Generalized Atanassov's Intuitionistic Fuzzy Index. Construction Method

Edurne Barrenechea Humberto Bustince Miguel Pagola Javier Fernandez Josean Sanz

Dpt. Automática y Computación, Universidad Pública de Navarra
Campus de Arrosadía, s/n
31006 Pamplona (Spain)
Email: {edurne.barrenechea,bustince,miguel.pagola,fcojavier.fernandez,joseantonio.sanz}@unavarra.es

Abstract— In this work we introduce the concept of *Generalized Atanassov's Intuitionistic Fuzzy Index*. We characterize it in terms of fuzzy implication operators and we propose a construction method with automorphisms. Finally, we study some special properties of the generalized Atanassov's intuitionistic fuzzy index.

Keywords— Atanassov's Intuitionistic Fuzzy Index, Fuzzy Implication Operator, Automorphisms.

1 Introduction

In 1983 Atanassov [1] introduced Atanassov's intuitionistic fuzzy sets ($A-IFSs$) in such a way that each element of the set has two values assigned, the membership degree and the non-membership degree. In this direction, Atanassov defined these sets also indicating that the index of intuitionism of each element obtained by subtracting the sum of membership and non-membership from one (a subtraction that should be positive and less than or equal to one), it is a measurement of the effect of working with Atanassov's intuitionistic fuzzy sets. We consider this intuitionistic index (or condition of intuitionism) a very important characteristic of $A-IFSs$ since from it we can obtain very valuable information of each element and taking on advantage of this potentiality in different applications. For example, in image processing the task of divide into disjoint parts a digital image is denoted as segmentation. The most commonly used strategy for segmenting images is global thresholding that refers to the process of dividing the pixels in an image on the basis of their intensity levels of gray. The experts have uncertainty when assigning the pixels either to the background or to the object trough the choice of the membership functions. Moreover, this choice has proven to be of uttermost importance regarding the algorithms performance. In order to overcome this problem, we consider using the Atanassov's intuitionistic fuzzy index values for representing the uncertainty of the expert in determining that the pixel belongs to the background or that it belongs to the object. From this point of view, we can consider the expert provides the degree of membership of an element to an $A-IFS$ and also the degree of intuitionism the expert has in given this membership degree (see [9]). This fact has led us to present the new concept of *Generalized Atanassov's Intuitionistic Fuzzy Index* that generalize the expression given by Atanassov. We also provide a characterization method by means of fuzzy implication operators. Moreover, we study a construction method using automorphisms that allow us to present simple expressions of said index.

Pankowska and Wygralak ([16, 17, 21]) proposed another generalization of the intuitionistic index based on strong negations and triangular norms. This approach is used to construct flexible algorithms of group decision making which involve relative scalar cardinalities defined by means of generalized sigma counts of fuzzy sets.

2 Preliminary definitions

Let U be an ordinary finite non-empty set. An *Atanassov's intuitionistic fuzzy set* ($A-IFS$) [1] in U is an expression A given by

$$A=\{(u,\mu_A(u),\nu_A(u))|u\in U\} \tag{1}$$

where

$$\begin{array}{l}\mu_A\ :\ U\longrightarrow[0,1]\\ \nu_A\ :\ U\longrightarrow[0,1]\end{array}$$

satisfy the condition $0\leq\mu_A(u)+\nu_A(u)\leq 1$ for all u in U.

The numbers $\mu_A(u)$ and $\nu_A(u)$ denote respectively the degree of membership and the degree of non-membership of the element u in set A. We will also use the notation $A(u)=(\mu_A(u),\nu_A(u))$. We will represent as $A-IFSs(U)$ the set of all the Atanassov's intuitionistic fuzzy sets in U.
Atanassov defined the *Atanassov's intuitionistic fuzzy index* of the element u in $A\in A-IFSs(U)$ as:

$$\Pi_A(u)=1-\mu_A(u)-\nu_A(u). \tag{2}$$

We know fuzzy sets are represented exclusively by the membership function degree, that is,

$$A=\{(u,\mu_A(u))|u\in U\}. \tag{3}$$

Hereinafter, fuzzy sets have associated a non-membership degree given by one minus the membership degree:

$$\begin{array}{ll}A= & \{(u,\mu_A(u),\nu_A(u))|u\in U\}=\\ & \{(u,\mu_A(u),1-\mu_A(u))|u\in U\}.\end{array} \tag{4}$$

Since $\mu_A(u)+\nu_A(u)=\mu_A(u)+1-\mu_A(u)=1$, in this sense fuzzy sets are considered as a particular case of Atanassov's intuitionistic fuzzy sets. We will represent as $FSs(U)$ the set of all the fuzzy sets in U.

We will call automorphism of the unit interval every function $\varphi:[0,1]\to[0,1]$ that is continuous and strictly increasing such that $\varphi(0)=0$ and $\varphi(1)=1$.

A function $n:[0,1]\to[0,1]$ such that $n(0)=1$ and $n(1)=0$ is called a *strong negation* whenever it is strictly decreasing, continuous and involutive. Trillas ([19, 20]) proved that $n:[0,1]\to[0,1]$ is a strong negation if and only if there exists an automorphism φ of the unit interval such that

ISBN: 978-989-95079-6-8

$n(x) = \varphi^{-1}(1 - \varphi(x))$. In this work we will only consider strong negations.

We denote as L^* the following set:

$$L^* = \{(x, y) | (x, y) \in [0, 1] \times [0, 1] \; and \; x + y \leq 1\} \quad (5)$$

and the elements $0_{L^*} = (0, 1)$ and $1_{L^*} = (1, 0)$.

For every $(x, y), (z, t) \in L^*$ the following expressions are known ([1]-[8],[12]-[14]):

- $(x, y) \leq_{L^*} (z, t) \; if \; and \; only \; if \; x \leq z \;\; and \; y \geq t$. This relation is transitive, reflexive and antisymmetric.
- $(x, y) = (z, t) \;\; if \;\; and \;\; only \;\; if \;\; (x, y) \leq_{L^*} (z, t) \;\; and \;\; (z, t) \leq_{L^*} (x, y)$.
- $(x, y) \preceq (z, t) \; if \; and \; only \; if \; x \leq z \;\; and \; y \leq t$.

  In [14] is proven $(L^*, \leq_{L^*})$ is a complete lattice.

Therefore, if $A \in FSs(U)$, then

$$A_c = \{(u, 1 - \mu_A(u), \mu_A(u)) | u \in U\}. \quad (6)$$

Next, we recall the definition of Atanassov's intuitionistic fuzzy t-norm and t-conorm and also the notion of t-representability, in such a way, Deschrijver *et al.* gave a construction method of Atanassov's intuitionistic fuzzy t-norms and t-conorms by means of t-norms and t-conorms on $[0, 1]$ (see [14]).

Definition 1 *A function* $\mathbf{T} : (L^*)^2 \to L^*$ *is said to be an Atanassov's intuitionistic fuzzy t-norm if it is commutative, associative, and increasing (in both arguments with respect to the order* $\leq_{L^*}$*), with neutral element* 1_{L^*}*. In the same way, a function* $\mathbf{S} : (L^*)^2 \to L^*$ *is said to be an Atanassov's intuitionistic fuzzy t-conorm if it is commutative, associative, increasing and with neutral element* 0_{L^*}*.*

Definition 2 *An Atanassov's intuitionistic fuzzy t-norm* $\mathbf{T}$ *is called t-representable if and only if there exist a t-norm* T *and a t-conorm* S *on* $[0, 1]$ *such that, for all* $(x, y), (z, t) \in L^*$

$$\mathbf{T}((x, y), (z, t)) = (T(x, z), S(y, t)) \in L^*. \quad (7)$$

An Atanassov's intuitionistic fuzzy t-conorm $\mathbf{S}$ *is called t-representable if and only if there exist a t-norm* T *and a t-conorm* S *on* $[0, 1]$ *such that, for all* $(x, y), (z, t) \in L^*$

$$\mathbf{S}((x, y), (z, t)) = (S(x, z), T(y, t)) \in L^*. \quad (8)$$

Example 1 *Let be* $(x, y), (z, t) \in L^*$, $T = min$ *and* $S = max$ *on* $[0, 1]$

$$(a) \quad \mathbf{min}((x, y), (z, t)) = (min(x, z), max(y, t)). \quad (9)$$

$$(b) \quad \mathbf{max}((x, y), (z, t)) = (max(x, z), min(y, t)). \quad (10)$$

Definition 3 *([10, 14]) An Atanassov's intuitionistic fuzzy negation is a function* $\mathbf{n} : L^* \to L^*$ *that is decreasing (with respect to* $\leq_{L^*}$*) such that* $\mathbf{n}(0_{L^*}) = 1_{L^*}$ *and* $\mathbf{n}(1_{L^*}) = 0_{L^*}$*. If for all* $(x, y) \in L^*$ $\mathbf{n}(\mathbf{n}((x, y))) = (x, y)$ *it is said that* $\mathbf{n}$ *is involutive.*

The characterization of the Atanassov's intuitionistic fuzzy negations and the following result are presented in [14], for interval-valued fuzzy sets are proven in [10].

Theorem 1 *A function* $\mathbf{n} : L^* \to L^*$ *is an involutive Atanassov's intuitionistic fuzzy negation if and only if there exists an involutive fuzzy negation* n *such that*

$$\mathbf{n}((x, y)) = (n(1 - y), 1 - n(x)). \quad (11)$$

Throughout this work we will restrict to involutive Atanassov's intuitionistic fuzzy negations $\mathbf{n}$ generated from a given negation n, as in Theorem 1.

3 Generalized Atanassov's Intuitionistic Fuzzy Index

In this section we propose the definition of generalized Atanassov's intuitionistic fuzzy index and we characterize such index by means of fuzzy implication operators and automorphisms.

