.33: A topological space T is called irreducible if for any decomposition $T=A_1\cup A_2$ where A_1, A_2 are closed subsets of T, then $T=A_1$ or $T=A_2$.

Theorem 3.34: Let Y be a closed subset of $\aleph$. If $J(Y)$ is L-prime submodule of M, then Y is irreducible.

Proof: Suppose that $J(Y)$ is L-prime submodule of M. Suppose $Y=Y_1\cup Y_2$, where Y_1,Y_2 are closed subsets of $\aleph$. Then $J(Y)\subseteq J(Y_1)$ and $J(Y)\subseteq J(Y_2)$. Also $J(Y)=J(Y_1\cup Y_2)=J(Y_1)\cap J(Y_2)$. Then $J(Y_1)oJ(Y_2)\subseteq J(Y_1)\cap J(Y_2)\subseteq J(Y).$ Since $J(Y)$ is L-prime submodule of M, then $J(Y_1)\subseteq J(Y)$ or $J(Y_2)\in(J(Y):1_M)$. If $J(Y_1)\subseteq J(Y)$, the proof is complete. If $J(Y_2)\in(J(Y):1_M)$, then $J(Y_2)\subseteq J(Y)$. This completes the proof.

Definition 3.35: Let Y be a closed subset of $\aleph$ and $\eta\in Y$. Then η is called a generic point of Y if $Y=\overline{\{\eta\}}$, the closure of $\eta.$

We know that ıf $Y\subseteq\aleph$ where $\aleph$ is a topological space, then Y is irreducible if and only if $\overline{Y}$ is irreducible.

Theorem 3.36: Let Y be a closed subset of $\aleph$ and $\eta\in Y$ be a generic point of Y Then Y is irreducible.

Proof: Since η is a generic point of Y, then $Y=\overline{\{\eta\}}$. Since $\{\eta\}$ is irreducible, $\overline{\{\eta\}}$ is irreducible. Therefore Y is irreducible

4 Conclusion

Letting $L-\xi(M)=\{V(\mu.1_M)\mid\mu\in LI(R)\}$. It is easy to prove that this set always induces a topology T on $L-spec(M)$. R. Ameri and R Mahjoob showed that $L-\xi(M)$ induces a topology which is called Zariski topology if and only if M is a top module. By following them we show that a topology T on $L-spec(M)$ exists if and only if $L(M)$ is L_T_p module. Under this condition topology T on $L-spec(M)$ is T_0 space. Behind the existing T_1 space, if M is a Noetherian $R-$module, then the topological space $(\text{L-Spec(M)}=\aleph,T)$ is compact.

Acknowledgement

"Bayram Ali Ersoy's work was supported by the Scientific and Technological Research Council of Turkey (TUBITAK); Dan Ralescu's work was supported by a Taft Travel for Research Grant ".

References

[1] C.P. Lu, Prime submodules of modules, *Comm. Math. Univ*. 33 (1987) 61-69.

[2] C.P. Lu, The Zariski topology on the spectrum of a modules, *Houston Journal of Mathematics*. 25 (3) (1999) 417-432.

[3] R.L. Mccasland, M.E. Moore, P.F. Smith, On the spectrum of modules over a commutative ring. *Comm. Algebra*, 25 (1) (1997) 417-432.

[4] L.A. Zadeh, Fuzzy sets, *İnform and Control* 8 (1965) 338-353.

[5] J.A. Goguen, L-fuzzy Sets, *J. Math. Appl.* 18 (1967) 145-174.

[6]A. Rosenfeld, Fuzzy groups, *J. Math. Anal. Appl*. 35 (1971) 512-517.

[7] C.V. Negoita, D.A. Ralescu, Application of fuzzy systems analysis, Birkhauser, Basel, 1975.

[8] F.Z. Pan, Fuzzy finitely generated modules, *Fuzzy Sets and Systems*, 21 (1987) 105-113.

[9] F.I. Sidky, On radical of fuzzy submodules and primary fuzzy submodules, *Fuzzy Sets and Systems* 119 (2001) 419-425.

[10] V.N. Dixit, R. Kumar, N. Ajmal, Fuzzy ideals and fuzzy prime ideals of a ring, *Fuzzy Sets and Systems*, 44 (1991) 127-138.

[11] T.K. Mukherjee, and M.K. Sen, On fuzzy ideals of a ring I, *Fuzzy sets and systems* 21 (1987) 99-104.