Definition 4 *A function* $\Pi_G : L^* \to [0, 1]$ *is called a generalized Atanassov's intuitionistic fuzzy index associated with the strong negation* n*, if it satisfies the following conditions:*

(i) $\Pi_G((x, y)) = 1$ *if and only if* $x = 0$ *and* $y = 0$*;*

(ii) $\Pi_G((x, y)) = 0$ *if and only if* $x + y = 1$*;*

(iii) If $(z, t) \preceq (x, y)$*, then* $\Pi_G((x, y)) \leq \Pi_G((z, t))$*;*

(iv) $\Pi_G((x, y)) = \Pi_G(\mathbf{n}((x, y)))$ *for all* $(x, y) \in L^*$ *such that* $\mathbf{n}$ *is generated from an involutive fuzzy negation* n*, as in Theorem 1.*

Example 2

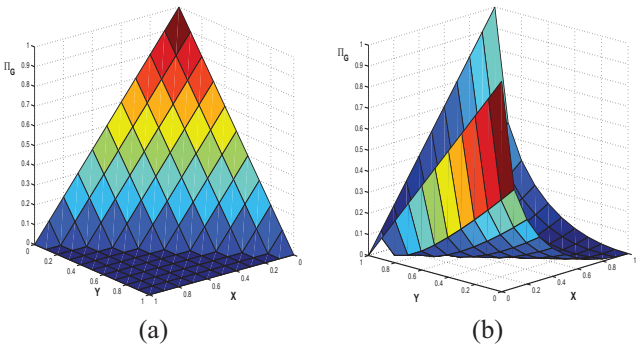
$$(a) \quad \Pi_G((x, y)) = 1 - y - x, \quad (12)$$

with $n(x) = 1 - x$ *for all* $x \in [0, 1]$*. As we can observe, this expression is equal to the expression (2) given by Atanassov.*

$$(b) \quad \Pi_G((x, y)) = ((1 - y)^{0.5} - x^{0.5})^2 \quad (13)$$

with $n(x) = (1 - x^{0.5})^2$ *for all* $x \in [0, 1]$.

Fig. 1 depicts the expressions given by Example 2.



(a) (b)

Figure 1: (a) $\Pi_G((x, y)) = 1 - y - x$ (b) $\Pi_G((x, y)) = ((1 - y)^{0.5} - x^{0.5})^2$

Next, we study the symmetry property of Π_G and we prove that only it is satisfied if we take as n the standard negation. Later, we introduce a construction method of the generalized Atanassov's intuitionistic fuzzy index using two automorphisms. In this way, it is quite simple to give different expressions of the intuitionistic index. Moreover, we analyze the case if the two automorphisms that we are taking are equal.

ISBN: 978-989-95079-6-8

Proposition 1 *Let Π_G be a generalized Atanassov's intuitionistic fuzzy index associated with the strong negation n. Then,*

$$\Pi_G((x,y)) = \Pi_G((y,x))$$

if and only if

$$n(x) = 1 - x;\ that\ is, \mathbf{n}((x,y)) = (y,x).$$

Proof. If $\Pi_G((x,y)) = \Pi_G((y,x))$ and by means of item (iv) of Definition 4 and Theorem 1, we obtain, $\Pi_G((x,y)) = \Pi_G((y,x)) = \Pi_G(\mathbf{n}((x,y))) = \Pi_G((n(1-y), 1-n(x)))$. Therefore, $n(1-y) = y$ and $n(x) = 1-x$.
On the other side, if n is the standard negation, then $\Pi_G((x,y)) = \Pi_G(\mathbf{n}((x,y))) = \Pi_G((n(1-y), 1-n(x))) = \Pi_G((y,x))$.

Proposition 2 *If φ_1, φ_2 are two automorphisms of the unit interval, then*

$$\Pi_G((x,y)) = \varphi_1^{-1}(\varphi_2(1-y) - \varphi_2(x)) \tag{14}$$

with $n(x) = \varphi_2^{-1}(1 - \varphi_2(x))$ is a generalized Atanassov's intuitionistic fuzzy index associated with the strong negation n.

Proof. We must proof the four properties of a generalized Atanassov's intuitionistic fuzzy index. The proof of properties (i) and (ii) is direct bearing in mind the bound conditions of automorphisms definition. (iii) if $(z,t) \preceq (x,y)$ then, $z \leq x$ and $t \leq y$. Automorphisms are strictly increasing therefore, $\varphi_2(z) \leq \varphi_2(x)$ and $\varphi_2(1-y) \leq \varphi_2(1-t)$, so we obtain $\Pi_G((x,y)) \leq \Pi_G((z,t))$. (iv) Bearing in mind Teorem 1 and also $n(x) = \varphi_2^{-1}(1-\varphi_2(x))$, we have $\Pi_G(\ \mathbf{n}((x,y))) = \Pi_G((n(1-y), 1-n(x))) = \varphi_1^{-1}(\varphi_2(n(x)) - \varphi_2(n(1-y))) = \varphi_1^{-1}(\varphi_2(\varphi_2^{-1}(1-\varphi_2(x))) - \varphi_2(\varphi_2^{-1}(1-\varphi_2(1-y)))) = \Pi_G((x,y))$.

Proposition 3 *In the conditions of Proposition 2:*

$$\Pi_G((0,y)) = 1 - y$$

if and only if

$$\varphi_1(x) = \varphi_2(x) \quad for\ all \quad x \in [0,1].$$

Proof. Direct.

For us, a fuzzy implication operator will be an implication in the sense of Fodor and Roubens [15], that is, a function $I : [0,1]^2 \to [0,1]$ that satisfies the following properties:

I_1. If $x \leq z$ then $I(x,y) \geq I(z,y)$ for all $y \in [0,1]$;

I_2. If $y \leq t$ then $I(x,y) \leq I(x,t)$ for all $x \in [0,1]$;

I_3. $I(0,x) = 1$ (dominance of falsity) for all $x \in [0,1]$;

I_4. $I(x,1) = 1$ for all $x \in [0,1]$.

I_5. $I(1,0) = 0$.

Depending on the application the following properties can also be demanded to a fuzzy implication operator:

I_6. $I(1,x) = x$ (neutrality of truth).

I_7. $I(x, I(y,z)) = I(y, I(x,z))$ (exchange property).

I_8. $I(x,y) = 1$ if and only if $x \leq y$.

I_9. $I(x,y) = I(n(y), n(x))$ (contraposition) with a strong negation n.

I_{10}. I is a continuous function.

In [6] is provide the following important result.

Proposition 4 *Let $I : [0,1]^2 \to [0,1]$. For all $x, y, z \in [0,1]$, the following properties hold:*

(i) *I satisfies I_1 if and only if $I(max(x,y), z) = min(I(x,z), I(y,z))$.*

(ii) *I satisfies I_1 if and only if $I(min(x,y), z) = max(I(x,z), I(y,z))$.*

(iii) *I satisfies I_2 if and only if $I(x, min(y,z)) = min(I(x,y), I(x,z))$.*

(iv) *I satisfies I_2 if and only if $I(x, max(y,z)) = max(I(x,y), I(x,z))$.*

Next, we will give a characterization of the generalized Atanassov's intuitionistic fuzzy index Π_G by means of a function $I : [0,1]^2 \to [0,1]$ that satisfies some properties of fuzzy implication operators.

Theorem 2 *Let n be a strong negation. A function $\Pi_G : L^* \to [0,1]$ is a generalized Atanassov's intuitionistic fuzzy index associated with the strong negation n*

if and only if

there exists a function $I : [0,1]^2 \to [0,1]$ satisfying $I1, I8, I9$, and $I(x,y) = 0$ if and only if $x = 1$ and $y = 0$, such that

$$\Pi_G((x,y)) = n(I(1-y, x)).$$

Proof. Sufficiency. We need to prove the four properties of Definition 4. Properties (i) and (ii) are direct. Property (iii) we must consider the result proven in [6] that if I satisfies I_1 and I_9, then I satisfies I_2. Property (iv) is direct bearing in mind that I satisfies I_9 and n is a strong negation. In order to prove necessity, let us suppose that Π_G is a generalized Atanassov's intuitionistic fuzzy index associated with the strong negation n. Define I

$$I(x,y) = \begin{cases} 1 \text{ if } x \leq y \\ n(\Pi_G((y, 1-x))) \text{ if } x > y. \end{cases} \tag{15}$$

Now it is easy to prove that I satisfies I_1, I_8, I_9 and also $I(x,y) = 0$ if and only if $x = 1$ and $y = 0$.

Example 3 *We build these generalized Atanassov's intuitionistic fuzzy indices associated with the strong negation $n(x) = 1 - x$ for all $x \in [0,1]$.*

(a) Lukasiewicz implication: $I(x,y) = min(1, 1-x+y)$. Then,

$$\Pi_G((x,y)) = n(I(1-y,x)) =$$

$$n(min(1, 1-1+y+x)) = n(y+x) = 1-y-x. \tag{16}$$

(b) Fodor implication:

$$I(x,y) = \begin{cases} 1 & \text{if } x \leq y \\ max(1-x,y) & \text{if } x > y \end{cases} \quad (17)$$

In this case we have

$$\Pi_G((x,y)) = \begin{cases} 0 & \text{if } x+y=1 \\ n(max(y,x)) & \text{if } x+y<1. \end{cases} \quad (18)$$

Corollary 1 *Let n be a strong negation. A continuous function $\Pi_G : L^* \to [0,1]$ such that $\Pi_G((x,0)) = n(x)$ for all $x \in [0,1]$ is a generalized Atanassov's intuitionistic fuzzy index associated with the function $I : [0,1]^2 \to [0,1]$ with I satisfying $I7$ and $I10$*

if and only if

there exists an automorphism φ of the unit interval such that

$$\Pi_G((x,y)) = \varphi^{-1}(\varphi(1-y) - \varphi(x)) \text{ and}$$

$$n(x) = \varphi^{-1}(1-\varphi(x))$$

Proof. It's enough to take into account Theorem 2, the relations between the properties of I studied in [6] and the following theorem proved in 1987 by Smets and Magrez (see [18]): a function $I : [0,1]^2 \to [0,1]$ verifies $I2$, $I7$, $I8$ and $I10$ if and only if there exists an automorphism φ of the unit interval such that $I(x,y) = \varphi^{-1}(min(1, 1-\varphi(x)+\varphi(y)))$.

Example 4 *if $\varphi(x) = x^2$ for all $x \in [0,1]$, then*

$$\Pi_G((x,y)) = ((1-y)^2 - x^2)^{0.5}, \quad (19)$$

with $n(x) = (1-x^2)^{0.5}$ for all $x \in [0,1]$.

Next, we provide the expressions of Π_G when it is applied to the meet operator min and the join operator max on L^*.

Theorem 3 *Under conditions of Theorem 2 the following items hold:*

$$(1)\ \Pi_G(\mathbf{max}((x,y),(z,t))) = min\{max(\Pi_G((x,y)), \Pi_G((x,t))), max(\Pi_G((z,y)), \Pi_G((z,t)))\}$$

$$(2)\ \Pi_G(\mathbf{min}((x,y),(z,t))) = max\{min(\Pi_G((x,y)), \Pi_G((x,t))), min(\Pi_G((z,y)), \Pi_G((z,t)))\}$$

Proof. (1) We must take into account the definition of meet operator min and the join operator max on L^*, also if I satisfies I_1 and I_9, then I satisfies I_2 and finally the items (i), (iv) of Proposition 4.

$\Pi_G(\mathbf{max}((x,y),(z,t))) = \Pi_G((max(x,z), min(y,t))) = n(I(1-min(y,t), max(x,z)))$. Therefore,

$$\begin{array}{ll} n(\Pi_G((max(x,z), min(y,t)))) & = \\ I(1-min(y,t), max(x,z)) & = \\ I(max(1-y,1-t), max(x,z)) & = \\ max\{I(max(1-y,1-t),x), & \\ I(max(1-y,1-t),z)\} & = \\ max\{min(I(1-y,x), I(1-t,x)), & \\ min(I(1-y,z), I(1-t,z))\} & = \\ max\{min(n(\Pi_G((x,y))), n(\Pi_G((x,t)))), & \\ min(n(\Pi_G((z,y))), n(\Pi_G((z,t))))\} & = \\ max\{n(max(\Pi_G((x,y)), \Pi_G((x,t)))), & \\ (n(max(\Pi_G((z,y)), \Pi_G((z,t))))\} & = \\ n(min\{max(\Pi_G((x,y)), \Pi_G((x,t))), & \\ max(\Pi_G((z,y)), \Pi_G((z,t)))\}). & \end{array}$$

(2) It is quite similar to the previous one but we must take into account items (ii) and (iii) of Proposition 4.

Note the result presented in Theorem 3 is not valid for any t-norm, t-conorm in L^*. The reason is that we have not defined the order relation $\leq_{L^*}$ in Definition 4, only we have defined the item (iii) If $(z,t) \preceq (x,y)$, then $\Pi_G((x,y)) \leq \Pi_G((z,t))$. That is, $\mathbf{max}((x,y),(z,t)) = (max(x,z), min(y,t)) \leq_{L^*} (S(x,z), T(y,t)) = \mathbf{S}((x,y),(z,t))$ and we obtain that $\Pi_G(\mathbf{max}((x,y),(z,t)))$ and $\Pi_G(\mathbf{S}((x,y),(z,t)))$ are imcomparable.

4 Conclusions

We define the concept of *generalized Atanassov's intuitionistic fuzzy index* and we give different construction methods. We consider this concept could be applied to image processing and we want to relate it with the concept of local contrast of a window of an image, that is, the local variations in brightness. In [11] we proposed an expression of the local contrast constructed from Atanassovs intuitionistic index by means of Atanassov's intuitionistic fuzzy S-implications. Local contrast, from our point of view, must satisfy some specific properties and we would like to study the performance of the new concept presented in this work with regard to our definition of local contrast of an image.

Acknowledgment

This paper has been partially supported by the National Science Foundation of Spain, Reference TIN2007-65981.

References

[1] K. T. Atanassov. Intuitionistic fuzzy sets. *In: V. Sgurev, ed., VII ITKR's Session, Sofia*, June 1983 (deposed in Central Science and Technical Library, Bulgarian Academy of Sciences, 1697/84, in Bulgarian).

[2] K. T. Atanassov. Intuitionistic Fuzzy Sets. *Fuzzy Sets and Systems*,20:87–96, 1986.

[3] K. Atanassov, G. Gargov. Elements of Intuitionistic Fuzzy Logic. Part I. *Fuzzy Sets and Systems*,95:39-52, 1998.

[4] P. Burillo, H. Bustince. Orderings in the referential set induced by an intuitionistic fuzzy relation. *Notes on Intuitionistic Fuzzy Sets*, 1:93-103, 1995.

ISBN: 978-989-95079-6-8

[5] H. Bustince, P. Burillo, V. Mohedano. About Intuitionistic Fuzzy Implication Operators. *JCIS 2002, Research Triangle Park*, 109–113, North Carolina, USA, 2002.

[6] H. Bustince, P. Burillo, F. Soria. Automorphisms, Negations and Implication Operators. *Fuzzy Sets and Systems*, 134:209-229, 2003.

[7] H. Bustince, E. Barrenechea, V. Mohedano. Intuitionistic Fuzzy Implication Operators. An expression and main properties. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 12(3):387–406, 2004.

[8] H. Bustince, V. Mohedano, E. Barrenechea, M. Pagola . A study of the Intuitionistic Fuzzy S-Implication. First Properties. *Procesos de Toma de Decisiones, modelado y Agregación de Preferencias*, 141–150, Eds.:E. Herrera-Viedma, 2005.

[9] H. Bustince, M. Pagola, P. Melo-Pinto, E. Barrenechea, P. Couto, Image Threshold Computation by modelizing Knowledge/Unknowledge by means of Atanassov's Intuitionistic Fuzzy Sets. *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models. Intelligent Systems from Decision Making to Data Mining, Web Intelligence and Computer Vision*, 621–638, Eds.: H. Bustince, F. Herrera, J. montero, (Springer-Verlag), 2008.

[10] H. Bustince, E. Barrenechea, M. Pagola. Generation of interval-valued fuzzy and Atanassov's intuitionistic fuzzy connectives from fuzzy connectives and from K_α operators: Laws for conjunctions and disjunctions, amplitude. *International Journal of Intelligent Systems*, 23(6):680-714, 2008.

[11] H. Bustince, E. Barrenechea, M. Pagola, J. Fernandez, J. Olagoitia, P. Melo-Pinto, P. Couto. Contrast computing using Atanassov's intuitionistic fuzzy sets. *In Proc. 8th International FLINS Conference on Computational Intelligence in Decision and Control* , Madrid: 259–264, 2008.

[12] C. Cornelis, G. Deschrijver. The Compositional Rule of Inference in an Intuitionistic Fuzzy Logic Setting. *In Proc. 6th ESSLLI Students Session*, 2001.

[13] C. Cornelis, G. Deschrijver, E. Kerre. Classification of Intuitionistic Fuzzy Implicators: an Algebraic Approach. *In Proc. 6th Joint Conf. Information Sciences*, Research Triangle Park, USA: 105–108, 2002.

[14] G. Deschrijver, C. Cornelis, E. Kerre. On the representation of intuitionistic fuzzy t-norms and t-conorms. *IEEE Transactions on Fuzzy Systems*, 12(1):45-61, 2004.

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

[16] A. Pankowska, M. Wygralak. On hesitation degrees in IF-set theory. *In: LNAI 3070,* Springer, 338–343, 2004.

[17] A. Pankowska, M. Wygralak. General If-sets with triangular norms and their applications to group decision making. *Information Sciences*, 176:2713–2754, 2006.

[18] P. Smets, P. Magrez. Implication in fuzzy logic. *International Journal of Approximate Reasoning*, 1:327-347, 1987.

[19] E. Trillas. Sobre funciones de negación en la teoría de conjuntos difusos. *Stochastica*, III(1):47-59, 1979 (in Spanish). Reprinted (English version) in: Advances of Fuzzy Logic. S. Barro et altri-Universidad de Santiago de Compostela, 31-43,1998.

[20] E. Trillas, C. Alsina, J.M. Terricabras. *Introducción a la lógica borrosa*, Ariel Matemática, 1995.

[21] M. Wygralak. Representing incomplete knowledge about fuzzy sets. *In Proc. 5th Eusflat Conf.*, Ostrava, Czech Rep.: Vol.II, 287–292, 2007.

ISBN: 978-989-95079-6-8

Time Series Analysis and Prediction Based on Fuzzy Rules and the Fuzzy Transform

Martin Štěpnička Viktor Pavliska Vilém Novák

Irina Perfilieva Lenka Vavříčková Iva Tomanová

Institute for Research and Applications of Fuzzy Modeling, University of Ostrava
Ostrava, Czech Republic
Email: {martin.stepnicka,viktor.pavliska,vilem.novak,irina.perfilieva, lenka.vavrickova, iva.tomanova}@osu.cz,

Abstract— *A new methodology for analysis and forecasting of time series is proposed. It directly employs two techniques: the fuzzy transform and the perception-based logical deduction. Due to the usage of both of them and due to the innovative approach consisting in a construction of several independent models, the methodology is successfully applicable for robust long time predictions.*

Keywords— Time series, fuzzy transform, perception-based logical deduction, fuzzy partition.

1 Introduction

We propose a new methodology for forecasting time series which is based on combination of two techniques: fuzzy transform and perception-based logical deduction. The proposed methodology consists of two phases: analysis of a time series and its forecast. In both phases, the above mentioned techniques play an essential role.

The organization of the work is as follows. After preliminaries and recalling the techniques in Section 2, we continue with a detailed description of the proposed approach in Section 3. Then we demonstrate the results in Section 4.

2 Preliminaries

Let us briefly recall the main tools employed in the suggested approach, the *fuzzy transform* (F-transform) [1] and the *perception-based logical deduction* [2, 3], in particular.

2.1 Fuzzy (F)-Transform

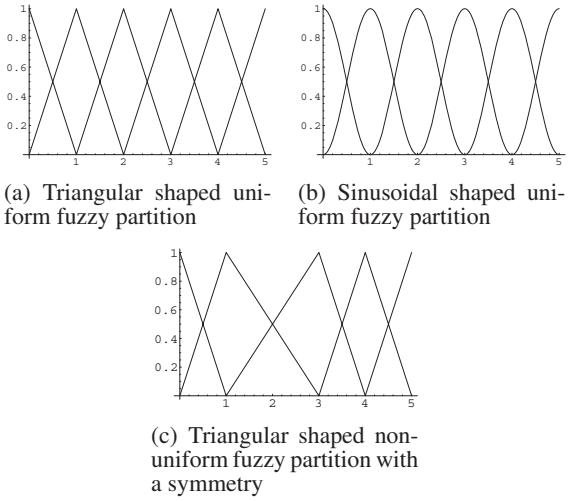
Fuzzy transform, in detailed described in [1], consists in two transformations. The first one, sometimes called the direct one, maps a continuous function defined on a fixed real interval $[a, b]$ into an n-dimensional vector. The inverse one maps the vector back to the space of continuous functions.