[12] S.K. Bhambri, R. Kumar, P. Kumar, Fuzzy prime submodules and radical of a fuzzy submodules, *Bull. Calcutta Math. Soc*. 87 (1993) 163-168.

[13] H. Hadji-Abadi, M.M. Zahedi, Some results on fuzzy prime spectrum of a ring, *Fuzzy Sets and Systems*, 77 (1996) 235-240.

[14] R. Kumar, Fuzzy prime spectrum of a ring, *Fuzzy Sets and Systems*, 46 (1992) 147-154.

[15] R. Kumar, J.K. Kohli, Fuzzy prime spectrum of a ring II, *Fuzzy Sets and Systems*, 59 (1993) 223-230.

[16] H.V. Kumbhojkar, Spectrum of prime fuzzy ideals, *Fuzzy Sets and Systems*, 62 (1994) 101-109.

[17] H.V. Kumbhojkar, Some comments on spectrum of prime fuzzy ideals of a ring, *Fuzzy Sets and Systems*, 88 (1997) 109-114.

[18] R. Ameri, R. Mahjoob, Spectrum of prime L-submodules, *Fuzzy Sets and Systems*, 5122 (2007) .

[19] J.N. Mordeson, D.S. Malik, Fuzzy Commutative Algebra, *Worlf Scientific Publishing*, Singapore, 1998.

[20] D.S. Malik, J.N. Mordeson, L-prime spectrum of a ring, *Fuzzy Information Processing Society NAFIPS '97. 1997 Annual Meeting of the North American*, (1997) 273-278.

Author Index

Abbasbandy, Saeid 642
Addabbo, Tindara 1276
Adlassnig, Klaus-Peter 1085
Afanasyeva, Tatyana 792
Agliardi, Elettra 1009
Aguiló, Isabel 333
Ahmad, Muhammad Zaini 1311
Aires, Lara 1358
Akbarzadeh-T, Mohammad-R. 381, 396, 443
Alboody, Ahed 1079
Albusac, Javier 375
Alcalá, Rafael 1146, 1740
Alcalde, Cristina 58
Alizadeh, Meysam 1745
Almeida, Cesar Guilherme 271
Almeida, Rui Jorge 1785, 1803, 1809
Alonso, José M. 119, 125
Alouane, Basma 420
Anderson, Derek 282, 288
Andreasen, Troels 1618
Angelov, Plamen 931
Angilella, Silvia 1194
Antonelli, Michela 1033
Araujo, Joao 1358
Ardehali, Iman Mohammadi 1142
Arenas, Mar 1486
Armengol, Eva 891
Asmuss, Svetlana 1735
Asmussen, Katrin 1480
Assaghir, Zainab 707
Atanassov, Krassimir T. 1797
Avazbeigi, Milad 1142, 1520
Baczynski, Michal 277
Badamchizadeh, Mohammad Ali 1270
Baek, Gyeongdong 1183
Baetens, Jan 201
Baez-Sanchez, A. David 1353
Ballagi, Áron 1468
Ban, Adrian 798
Barranco, Carlos 903, 957
Barrenechea, Edurne 478, 963, 1164
Barros, Laécio C. 240, 265
Baruch, Ieroham 207, 460
Bassanezi, Rodney C. 265
Bede, Barnabas 730, 760
Bedregal, Benjamin 1
Begue, Agnes 1496
Behounek, Libor 449
Beliakov, Gleb 1370
Benferhat, Salem 1223, 1712
Benítez, José Manuel 1740
Benítez, Alicia D. 1751
Benrejeb, Mohamed 1655
Bereketli, İlke 1439
Berlanga, Francisco José 42
Berlik, Stefan 1246
Berthold, Michael 944
Bilbao, Amelia 1486
Biletska, Krystyna 849
Billaudel, Patrice 1015
Bisserier, Amory 1305
Bobillo, Fernando 189
Bocklisch, Steffen F. 885, 974
Bodenhofer, Ulrich 1376
Bonissone, Piero 1553
Bordogna, Gloria 867, 915
Borgelt, Christian 968
Bosc, Patrick 259
Bosteels, Klaas 25
Botzheim, János 1667, 1833
Bou, Félix 1541
Bouchon-Meunier, Bernadette 1625
Boukezzoula, Reda 1305, 1851
Bracho, Carlos Campos 1875
Braga, Antonio 1490
Brando, Carmen 259
Brito, Julio 1547
Brockmann, Werner 466
Bronselaer, Antoon 663
Bugarín, Alberto 1559
Bulskov, Henrik 1618
Burusco, Ana 58
Bustince, Humberto 300, 478, 754, 963
Butkiewicz, Bohdan 1188
Cabrerizo, Francisco Javier 13
Cadenas, José M. 669, 1553
Caetano, Mario 1051
Caldas Pinto, João 1857
Calvo, Tomasa 1370
Camargo, Heloisa 742
Campaña, Jesús R. 903
Campos, Clara M. 1706
Canós, Lourdes 1706
Canós, María J. 1706
Cardin, Marta 1637
Carrara, Paola 915
Carrillo, Juan V. 669
Carvalho, João Paulo 178, 1821
Casasnovas, Jaume 647
Casillas, Jorge 1631, 1751
Castiello, Ciro 339
Castillo, Oscar 809, 815, 1382
Castro, Jose Jesús 375
Castro, Juan Ramon 774, 1382
Celikyilmaz, Asli 1388
Ceravolo, Paolo 1169
Chabchoub, Habib 213
Chalco-Cano, Yurilev 231, 703
Chamorro-Martinez, Jesus 1763
Chen, Guoqing 1462
Chiclana, Francisco 980
Cho, Jaewoo 1183
Choi, Moon-Ho 1183
Chountas, Panagiotis 1797
Cintra, Marcos 742
Cintula, Petr 519
Colubi, Ana 1428, 1433
Combarro, Elias F. 64

Cordón, Oscar 195, 1514
Cornelis, Chris 408, 925
Coroianu, Lucian 798
Couceiro, Miguel 525, 531
Coupland, Simon 288, 606
Couso, Ines 803
Couto, Pedro 963, 1164
Cruz, Carlos 1158, 1398
Damas, Sergio 195
Damiani, Ernesto 1169
Daňková, Martina 585, 693
De Baets, Bernard 201, 780, 786, 1311
De Cock, Martine 1347
de la Fuente, David 544
de Lima, Beatriz S. L. P. 768
De Maeyer, Philippe 1480
De Tre, Guy 663, 1404, 1480
Dehghan, F. 381
Del Jesus, Maria Jose 42
Delafontaine, Matthias 1480
Delgado, Miguel 253, 1211
Dellunde, Pilar 414
Denoeux, Thierry 849
D'Errico, Liliana 1136
Deschrijver, Glad 184
di Martino, Ferdinando 53
Dias, Marina R. B. 240
Díaz, Susana 780, 786
Díaz, Ramón Andrés 1553
Dieulot, Jean-Yves 1068
Dimuro, Graçaliz 1
Divari, Maria 88
Djouadi, Yassine 420
Domech, Jesus 1102
Dourado, Antonio 1358
Dragicevic, Suzana 1404
Driankov, Dimiter 7
Dubois, Didier 1021
Ducange, Pietro 1033
Dujmovic, Jozo 1404
Dvorak, Antonin 357, 507
Dyba, Martin 1422
Eckhardt, Alan 938
El Hajj, Mahmoud 1496
Elfelly, Nesrine 1068
Elmorsy, Shamiaa 825
Endo, Yasunori 861, 1152
Entani, Tomoe 155
Eriz, Mariano 173
Erman, Maria 763
Erol Genevois, Müjde 1126, 1439
Ersoy, Bayram Ali .....
Espinilla, Macarena 108
Esteva, Francesc 891, 1541
Etminani, Kobra 396
Evsukoff, Alexandre G. 768
Facchinetti, Gisella 1276
Faigle, Ulrich 161
Fan, Tuan-Fang 294
Fanelli, Anna M. 339
Fazel Zarandi, Mohammad Hossein 30, 1520
Felix, Rudolf 1004
Fenza, Giuseppe 1315
Fernandez, Javier 300, 478
Fernández, Alberto 42
Fernandez-Escribano, Gerardo 19
Figueiredo, Karla 1096
Fober, Thomas 1299
Fodor, Janos 653
Földesi, Péter 1667
Fonte, Cidália C. 1051
Frélicot, Carl 855, 1200
Fuchs, Martin 345
Fuentes-Gonzalez, Ramon 58
Fujimoto, Katsushige 636
Fukuda, Shinji 1258
Gabr, Mahmoud 825
Gabriel, Thomas 944
Gacto, María José 1146
Gągolewski, Marek 1691
Gál, László 1508, 1833
Galar, Mikel 754, 1164
Galichet, Sylvie 768, 1305, 1851
Galvan-Guerra, Rosalba 460
Garcia-Alegre, Maria C. 1729
Garcia-Cascales, M.Socorro 675
García-Cerdaña, Àngel 891
García-Honrado, Itziar 897
García-Lapresta, José Luis 1527
Garmendia, Luis 837
Garrido, Antonio 1063
Garrido, Carmen 669, 1553
Gedeon, Tom 1468
Georgescu, Vasile 1679, 1791
Gerla, Brunella 713
Ghaemi, Sehraneh 1270
Ghisalberti, Giorgio 867
Gil, M. Ángeles 1428
Gil-Lafuente, Anna M. 1661
Giove, Silvio 1637
Girardin, Philippe 707
Godo, Lluis 1541
Gomide, Fernando 1114
Goncalves, Paulo 1857
Gonçalves, Luisa M. S. 1051
Gongora, Mario 697
González del Campo, Ramón 837
González- Hidalgo, Manuel 630
González-Rodríguez, Gil 1428, 1433
Gottwald, Siegfried 386
Grabisch, Michel 76, 131, 161, 363
Graves, Daniel 1108
Greco, Salvatore 1194
Greenfield, Sarah 980
Grzegorzewski, Przemysław 1531, 1535, 1691
Guadarrama, Sergio 195, 897, 1410
Guerra, Thierry M. 736
Guerra, Carlos 754
Guerra, Maria Letizia 1009
Guillaume, Serge 1496
Guinea, Domingo Miguel 1729
Guinea, Domingo 1729
Gürbüz, Tuncay 1126, 1439
Hadjali, Allel 259, 501
Hagras, Hani 537
Halouani, Nesrin 213