The F-transform is defined with respect to *basic functions* forming a fuzzy partition.

Definition 2.1 Let $c_1 < \cdots < c_n$ be fixed nodes within $[a, b]$, such that $c_1 = a, c_n = b$ and $n \geq 2$. We say that fuzzy sets $A_1, \ldots, A_n \in \mathcal{F}([a, b])$ are *basic functions* forming a fuzzy partition of $[a, b]$ if they fulfill the following conditions for $i = 1, \ldots, n$:

1. $A_i(c_i) = 1$;
2. $A_i(x) = 0$ for $x \notin (c_{i-1}, c_{i+1})$ where for the uniformity of the denotation we put $x_0 = x_1 = a$ and $x_{n+1} = x_n = b$;
3. A_i is continuous
4. A_i strictly increases on $[c_{i-1}, c_i]$ and strictly decreases on $[c_i, c_{i+1}]$;
5. for all $x \in [a, b]$

$$\sum_{i=1}^{n} A_i(x) = 1. \tag{1}$$



(a) Triangular shaped uniform fuzzy partition

(b) Sinusoidal shaped uniform fuzzy partition

(c) Triangular shaped non-uniform fuzzy partition with a symmetry

Figure 1: Graphic presentation of distinct fuzzy partitions.

Usually, the *uniform fuzzy partition* is used i.e. n equidistant nodes $c_i = c_{i-1} + h$ are fixed. Let us remark that the shape of the basic functions is not predetermined and it can be chosen based on further requirements. For some examples of fuzzy partitions fulfilling the Definition 2.1, see Fig. 1.

Definition 2.2 Let a fuzzy partition of $[a, b]$ be given by basic functions $A_1, \ldots, A_n, n \geq 2$ and let $f : [a, b] \to \mathbb{R}$ be an arbitrary continuous function. The n-tuple of real numbers $[F_1, \ldots, F_n]$ given by

$$F_i = \frac{\int_a^b f(x)A_i(x)dx}{\int_a^b A_i(x)dx}, \ i = 1, \ldots, n \tag{2}$$

is the *direct fuzzy transform (F-transform)* of f with respect to the given fuzzy partition. $F_1, \ldots, F_n$ are the *components* of the F-transform of f.

ISBN: 978-989-95079-6-8

In practice, function f is usually not given analytically but we are at least provided by some data obtained, e.g., by some measurements. In this case, Definition 2.2 can be modified in such a way that finite summations replace definite integrals in formula (2).

Let us be given n basic functions forming a fuzzy partition of $[a, b]$ and let the function f be given at $T > n$ fixed nodes $x_1, \ldots, x_T \in [a, b]$. We say that the set of nodes $\{x_1, \ldots, x_T\}$ is *sufficiently dense with respect to the fuzzy partition* if

$$(\forall i)(\exists t) \quad A_i(x_t) > 0. \tag{3}$$

Definition 2.3 Let a fuzzy partition of $[a, b]$ be given by basic functions $A_1, \ldots, A_n, n \geq 2$ and let $f : [a, b] \to \mathbb{R}$ be a function known at a set $\{x_1, \ldots, x_T\}$ of sufficiently dense nodes with respect to the given fuzzy partition. The n-tuple of real numbers $[F_1, \ldots, F_n]$ given by

$$F_i = \frac{\sum_{t=1}^{T} f(x_t)A_i(x_t)}{\sum_{t=1}^{T} A_i(x_t)}, \; i = 1, \ldots, n \tag{4}$$

is the *(discrete) direct F-transform* of f with respect to the given fuzzy partition. $F_1, \ldots, F_n$ are the *components* of the F-transform of f.

Since this paper deals with the application of the F-transform to the analysis and prediction of time series, i.e., discrete problem, there is no danger of confusion and we may easily, for the sake of simplicity, omit the word "discrete" and simply talk about the "F-transform" in the rest of the paper.

The F-transform of f with respect to $A_1, \ldots, A_n$ will be denoted by $F_n[f] = [F1, \ldots, Fn]$. It has been proved [1] that the components of the F-transform are the weighted mean values of an original function where the weights are given by the basic functions.

The original function f can be approximately reconstructed (with the help of the inversion formula) from its fuzzy transform $F_n[f]$. The function represented by the inversion formula is called the *inverse F-transform*.

Definition 2.4 Let $F_n[f]$ be the direct F-transform of f with respect to $A_1, \ldots, A_n \in \mathcal{F}([a, b])$. Then the function $f_{F,n}$ given on $[a, b]$ as follows

$$f_{F,n}(x) = \sum_{i=1}^{n} F_i A_i(x), \tag{5}$$

is called the *inverse F-transform* of f.

The inverse F-transform is a continuous function on $[a, b]$ no matter whether we have started from analytically given continuous f or we have been given only function values on a sufficiently dense set. Anyhow, for the time series problem, we consider the inverse F-transform at the same points where the original function is given.

2.2 Perception-based Logical Deduction

In this subsection, we briefly recall the main idea of the *perception-based logical deduction* (PbLD). For details, let us refer to [2, 3].

The PbLD is a method of deducing conclusions on the basis of a *linguistic description* which is a set of m fuzzy/linguistic rules:

IF x_1 is $\mathcal{A}_{11}$ **AND** $\cdots$ **AND** x_q is $\mathcal{A}_{1q}$ **THEN** z is $\mathcal{B}_1$
$\ldots$
AND
$\ldots$
IF x_1 is $\mathcal{A}_{m1}$ **AND** $\cdots$ **AND** x_q is $\mathcal{A}_{mq}$ **THEN** z is $\mathcal{B}_m$ (6)

where $\mathcal{A}_{11}, \ldots, \mathcal{A}_{nq}, \mathcal{B}_1, \mathcal{B}_m$ are specific linguistic expressions called *evaluative expressions* which have the form

$$\langle \text{ linguistic hedge } \rangle \; \{ \textit{ small, medium, big } \}.$$

Note that extension of $\mathcal{A}$ is a functional value of its intension in a point w. These extensions are specific fuzzy sets having the form depicted in Fig.2. Let us shortly mention, that formal theory of evaluative linguistic expression is a deep and firm theory and it would be out of range of the conference paper to recall it in a more detailed way to readers. Therefore, we refer to its detailed description and justification published in [4].

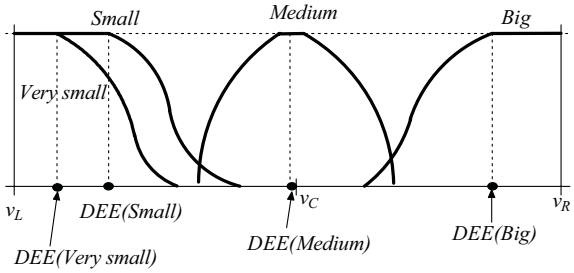


Figure 2: Fuzzy sets representing typical evaluative linguistic expressions.

In the frame of the above outlined theory, that the rules in (6) are taken as genuine conditional expressions of natural language that are interpreted accordingly, see [5], i.e. intension and extensions of them are also defined. The extension of each rule is a special fuzzy relation computed using a t-norm (usually minimum) and Łukasiewicz (residual) implication.

By *perception* we understand an evaluative expression assigned to the given value in the given context. The choice of perception is not arbitrary and besides the context, it also depends on the topic of the specified linguistic expression. The PbLD method uses a sophisticated algorithm for choosing the best fitting rule(s). The most proper crisp output is obtained using a DEE (Defuzzification of Evaluative Expressions) method [2].

2.3 Time Series

Let

$$\{x_t | t = 1, \ldots, T\} \subset \mathbb{R}, \quad T \geq 3 \tag{7}$$

be a given time series. The task is to analyze and forecast the time series, i.e., to determine values

$$\{x_t | t = T + 1, \ldots, T + k\} \subset \mathbb{R}, \quad k \geq 1. \tag{8}$$

There are two standard approaches to forecast a time series. The first one uses autoregressive and moving averages models of the so called *Box-Jenkins methodology* [6]. This approach,

ISBN: 978-989-95079-6-8

no matter how useful and successful in forecasts it may be, besides stationarity and seasonality provides us with a non-transparent and non-interpretable analysis. Therefore, it is not appropriate for an extension by means and tools directly using the fuzzy sets theory, where the interpretability and transparency play a crucial role.

Remark 2.5 A study presenting Takagi-Sugeno rules [7] in the view of the autocorrelation Box-Jenkins methodology has been published [8]. Nevertheless, the Takagi-Sugeno rules are rather regression-based than linguistic-based in comparison with (6). Analogously, distinct neuro-fuzzy approaches, which are on the border of neural networks, Takagi-Sugeno models and evolving fuzzy systems, are very often and successfully used [9, 10]. But well tuned Gaussian fuzzy sets with a centroid at node 5.6989 and a width parameter equal to 2.8893 (see [10]) constructed as a product of an employed optimization technique are undoubtably far from the interpretability of systems using models of fragments of natural language. Therefore, these approaches, no matter how effective and powerful, are closer to standard regression methods.

The second approach consists in the so called *decomposition* [11] of a time series where each element x_t of the time series is decomposed into the following components

$$x_t = Tr_t + S_t + C_t + E_t \tag{9}$$

where Tr_t, S_t, C_t, E_t are the t-th *trend, seasonal, cyclic* and *error* components, respectively.

Trend and seasonal components may be analyzed. The cyclic component is a bit problematic. The name "cyclic" comes from the economical cycles which are not regular and they are dependent on many outer factors to be analyzed from the past. The error component is a random noise which basically cannot be forecasted and therefore is omitted from our further considerations.

After these omissions, for further investigation, usually the following simplified decomposed model is considered

$$x_t = Tr_t + S_t. \tag{10}$$

Remark 2.6 Formulas (9) and (10) describe so called additive decomposition. Using multiplications instead of summations, we would get the multiplicative decomposition. Since the treatment is analogous, for the sake of simplicity we restrict our focus on the additive ones but the further methods and results are similar.

3 Proposed approach

Let us introduce our suggested approach directly employing the techniques briefly recalled in Section 2.