Hamasuna, Yukihiro....1152
Hamilton, Sam....70
Hamilton, Wendy....70
Hamrawi, Hussam....606
Haned, Faiza....1712
Hartert, Laurent....1015
Hassan, Yasser....825
Hasuike, Takashi....1120
Heinonen, Petri....1685
Helmer, Sven....957
Hempel, Arne-Jens....885
Herbst, Gernot....974
Herrera, Francisco....42, 1027, 1146, 1740
Herrera-Viedma, Enrique....13
Hochreiter, Sepp....1376
Holcapek, Michal....357, 507, 1565
Holecek, Pavel....681
Honda, Aoi....636
Huellermeier, Eyke....1294, 1299
Hurtado Martin, German....408
Hüsam, Philipp....386
Ibáñez, Oscar....195
Ida, Masaaki....1343
Iliev, Boyko....7
Imai, Hideyuki....1612
Inglada, Jordi....1079
Inuiguchi, Masahiro....873
Ionescu, Mihaela....1376
Ishibuchi, Hisao....1649, 1839
Ishii, Hiroaki....1120
Iskala, Hannu....1685
Ivorra, Carlos....669
Jacas, Joan....1696
Jafelice, Rosana M. 265, 271
Jaime-Castillo, Sergio....903, 1063
James, Simon....1370
Janssen, Jeroen....1347
Jassbi, Javad....1217
Jaudoin, Hélène....259
Jayaram, Balasubramaniam....277, 432
Jeon, Tae-Ryong....1183
Jiménez, Luis....19, 375
Jiménez, Jorge....1073
Jiménez, Mariano....1486
Jimenez de la C., Guillermo....1815
Jiménez-Gamero, María Dolores....703
John, Robert....697, 980
Júlio, Eduardo N. B. S.....1051
Jurio, Aranzazu....831, 1164
Juuso, Esko....1685, 1779
Kacprzyk, Janusz....1288, 1321, 1416
Kadmiry, Bourhane....7
Kanzawa, Yuchi....861
Kao, Bo-Ruei....369
Kasperski, Adam....390, 426
Kawasaki, Yuichi....618
Kaymak, Uzay....351, 1673, 1785, 1803, 1809
Keller, James....282, 288
Kerre, Etienne....25, 167, 1474
Khaleghei Ghoshe Balagh, Akram....1745
Khanmohammadi, Sohrab....1217, 1270
Kielan, Krzysztof....1717
Kim, Sungshin....1183
Kim, Eunjin....1773
Kimala, Vilany....92
Kindermann, Stefan....950
Kiseliova, Tatiana....1577
Klement, Erich Peter....114
Kobayashi, Ichiro....234
Koczy, Laszlo T.....1468, 1508, 1667, 1833
Kohout, Ladislav J. 1773
Kokuhuda, Mitsuru....657
Kolesárová, Anna....114
Kop, Yeşim....1450
Koprowski, Robert....1444
Kosheleva, Olga....438, 455, 1394
Krajci, Stanislav....1252
Kreinovich, Vladik....1264
Kridlo, Ondrej....1252
Kroupa, Tomas....713
Kruse, Rudolf....1177
Kubota, Naoyuki....1869
Kukkurainen, Paavo....821
Kupka, Jiri....589
Kwiatkowska, Mila....1717
Labreuche, Christophe....363, 567
Lai, K. Robert....369
Lamara, Hakim....736
Lamata, Teresa....669, 675
Lang, Tiziana....1276
Lazzari, Luisa Lucila....173
Lazzerini, Beatrice....1033
Le Capitaine, Hoel....855, 1200
Legind Larsen, Henrik....402
Leida, Marcello....1169
Li, Zhan....1583
Liau, Churn-Jung....294
Liern, Vicente....669, 1706
Liétard, Ludovic....513
Lin, Menq-Wen....369
Liu, Jun....213
Liu, Duen-Ren....294
Llamazares, Bonifacio....1527
Lloyd, Christopher D.....351
Lodwick, Weldon....321
Loia, Vincenzo....53, 1315
Lopez, Miguel....809
López López, Lorenzo Manuel....375
López-Díaz, Miguel....1327
Lopez-Molina, Carlos....831
Loreti, Michele....1136
Lotfi, Ehsan....381
Lovassy, Rita....1508
Lovric, Milan....1803
Lu, Jie....555
Luengo, Julián....1027
Lughofer, Edwin....931, 950
Luke, Robert....282, 288
Ma, Jun....555
Machado, Liliane....102
Magdalena, Luis....119, 125
Mannelin, Marjo....1685
Marcelloni, Francesco....1033
Mariaca Gaspar, Carlos Román....207
Marichal, Jean-Luc....525, 531
Marin, Nicolás....1039

Márquez, Antonio Angel 1045
Márquez, Francisco Alfredo 1045
Marsala, Christophe 1625
Martel, Jean-Marc 213
Martin, Trevor 742, 909, 986
Martin, David 1729
Martín, Javier 1723
Martínez, Luis 108, 213
Martínez, Raquel 1553
Martínez-Jimenez, Pedro M. 1763
Martínez-Moreno, Juan 1702
Massanet, Sebastià 573
Masson, Marie-Hélène 849
Mastroleo, Giovanni 1276
Matarazzo, Benedetto 1194
Mauris, Gilles 1589
Mayag, Brice 363
Mayor, Gaspar 1723
Medina, Juan Miguel 903, 1063
Medina, Jesús 1091
Mehdizadeh, Esmaeil 1768
Mélange, Tom 167, 1474
Melin, Patricia 774, 809, 815, 1382
Melo-Pinto, Pedro 831, 963
Mencar, Corrado 339
Mendoza, Olivia 774, 1382
Merigó, José M. 1364, 1661
Mesiar, Radko 114, 300
Meyer, João Frederico C. A. 271
Michalik, Krzysztof 1717
Midenet, Sophie 849
Milea, Viorel 1785
Miller, Simon 697
Miranda, Pedro 64, 76
Mir-Torres, Arnau 630
Miyamoto, Sadaaki 612, 618, 861, 1152
Miyasaki, Rodolfo 1757
Mockor, Jiri 1282
Mohammed, Abbas 763
Mohd Sharef, Nurfadhlina 909
Mokhtari, Amine 501
Mokhtari, Aïcha 1712
Molina, Carlos 1039, 1571
Monreal, Amadeu 1696
Montero, Javier 300
Montes, Susana 780, 786, 1073
Montseny, Eduard 1863
Moraes, Ronei 102
Moraga, Claudio 306, 687, 879
Moreno, Juan Manuel 13
Moreno, José A. 1547, 1706
Moreno-Garcia, Juan 19
Mouliá, Patricia Inés 173
Mucientes, Manuel 1559
Muniz, Cyro 1096
Muñoz Hernández, Susana 1827
Nachtegael, Mike 167, 1474
Naessens, Helga 408
Napoli, Amedeo 707
Narukawa, Yasuo 624, 1132
Nasiri, Maryam 1246
Navara, Mirko 550
Neumann, Bastian 466
Nguyen, Hung 1264
Nieradka, Grzegorz 1188
Nin, Jordi 489
Niskanen, Vesa 561
Noguera, Carles 519
Nojima, Yusuke 1649, 1839
Novak, Vilem 483, 1422
Nunes-Silva, Geraldo 231
Ohnishi, Shin-ichi 1612
Ojeda-Aciego, Manuel 1091
Okumura, Naoko 234
Onisawa, Takehisa 1595
Orduna, Raul 300
Osone, Keisuke 1595
Pablos Ceruelo, Victor 1827
Pacheco, Marco 1102
Pagani, Marco 867, 915
Pagola, Miguel 478, 831, 963, 1164
Pais, Tiago 719
Palacios, Ana 803
Palm, Rainer 7
Pampalk, Elias 25
Pardo, María José 544
Pasternak-Winiarska, Karolina 1531
Paternain, Daniel 754, 831
Patrascu, Vasile 48
Pavlacka, Ondrej 495
Pavliska, Viktor 483
Pedro, Joao 178
Pedrycz, Witold 748, 1108, 1490
Pękala, Barbara 1206
Pelta, David 472
Peña, Teresa 1527
Pepe, Monica 915
Peregrín, Antonio 1045
Pereira, Martín 992
Perfilieva, Irina 483, 585, 1601
Petrík, Milan 595
Pinchuck, Andrew 312
Pires, Joao 178
Pivert, Olivier 259, 501
Pizzi, Nick 748
Pons, Olga 1039
Popova, Viara 697
Pourreza, Hamid Reza 381
Prade, Henri 420, 1021
Prados-Suarez, Belén 1863
Psaila, Giuseppe 867
Qiang, Yi 1480
Queiroz, Francisco 1490
Quirin, Arnaud 1514
Rada, Roy 1745
Raeeji Yanehsari, Noorali 396
Rakus-Andersson, Elisabeth 225, 763
Ralescu, Dan .....
Ramík, Jaroslav 98
Rampini, Anna 915
Recasens, Jordi 837, 1696
Reformat, Marek 1583
Reiser, Renata 1
Renčová, Magdaléna 1882
Renedo, Eloy 879, 897
Ribeiro, Rita 719