3.1 Trend and Seasonal Analysis

Traditional approach to the trend analysis assumes the trend to be an a priori given function, e.g., linear, polynomial, exponential or a kind of a saturation function such as sigmoidal function, for instance. This approach simplifies the analysis, which consists in a regressive determination of parameters of the predetermined function, and even the prediction, which consists in a simple prolongation – i.e. in an evaluation of the determined trend function at nodes $T+1, \ldots, T+k$.

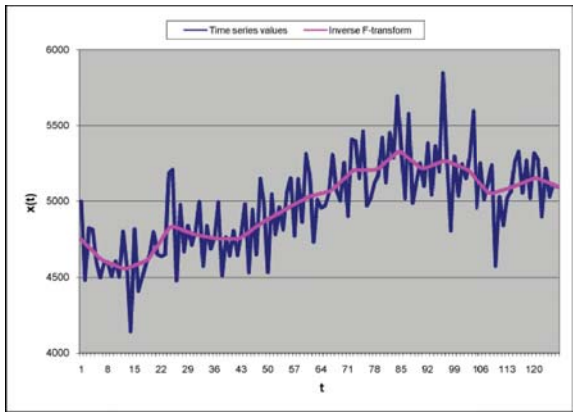


Figure 3: Time series with its inverse F-transform. Standard trend analysis would be inappropriate due to irregular cyclic changes.

But such an approach is not always the most proper one, see Fig. 3. The problem is that such an approach where we fix a trend function for the whole domain is too restrictive. The trend may change during the measurements of a time series especially in case of a long time series. For such cases, complicated adaptive trend changing models are constructed. As a typical example, we may especially in these days of the financial crisis recall the equity indexes where we may hardly simply prolong the trends while very often after robust growth radical and dramatic falls come followed by stagnation and then again by growths, due to the cyclic influences. Here, the prolongation might be very often the worst thing we may apply in predictions.

We propose the F-transform method for the trend description. The F-transform does not fix any shape of the curve, but it possess powerful approximation, noise reduction and computational properties [12].

The time series x_t may be viewed as a function x on the interval $[0, T]$ which we are not given analytically, but we are provided by measurements $x(t) = x_t$ at nodes $t = 1, \ldots, T$. Let us build a uniform fuzzy partition according to Definition 2.1 such that each basic function $A_2, \ldots, A_{n-1}$ "cover" the number of nodes equal to the number of nodes belonging to a season. For example, in case of a time series on the monthly basis, each basic function covers 12 point except basic functions A_1, A_n which cover the first and the last 6 nodes, respectively. So, the set of nodes is sufficiently dense with respect to the fuzzy partition. From now on, we will explicitly refer to such a time series on the monthly basis since everything may be easily generalized for a different seasonality.

Remark 3.1 Usually, we are provided by some additional information about a time series, for example that it is a GDP growth per month. In cases, when such information is not at disposal or does not lead to a sufficient knowledge to determine the seasonality (diesel engine pressure etc. [13]), the perception-based logical deduction with appropriate rules may be successfully used. For details, we refer the reader to [14].

Let $F_n[x] = [X_1, \ldots, X_n]$ be the F-transform of the function x w.r.t. the given fuzzy partition and let $x_{F,n}$ be its inverse F-transform. The inverse F-transform will serve us as a model

ISBN: 978-989-95079-6-8

of the trend component (including cyclic changes). It does not fix a priori some shape of the function. Therefore, we may easily from (9) with omitted error component determine the seasonal components S_t

$$S_t = x_t - x_{F,n}(t) \tag{11}$$

where $x_{F,n}(t) = Tr_t + C_t$.

3.2 Trend Forecast

The suggested approach to the trend analysis implicitly treats also possible cyclic component influences without complicated adaptive trend changing mechanisms. This is a significant difference to the classical approach, where we first model the trend only and then determine the seasonal components which are influenced by the cyclic irregular changes. Our approach treats the problem the other way around and the trend model $x_{F,n}$ primarily serves us to get pure seasonal components without the cyclic influences.

On the other hand, we cannot easily forecast such a trend model by the prolongation, i.e. by the evaluation of the predetermined fixed trend function at nodes $t = T+1, \ldots, T+k$. Due to the drawback of such a traditional approach, it is not disadvantage but vice-versa, as explained below.

We follow the idea of [15] and for the trend forecast, we employ the perception-based logical deduction. As antecedent variables, we consider the F-transform components of the given time series X_i, $i = 1, \ldots, n$ as well as their differences of the first and the second order

$$\begin{array}{rcll} \Delta X_i & = & X_i - X_{i-1}, & i = 2, \ldots, n \\ \Delta^2 X_i & = & \Delta X_i - \Delta X_{i-1}, & i = 3, \ldots, n \end{array} \tag{12}$$

respectively.

So, we will deal, e.g., with fuzzy rules such as the following one

$$\textbf{IF } \Delta X_{i-1} \text{ is } \mathcal{A}_{\Delta i-1} \textbf{ AND } X_i \text{ is } \mathcal{A}_i \textbf{THEN } \Delta X_{i+1} \text{ is } \mathcal{B}. \tag{13}$$

The differences of the F-transform components expressing the time series trend (and the cycle) of distinct orders are capable to describe the dynamics of the time series better than the F-transform components itself. Fuzzy rules such as (13) may describe logical dependencies of the trend changes (hidden cycle influences) which is highly desirable and suggested in comparison with the standard prolongation of the trend observed in the past. The advantage of the transparently interpretable form of the rules using fragments of natural language is doubtless. This might be helpful in better understanding of functionalities and motive factors determining the changes in a process yielding the analyzed times series.

Let us mention that fuzzy rules such as (13) are automatically generated by the *linguistic learning* algorithm [16] implemented in the software package LFLC 2000 [17] from the F-transform components of the time series and their differences.

Remark 3.2 Although the fuzzy rules are also generated from the past, the suggested approach may be able to learn, describe and successfully predict the future of equity indexes mentioned as a motivation example at the beginning of Subsection 3.1. Of course, if a similar progress has been observed and measured in the past. Prolongation of a trend function is not capable of this task.

3.3 Independent Models

Fuzzy rules and the perception-based logical deduction may help us to forecast next F-transform components $X_{n+1}, \ldots, X_{n+\ell}$ of the time series from which we easily determine trend components of the time series as the values of the inverse F-transform $x_{F,n+\ell}(T+1), \ldots, x_{F,n+\ell}(T+k)$, where $\ell < k$.

The problem may appear when the forecast is supposed to be for a too long term. In case if we want to predict the time series x_t on a monthly basis for the next 18 months, we first have to forecast X_{n+1}. For this, we use components $X_1, \ldots, X_n$ and differences $\Delta X_2, \ldots, \Delta X_n$, $\Delta^2 X_3, \ldots, \Delta^2 X_n$. Then, using the same fuzzy rules and antecedent variables $X_1, \ldots, X_{n+1}$, $\Delta X_2, \ldots, \Delta X_{n+1}$ and $\Delta^2 X_3, \ldots, \Delta^2 X_{n+1}$ we may forecast X_{n+2}. And finally, we are able to forecast X_{n+3} based on $X_1, \ldots, X_{n+2}$, $\Delta X_2, \ldots, \Delta X_{n+2}$ and $\Delta^2 X_3, \ldots, \Delta^2 X_{n+2}$.

Clearly, there is a danger of the forecast error propagation since we forecast from forecasted values. The longer-term prediction, the higher is the damage.

To overcome this problem, we suggest to build a finite number of independent trend forecasting models using the technique described in Subsection 3.2. For the sake of simplicity, let us consider ΔX_{i-1} and X_i as the antecedent variables. Then set of rules (13) will serve us as a model for the prediction the next F-transform component X_{n+1}.

Similarly, a set of rules

$$\textbf{IF } \Delta X_{i-1} \text{ is } \mathcal{A}_{\Delta i-1} \textbf{ AND } X_i \text{ is } \mathcal{A}_i \textbf{ THEN } \Delta X_{i+2} \text{ is } \mathcal{B} \tag{14}$$

generated again by the linguistic learning algorithm from the differences of the F-transform components, will serve us as a model for the prediction the next but one F-transform component X_{n+2}. And analogously, we may build the third model

$$\textbf{IF } \Delta X_{i-1} \text{ is } \mathcal{A}_{\Delta i-1} \textbf{ AND } X_i \text{ is } \mathcal{A}_i \textbf{ THEN } \Delta X_{i+3} \text{ is } \mathcal{B} \tag{15}$$

and further models, if necessary because of a longer-term forecast than just to next 18 values $x_{T+1}, \ldots, x_{T+18}$.

The suggested approach fully avoids the problem of the error accumulation due to the "forecasting from forecasted".

3.4 Seasonal Component Forecast

Forecasted F-transform components $X_{n+1}, \ldots, X_{n+\ell}$ determine the trend and cyclic components of the time series predicted for the future which are given as the values of the inverse F-transform $x_{F,n+\ell}(T+1), \ldots, x_{F,n+\ell}(T+k)$. To forecast the seasonal components and to compose them with the latter values is the last step in the time series prediction procedure.

The seasonal components are forecasted as follows. Let

$$\mathbf{S}_\xi = [S_{p\cdot(\xi-1)}, S_{p\cdot(\xi-1)+1}, S_{p\cdot(\xi-1)+2}, \ldots, S_{p\cdot(\xi-1)+p-1}] \tag{16}$$

be the ξ-th vector of the seasonal components (11) where p denotes the period of the seasonality, i.e., in our case of a time series on a monthly basis $p = 12$ and the $\mathbf{S}_\xi$ is the vector of the January, February etc. to December measurements of the ξ-th year.

Stationarity assumption on the seasonal component of the time series is considered, i.e., we assume that $\mathbf{S}_\xi$ is a linear

ISBN: 978-989-95079-6-8

combination of previous θ vectors $\mathbf{S}_{\xi-\theta}, \ldots, \mathbf{S}_{\xi-1}$. It means that we generate the following system of equations

$$\mathbf{S}_\xi = \sum_{j=1}^{\theta} d_j \cdot \mathbf{S}_{\xi-j}, \quad \xi > \theta \tag{17}$$

to which we search for the optimal solution with respect to coefficients $d_1, \ldots, d_\theta$. The computed coefficients are then used to determine the $\mathbf{S}_\xi$

Let us mention that the stationarity is a standard assumption which might be easily checked, see [6, 18].

Composition of both forecasts, of the forecasted F-transform components and the seasonal components is the last step to get the overall time series forecast. It is done inversely to the original decomposition which was either additive or multiplicative.

3.5 Optimization

There are some unknown parameters in the whole procedure which are to be determined individually for every single time series. Basically, it is the list of antecedent variables for the prediction of the F-transform components, the number of the antecedent variables and the parameter θ.

The time series is divided into the *learning set* and the *validation set* in such a way that the latter one is given by the last values of the times series of the length equal to the expected forecast. It means, the learning set is given by $\{x_1, \ldots, x_{T-k-1}\}$ and we cut $\{x_{T-k}, \ldots, x_T\}$ off the time series to determine the validation.

There is a software [19], developed at the *Institute for Research and Applications of Fuzzy Modeling*, where the parameters θ and the maximal number of antecedent variables are set up by a user. Then the software computes all possible combinations of antecedent variables up to the maximal number and combines it with seasonal components determined based on the optimal solution of (17) for all possible θ, again up to the pre-specified maximal one. For this computations, only the learning set is used.

All computed models are used to forecast $\{x'_{T-k}, \ldots, x'_T\}$ and these forecasts are compared with the validation set $\{x_{T-k}, \ldots, x_T\}$. As a suggestion to a user, all models are ordered according to a pre-specified error criterion which may be generally arbitrarily. A user may then employ any of the optimized and tuned models for the forecast of $x_{T+1}, \ldots, x_{T+k}$.

4 Demonstration

We applied our approach to analyze and forecast 111 time series on a monthly basis from the *Artificial Neural Network & Computational Intelligence Forecasting Competition NN3*. All the time series were "real-world" and chosen from different fields such as economy, meteorology, industry etc. while no such an information about the meaning or an origin of any time series was provided to participants.

There were 580 participants who registered to the competition and downloaded the time series but only 25 finally afforded to submit their results which illustrates the complicacy of the task to deal with such a huge and heterogenous set of time series. This fact is even supported by the average 15.44% error result (measured by SMAPE - Symmetric Mean Absolute Percent Error) obtained by commercial statistical software *ForecastPro* for the whole set of time series. Let us note that the *ForecastPro* software was used only as a benchmark together with five other benchmarks. Let us also remark that it employs all the existing standard techniques from the Box-Jenkins methodology while participants were limited by a single method of a non-statistical nature.

Our approach (without the construction of independent models at that times) took the 12th position with the overall average error 18.81%. In an unofficially evaluated subset of time series with a seasonal character, our approach took the 3rd rank among all participants and even beaten four of the six statistical benchmarks. The average error in this category measured by SMAPE was 12.7%.

The construction of independent models may even improve the results. For an illustration, let us turn the attention to the time series NN3_106 from the competition, see Fig. 4.

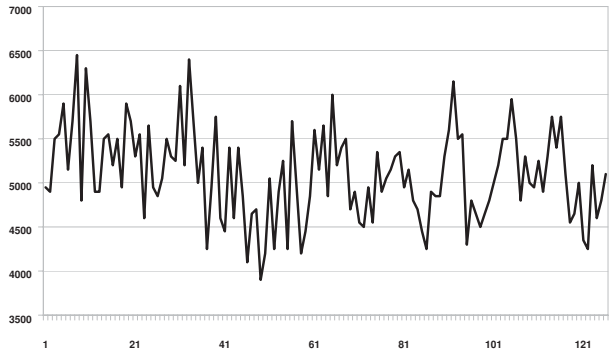


Figure 4: Time series NN3_106 from the NN3 competition.

It is a 126 months long time series with a hardly determinable trend - a typical representative of an appropriate time series for our approach. We cut off last 18 months to create a validation set and we use the first 114 months as a learning set to determine our model. The model may be used to forecast the last 18 months and then compared with the original time series values forming the validation set. There is the black line displaying the validation set of the original time series in Fig. 5.

The dotted grey line in the same figure denotes the forecast made by our methodology with a single trend prediction model optimized as described in Subsection 3.5. The generated model consisted of 15 fuzzy rules with the following 3 input variables $X_i, \Delta^2 X_i, \Delta^2 X_{i-1}$ and the output variable ΔX_{i+1}. The SMAPE error was 5.70%.

If we generate 3 independent models – all of them with the following 3 input variables $X_i, \Delta^2 X_i, \Delta^2 X_{i-1}$, every one of them predicting a different F-transform component – $\Delta X_{i+1}, \Delta X_{i+2}, \Delta X_{i+3}$, respectively, we improve the results. Due to the F-transform, the above mentioned 3 models predicting independently 3 first order differences of the F-transform components are fully sufficient to forecast values

ISBN: 978-989-95079-6-8

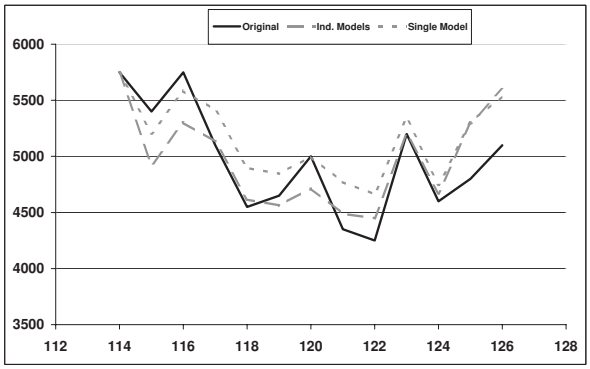


Figure 5: Time series NN3_106 (black line) and its two forecasts, one based on our methodology with a single model (dotted grey line), one forecast employs our methodology with 3 independent models (dashed grey line).

of 18 months ahead. The improve results are depicted by the grey dashed line on Fig. 5 and yielded error 4.77%.

5 Conclusions

We have proposed a novel fuzzy approach to analysis and prediction of time series. It directly uses two techniques, the F-transform and the perception-based logical deduction. Their combination helps to avoid the classical problem with prolongation of a pre-specified curve. Moreover, due to the interpretability of fuzzy rules used by the perception-based logical deduction, a user may get into a deeper understanding not only of the automatically generated model but also to the whole process whose measured samples yielded the analyzed time series.

Furthermore, due to the F-transform and the possibility to construct several independent trend forecasting models using fuzzy rules, we may easily produce long-time forecasts without the influence of the propagated and accumulated error due to the forecasting from forecasted values.

Acknowledgment

This work has been supported by project DAR 1M0572 of the MŠMT ČR.

References

[1] I. Perfilieva. Fuzzy transforms: theory and applications. *Fuzzy Sets and Systems*, 157:993–1023, 2006.

[2] V. Novák and I. Perfilieva. On the semantics of perception-based fuzzy logic deduction. *International Journal of Intelligent Systems*, 19:1007–1031, 2004.

[3] V. Novák. Perception-based logical deduction. In B. Reusch, editor, *Computational Intelligence, Theory and Applications (Advances in Soft Computing)*, pages 237–250. Springer, Berlin, 2005.

[4] V. Novák. A comprehensive theory of trichotomous evaluative linguistic expressions. *Fuzzy Sets and Systems*, 159(22):2939–2969, 2008.

[5] V. Novák and S. Lehmke. Logical structure of fuzzy IF-THEN rules. *Fuzzy Sets and Systems*, 157:2003–2029, 2006.

[6] G. Box and G. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, 1970.

[7] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man and Cybernetics*, 15:116–132, 1985.

[8] J.L.M. Aznarte, J.M. Bentez, and J.L. Castro. Fuzzy identification of systems and its applications to modeling and control. *Fuzzy Sets and Systems*, 158:2734–2745, 2007.

[9] Hai-Jun Rong, N. Sundararajan, Guang-Bin Huang, and P. Saratchandran. Sequential adaptive fuzzy inference system (safis) for nonlinear system identification and prediction. *Fuzzy Sets and Systems*, 157:1260–127, 2006.

[10] G. Leng, T.M. McGinnity, and G. Prasad. An approach for on-line extraction of fuzzy rules using a self-organising fuzzy neural network. *Fuzzy Sets and Systems*, 150:211–243, 2005.

[11] A. Bovas and J. Ledolter. *Statistical Methods for Forecasting*. Wiley, New York, 2003.

[12] I. Perfilieva and R. Valášek. Fuzzy transforms in removing noise. In B. Reusch, editor, *Computational Intelligence, Theory and Applications (Advances in Soft Computing)*, pages 221–230. Springer, Berlin, 2005.

[13] M. Štěpnička, E. Lughofer, and V. Pavliska. Comparison of data-driven fuzzy modelling methods tested on nox data. In *Abstracts of the 2nd FLLL/SCCH Master and PhD Seminar*, pages 33–39, 2006.

[14] V. Novák, M. Štěpnička, I. Perfilieva, and V. Pavliska. Analysis of periodical time series using soft computing methods. In Da Ruan, J. Montero, J. Lu, L. Martí nez, P. D'hondt, and E.E. Kerre, editors, *Computational Intelligence in Decision and Control*, pages 55–60. World Scientific, New Jersey, 2008.

[15] I. Perfilieva, V. Novák, V. Pavliska, A. Dvořák, and M. Štěpnička. Analysis and prediction of time series using fuzzy transform. In *Proc. IEEE World Congress on Computational Intelligence*, pages 3875–3879, 2008.

[16] R. Bělohlávek and V. Novák. Learning rule base of the linguistic expert systems. *Soft Computing*, 7:79–88, 2002.

[17] A. Dvořák, H. Habiballa, V. Novák, and V. Pavliska. The software package LFLC 2000 - its specificity, recent and perspective applications. *Computers in Industry*, 51:269–280, 2003.

[18] P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods (Springer Series in Statistics), 2nd ed.* Springer-Verlag, Heidelberg, 1998.

[19] H. Habiballa, V. Pavliska, and A. Dvořák. Software system for time series prediction based on F-transform and linguistic rules. *APLIMAT - Journal of Applied Mathematics*, accepted, 2009.