Ribeiro, Nicholas....1102
Rickard, John T....70
Riečan, Beloslav....1880
Riera, J. Vicente....647
Riera, Bernard....1845
Rifqi, Maria....1294, 1625
Robles, Ignacio....1740
Rocacher, Daniel....513
Rodriguez, Rosa Mª....108
Rodriguez, Ricardo Oscar....1541
Rodriguez-Benitez, Luis....19
Rodriguez-Diaz, Antonio....1382
Rodríguez-Fdez, Ismael....1559
Rodríguez-Muñiz, Luis J.....1327
Rodriguez-Uria, Maria Victoria....1486
Rogova, Ermir....1797
Roldán, Antonio....1702
Roldán, Concepcion....1702
Román-Flores, Heriberto....231, 703
Romaní, Santiago....1863
Rommelfanger, Heinrich....317
Ros, María....1211
Rosemann, Nils....466
Ruan, Da....555
Ruano, Antonio....1833
Rudas, Imre J....653, 760
Ruiz, M. Dolores....253
Ruiz-Aguilera, Daniel....630
Runkler, Thomas A.....149
Ruzha, Vechislav....1735
Ruz-Hernandez, Jose A.....1815
Saad, Emad....825
Saeys, Yvan....925
Sahba, Farhang....1164
Sakly, Anis....1655
Salazar M., Ruben....1815
Salski, Arkadiusz....725
Sanchez, Daniel....253, 1571
Sanchez, Luciano....803
Santamaría, José....195
Santos, Jorge....1240
Santos, Dolores....1706
Sanz, Josean....478, 754
Sanz, Veronica....1502
Sarkoci, Peter....595
Sawazaki, Keisuke....618
Sayed Mouchaweh, Moamar....1015, 1845
Schjær-Jacobsen, Hans....1057
Schockaert, Steven....1347
Schwarzbauer, Karin....1376
Sedes, Florence....1079
Seising, Rudolf....1332, 1338
Senatore, Sabrina....1315
Serrano, Jose-Maria....1571
Šešelja, Branimir....1073
Sessa, Salvatore....53
Seyyed Esfahani, Mir Mehdi....1745
Shelomov, Evgen....1815
Shen, Yun....909, 986
Silva, Ricardo C.....1158, 1234, 1398
Sobrevilla, Pilar....1863
Sobrino, Alejandro....306, 327, 992
Solana-Cipres, Cayetano....19
Sorini, Laerte....600, 1643
Sorsa, Aki....1685
Sostak, Alexander....1606
Soto-Hidalgo, Jose M.....1763
Sousa, João M. C.....1673, 1785, 1857
Spronk, Jaap....1803
Stefanini, Luciano....579, 600, 730, 1009, 1535
Steinbrecher, Matthias....1177
Stepnicka, Martin....432, 483, 693
Stichelbaut, Birger....1480
Stone, Erik....282
Straccia, Umberto....189
Strass, Hannes....1827
Stupňanová, Andrea....114
Sun, Feng....82
Suñer, Jaume....333
Sussner, Peter....167, 1757
Szmidt, Eulalia....1416
Szwarcman, Dilza....1102
Tabia, Karim....1223
Takahagi, Eiichiro....843
Talasova, Jana....495, 681
Tang, Xiaohui....1462
Tanscheit, Ricardo....1096, 1102
Tatari, Farzaneh....443
Tavakkoli-Moghaddam, Reza....1768
Tepavčević, Andreja....1073
Teran, Pedro.....
Tercan, A. Erhan....351
Thiel, Kilian....944
Thipwiwatpotjana, Phantipa....321
Tinati, Mohammad Ali....1270
Tizhoosh, Hamid....36, 1164
Tomanova, Iva....483
Tome, Jose....1821
Torra, Vicenç....489, 624, 1132
Torrens, Joan....333, 573, 630
Trawinski, Krzysztof....1514
Trillas, Enric....306, 687, 879, 897
Trivino, Gracian....327
Trutschnig, Wolfgang....1428, 1433
Turcan, Matej....1565
Turksen, I. Burhan....30, 1388, 1520
Tutmez, Bulent....351
Uljane, Ingrida....921
Ulukan, H. Ziya....1450
Unehara, Muneyuki....92
Unusov, Tagir....792
Vachkov, Gancho....1456
Valdez, Fevrier....815
Valle, Marcos Eduardo....247, 1757
Vallejo, David....375
Valls, Aida....489
Van de Weghe, Nico....1480
Vanekova, Veronika....998
Vasconcelos, Heraldo L....271
Vavrickova, Lenka....483
Vellasco, Marley....1096, 1102
Verbiest, Nele....925
Verdegay, José Luis....1158, 1547
Vermeir, Dirk....1347
Vermeiren, Laurent....736
Vertan, Constantin....1229