ISBN: 978-989-95079-6-8

On the use of aggregation operators for location privacy

Aida Valls[1], Jordi Nin[2] and Vicenç Torra[2]

1. ITAKA Research Group - Intelligent Tech. for Advanced Knowledge Acquisition
Department of Computer Science and Mathematics
Universitat Rovira i Virgili
43007 Tarragona, Catalonia, Spain

2. IIIA, Artificial Intelligence Research Institute
CSIC, Spanish National Research Council
Campus UAB s/n
08193 Bellaterra, Catalonia, Spain

Email: aida.valls@urv.cat, {jnin,vtorra}@iiia.csic.es

***Abstract*—**
Nowadays, the management of sequential and temporal data is an increasing need in many data mining processes. Therefore, the development of new privacy preserving data mining techniques for sequential data is a crucial need to ensure that sequence data analysis is performed without disclosure sensitive information. Although data analysis and protection are very different processes, they share a few common components such as similarity measurement.

In this paper we propose a new similarity function for categorical sequences of events based on OWA operators and fuzzy quantifiers. The main advantage of this new similarity function is the possibility of incorporating the user preferences in the similarity computation. We describe the implications of the application of different user preference policies in the similarity measurement when microaggregation, a well-known data protection method, is applied to sequential data.

***Keywords*—** Microaggregation, Privacy, Sequence aggregation, OWA operators.

1 Introduction

The development of new methods for managing sequential and temporal relationships is becoming an strategic area. Nowadays, thanks to the new technologies, categorical sequences of data are easily available in many scenarios. Initially, sequential data was originated by the study of gene information (DNA), but novel classes of applications based on geo-spatial information are currently appearing [1]. For instance, many applications use personal or vehicular tracking data, this information allows us to find interesting patterns to be used in many different applications, such as traffic control, sustainable mobility management, accessibility of services or even tracking patients with Alzheimer [2]. This type of data is known as Event Sequences [3].

Classical data mining algorithms and decision making techniques were developed for static data (*i.e.* data whose feature values do not depend on time). Consequently, the sequential nature of temporal data makes difficult to apply classical data mining and decision making methods. Some areas of Artificial Intelligence have started to develop new methods to deal with event sequences, such as temporal data mining algorithms [4, 5] or sequence data analysis, both interested in studying how feature values change with regards to time, in order to identify interesting temporal patterns.

As in classical data mining, in order to ensure that sequence data analysis is performed without disclosure sensitive information of data owners, statistical disclosure control (SDC) [6] and privacy preserving data mining (PPDM) [7] techniques should be applied. In this way, the privacy of the individuals can be guaranteed. A very common way to achieve a certain level of privacy on a database is to ensure k-anonymity [8, 9]. Microaggregation [10, 11] is the standard SDC technique to achieve the k-anonymity in a database. Essentially, microaggregation builds clusters of at least k records and replaces the original records with the centroid of the cluster that the record belongs to. In this way, k-anonymity is ensured because in the protected database there are at least k identical records.

In this paper we address the problem of measuring the similarity between sequences of events. Similarity measurement is a key point in data analysis but also in microaggregation. This problem was already studied in [12], where a new approach to calculate the similarity for event sequences was presented: the *Ordered-based Sequence Similarity* (OSS). This similarity is greately improved in this paper. The new version presented in this paper, applies the OWA operator to let the user to define different preference policies for the similarity comparison of two event sequences, thus, it is called OSS_{OWA}. Preference policies can be expressed using different fuzzy measures. We will present the implications that some policies (fuzzy measure) have on the results of a microaggregation procedure with an illustrative example.

This paper is organized as follows. Firstly, in Section 2 we give some preliminaries about aggregation functions and microaggregation methods for SDC. Then, in Section 3 we present a detailed description of our new similarity function for categorical sequences of events. Section 4 provides some experiments combining our similarity function with microaggregation. Finally, Section 5 draws some conclusions and describes some lines for future work.

ISBN: 978-989-95079-6-8

2 Preliminaries

In this section we give a short overview of aggregation operators, focused on the OWA operator, and microaggregation (a data protection method).

2.1 Aggregation functions

Aggregation functions [13] are functions used for information fusion. They typically combine N data values supplied by N data sources into a single datum.

In this paper, we use them to define a new similarity measure among sequences of categorical events. In particular, we consider the Ordered Weighted Aggregation (OWA) operator.

Two different definitions for this operator can be found in the literature. One applicable when the number of data sources is known in advance, and another that does not require to know how many data sources are combined. We use this latter definition, which is based on fuzzy quantifiers.

Definition 1 *A function $Q : [0, 1] \to [0, 1]$ is a* regular monotonically non-decreasing fuzzy quantifier *(non-decreasing fuzzy quantifiers for short) if it satisfies: (i) $Q(0) = 0$; (ii) $Q(1) = 1$; (iii) $x > y$ implies $Q(x) \geq Q(y)$.*

Two examples of families of fuzzy quantifiers Q_1 and Q_2 are given below. Q_1 corresponds to Yager α-quantifiers, for $\alpha > 0$.

$$Q_1^\alpha(x) = x^\alpha \tag{1}$$

$$Q_2^\alpha(x) = \begin{cases} 0 & \text{if } x = 0 \\ 1/(1 + e^{(\alpha - x)*10}) \text{ for } \alpha > 0 & \text{if } 0 < x < 1 \\ 1 & \text{if } x = 1 \end{cases} \tag{2}$$

A graphical representation of these two fuzzy quantifiers is given in Figure 1 for some particular α. $\alpha = \{0.2, 0.4, \ldots, 1.8, 2.0\}$ are used for $Q_1^\alpha(x)$, and $\alpha = \{0, 0.1, \ldots 0.9\}$ are used for $Q_2^\alpha(x)$. We can observe that for small α values, the function increases quickly near $x = 0$, whereas the increase is smoothly for larger values of α.

Using fuzzy quantifiers, the OWA operator [14, 15] is defined as follows.

Definition 2 *Let Q be a non-decreasing fuzzy quantifier, then $OWA_Q : \mathbb{R}^N \to \mathbb{R}$ is an* Ordered Weighted Averaging (OWA) operator *if*

$$OWA_Q(a_1, ..., a_N) = \sum_{i=1}^{N} (Q(i/N) - Q((i-1)/N)) a_{\sigma(i)}$$

where σ is a permutation such that $a_{\sigma(i)} \geq a_{\sigma(i+1)}$.

The interest of the OWA operators is that they permit the user to aggregate the values giving importance to large (or small) values. In the case of the quantifiers given above, the smaller the α is, the larger the importance for the largest values being aggregated. In contrast, the higher the α is, the lower the importance of the largest values (and the higher the importance given to low values). This different behaviour can be seen in Figure 1.

2.2 Microaggregation

Microaggregation [10, 11] is one of the most commonly employed data protection methods [16]. From the operational point of view, given a database with A attributes, microaggregation proceeds by building clusters of at least k records and, then, replacing each record by the centroid of the cluster to which the record belongs to. A certain level of privacy is ensured because for each attribute, there are always k records with an identical value. This assures the k-anonymity property at the attribute level [8, 9]. The goal of a microaggregation is to find clusters of at least k records while minimizing the total Sum of the Square Error:

$$SSE = \sum_{i=1}^{c} \sum_{x_{ij} \in C_i} (x_{ij} - \bar{x}_i)^T (x_{ij} - \bar{x}_i), \tag{3}$$

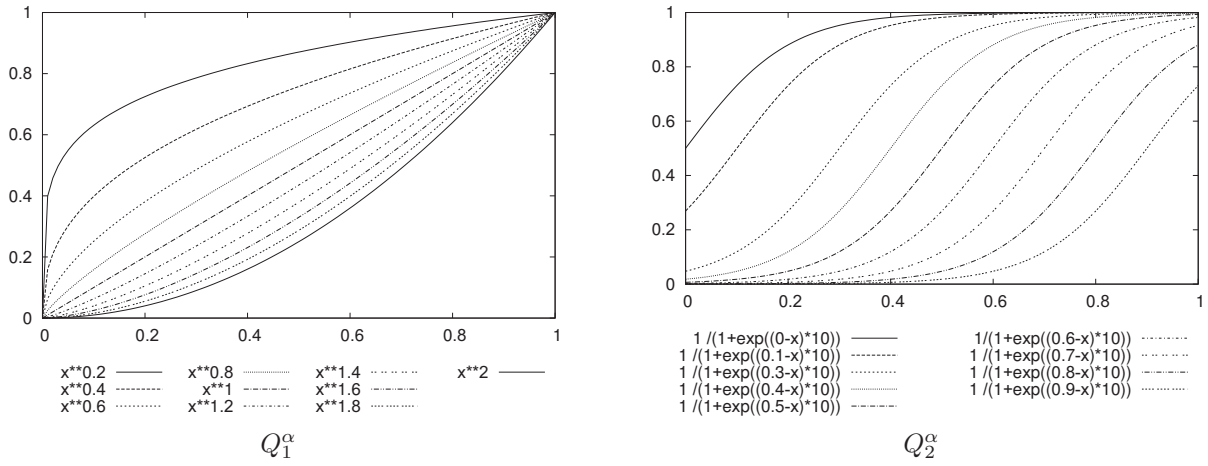
where c is the total number of clusters, C_i is the i-th cluster and $\bar{x}_i$ is the centroid of C_i. The restriction is $|C_i| \geq k$, for all $i = 1, \ldots, c$.

In the simplest case (when $A = 1$), there are polynomial time algorithms to obtain the optimal microaggregation [17]. However, for the general case ($A > 1$), the problem of finding the optimal microaggregation is NP-hard [18]. For this reason, microaggregation methods are heuristic.

It is worth to mention that most of the classical microaggregation techniques are based on the assumption that an average record for all the elements in the database can be computed, like MDAV algorithm [19]. For the case of dealing with sequences of events, this is not a feasible assumption, since sequences can be very diverse and the definition of a single average sequence is not meaningful. Other methods rely on an Euclidean space, which is also not the case of the event sequences space.