Vetterlein, Thomas....1085
Vieira, Vinicius da F. 768
Vieira, Susana....1673, 1785
Vila, Amparo 1039, 1211, 1571
Villaverde, Karen 455
Vivona, Doretta 88
Vojtas, Peter 938, 998
Wagner, Christian 537
Walker, Carol....219
Walker, Elbert....219
Walter, Igor....1114
Wang, Xue-ping 82
Wang, Xiaomeng....968
Wei, Qiang....1462
Wilbik, Anna 1321
Wrobel, Zygmunt 1444
Yager, Ronald R. 219, 472, 1583
Yaghoubi, Mona....381
Yamada, Koichi 92
Yamakami, Akebo 1158, 1234, 1398
Yamanoi, Takahiro....1612
Yarushkina, Nadezhda 792
Yasunobu, Seiji....657
Yeh, Chi-Tsuen....143
Yorita, Akihiro 1869
Yoshida, Yuji 137
Zadrozny, Slawomir 1288
Zahra, Basma 1655
Zarinbal, Marzie 30
Zeddigha, Ismahane 1712
Zhang, Guangquan 555
Zielinski, Pawel....390, 426



ISBN: 978-989-95079-6-8

ANA Aeroportos de Portugal
TAP PORTUGAL
FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR
FUNDAÇÃO LUSO-AMERICANA
Lisboa
FUNDAÇÃO CALOUSTE GULBENKIAN