In [3] the use of hierarchical clustering algorithms for sequences of events was presented. Those methods are sequential algorithms based on a similarity measure between pairs of elements [20]. The result is a hierarchical tree of non-overlapping clusters that can be cut at any level to obtain a partition. For the experiments of this paper, the k-Sized Hierarchical Clustering method (KSHC) has been applied [21]. This method is a modification of the classical hierarchical sequential agglomerative non-overlapping clustering methods in data mining [20]. The clustering process is structured in two main steps: (1) the construction of a dissimilarity matrix for all pairs of individuals, using an appropriate similarity measure according to the type of data and problem characteristics and (2) the exploitation of the similarity matrix to build a tree of non-overlapping clusters. In this second stage, an iterative process is executed, which consists of finding the pair of elements (i.e. individuals or clusters) that are at minimum dissimilarity (i.e. the most similar) and build a new cluster with these elements. Then, those elements are removed from the dissimilarity matrix, and the dissimilarity of the new cluster with respect to the rest of elements is computed. Different methods for calculating this dissimilarity have been defined: Single Linkage (if the minimum distance with respect to the elements in the cluster is taken), Complete Linkage (if the maximum distance is taken), Average Linkage, Median Linkage, etc. The resulting tree can be cut at any level to produce a partition of the individuals. The KSHC approach keeps the

ISBN: 978-989-95079-6-8



Q_1^α Q_2^α

Figure 1: Graphical representation of Q_1^α and Q_2^α.

tradictional clustering process, but modifies it properly in order to produce a partition that respects the k-anonymity property.

2.3 The Cluster Prototype

Methods for defining cluster prototypes (or centroids) are needed in microaggregation as well as in some unsupervised machine learning techniques (*e.g.* in clustering). In general, a prototype is defined as a typical example, basis, or standard for other elements of the same cluster. Prototypes combine the most representative values for the attributes of the elements that belong to the cluster. Consequently, prototypes are typical instances that represent the content of the cluster. They are calculated using an averaging function (in most of the cases, a weighted mean is used). However, when dealing with categorical sequences of events, those averaging functions are not applicable.

In [22] a new method for generating prototypes of event sequences was presented. This method is based on the calculation of an *Element Scoring Table (EST)*, where the elements can be either individual events or patterns of consecutive events considered as an indivisible unit. So, the method is called *Ordered Element Scoring Prototyping (OESP)*. Different scores are assigned to each event depending on its position in the sequences of the cluster. The ranking given by those scores is used to iteratively append an event to the prototype (initially the prototype is an empty sequence) until the prototype has a length similar to the average lengths of the sequences in the cluster.

This method will be used to calculate the centroids of the clusters obtained by the new similarity measure, which is needed to calculate the Sum of Square Error (3).

3 OSS_{OWA}: A new dissimilarity function for categorical sequences of events

From a formal point of view, a dissimilarity function f over two sequences of events s_1 and s_2 has to fulfill the following conditions:

1. Symmetry: $f(s_1, s_2) = f(s_2, s_1)$
2. Positivity: $f(s_1, s_2) \geq 0$ for any s_1 and s_2
3. Reflexivity: $f(s_1, s_1) = 0$

Some dissimilarity measures for sequences of symbols have been defined. For instance, Hamming distance [23] or Edit (Levenstein) distance [24] are usually considered [3]. However, such distances are quite simple when it is necessary to deal with the relative ordering positions of the symbols in the sequence, as it is needed in tracking studies such as the ones presented in the introduction. Another drawback is the difficulty of incorporating user (expert) preference policies in the computation of these measures.

As we are interested in comparing temporal event sequences, the results of those traditional categorical distances are not good since they disregard a lot of information that should be taken into account to obtain a more appropriate similarity value [12]. Two good examples of temporal event sequences are web logs and the itineraries of the tourists. In these scenarios, there are two key issues to be considered:

- The number of common elements inside the sequences.
- The order (position) among the common elements inside the sequences.

On one hand, the first consideration allows us to evaluate the common part of two different sequences, if this common part is large, then the two sequences have to be similar, because such sequences show that their owners have done the same things. On the other hand, the second consideration takes into account whether the events have been done in the same order or not. This second information is very useful when applying data mining processes. For example, when detecting common itineraries for planning new bus lines. Note that in this latter case, if two sequences have the same events placed in the same order, they have to be more similar than two sequences in which the events are placed in different order. Based on those two assumptions, in [12], the Ordered-based Sequence Similarity was defined (OSS).

However, in the context of event sequence comparison, apart from these two issues, it is also interesting to incorporate any user (expert) preferences on the similarity function. In some scenarios, it is more important to group sequences sharing a large number of events even though the events are not in the same order. On the contrary, in other scenarios, the

ISBN: 978-989-95079-6-8

user is interested in tracking the sub-sequences of common events, considering that two sequences with a few events in the same order are more similar than two sequences with a larger common set of unsorted events. Let us take as an example, an analysis of the tourist behaviour in a theme park. In the first case, the manager wants to find groups of visitors that have been at the same attractions (same locations) in spite of the order of visiting them. In the second case, the manager is interested in studying the flow of visitors in the park, and to know the common paths between attractions for planning or user-recommendation purposes.

The OWA operator with a fuzzy quantifier can be used to introduce the user aggregation policy into the computation of the similarity between pairs of temporal event sequences. This permits us to tackle the issues explained before. Based on this approach, a new similarity function can be defined, $d_{OSS-OWA}$. It is calculated in two steps.

Firstly, partial distances between the elements of the two sequences are calculated as follows.

Definition 3 *Let s and r be two event sequences. Let s_i be the i-th event of the sequence s, then the event distance of s_i with respect to r is defined as*

$$d(s_i, r) = \begin{cases} \frac{min_{s_i=r_j}|i-j|}{|s|+|r|} & if\ s_i \in r \\ 1 & otherwise \end{cases} \quad (4)$$

where $|s|$ and $|r|$ stand for the sequence lengths of sequences s and r respectively.

If s_i appears n times in s and m times in r, the occurrences in s and r are matched forming pairs that minimize $|i-j|$, if some occurrences do not have a couple ($n \neq m$), they are treated as if they do not appear in r, that is, they have a distance to r of 1.

Then, those partial distances are aggregated using the OWA operator, which permits to apply different aggregation policies[14].

Formally, given sequences s and r, we compute the event distances $d(s_i, r)$ for all $i = 1, \ldots, |s|$ and $d(r_i, s)$ for all $i = 1, \ldots, |r|$. Then, the OSS_{OWA} measure is defined as follows.

Definition 4 *Let s and r be two temporal event sequences with length $|s|$ and $|r|$, respectively, then the $d_{OSS-OWA}$ dissimilarity function of s and r is defined as*

$$d_{OSS-OWA(Q)}(s, r) =$$

$$OWA_Q(d(s_1, r), \ldots, d(s_{|s|}, r), d(r_1, s), \ldots, d(r_{|r|}, s))$$

where Q is a non-decreasing fuzzy quantifier, and $d(s_i, r)$ stands for the event distance of s_i with regards to r.

Following Definition 3, all event distances equal to 1 represents the case of two sequences with non-common events. On the contrary, distances smaller than 1 represent the shift between the common events in both sequences. In this case, the larger the event distance, the larger the shift. OWA operators permit the user to aggregate the values giving more importance to large or small values when appropriate quantifiers are selected. This stands for considering more important non-common events (large distance values) or common events (small distance values). In this way we are modeling the two scenarios described above.

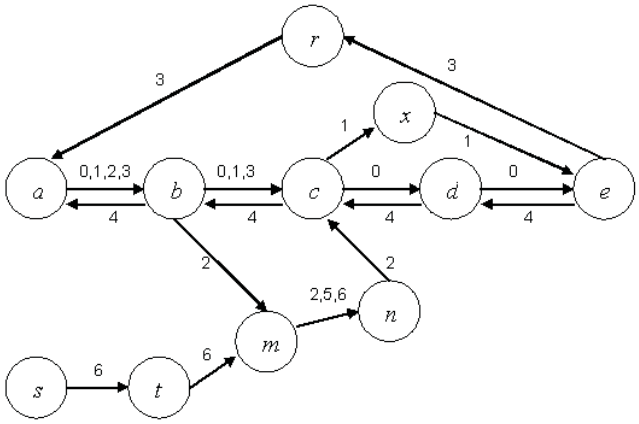


Figure 2: Example with 6 sequences.

Table 1: Original sequences.

id	Sequence
0	abcde
1	abcxe
2	abmnc
3	erabc
4	edcba
5	mn
6	stmn

4 Experiments

In this section, we present the results obtained with a synthetic sequences database (see Table 1). Figure 2 shows a graph representation of those sequences as paths between a set of 11 locations. These could be places inside a theme attractions park. The arrows indicate a transition from one attraction to the next one. The labels refer to the tourists that have followed each path step.

In order to test the $d_{OSS-OWA}$ function, the k-sized hierarchical clustering (KSHC) microaggreagtion algorithm has been used on this database. It has been applied to a k-anonymity value of $k = 2$, which means that cluster of 2 or 3 elements can be generated (recall that the number of elements is constrained by k and $2k-1$). Then, the method explained in section 2.3 for prototype generation has been performed to obtain the centroids of the clusters and calculate the information loss, that is, the SSE (3) .

Six different fuzzy quantifiers have been tested: $Q_1^{0.4}$, Q_1^1, Q_1^2, $Q_2^{0.1}$, $Q_2^{0.5}$ and $Q_2^{0.8}$. They define different aggregation policies in order to show how to introduce expert knowledge in the sequence similarity computation. Quantifiers with small α parameter consider more similar those sequences with a large set of common events. On the contrary, quantifiers with large α parameter consider more similar those sequences with few common events but placed in the same order. Note that, quantifiers with medium α parameter consider common events as important as events placed in the same order.

Table 2 shows the results obtained for the dataset given in Table 1. Each row corresponds to a different parametrization of the fuzzy measure used to compute the OSS_{OWA} dissimilarity function between pairs of sequences. The first three

ISBN: 978-989-95079-6-8

Table 2: Microaggregated sequences with $k = 2$.

Q	SSE	Sequences
$Q_1^{0.4}$	0.515	abcxe, erabc
	0.767	abcde, edcba, abmnc
	0.907	mn, stmn
Q_1^{2}	0.032	erabc, edcba
	0.063	abmnc, abcde, abcxe
	0.081	mn, stmn
Q_1^{4}	0.002	erabc, edcba
	0.0	abcde, abcxe
	0.157	mn, abmnc, stmn
$Q_2^{0.1}$	0.934	abcxe, erabc
	1.137	abcde, edcba, abmnc
	1.614	mn, stmn
$Q_2^{0.5}$	0.087	erabc, edcba
	0.162	abmnc, abcde, abcxe
	0.159	mn, stmn
$Q_2^{0.8}$	0.002	erabc, edcba
	0.0	abcde, abcxe
	0.146	mn, abmnc, stmn

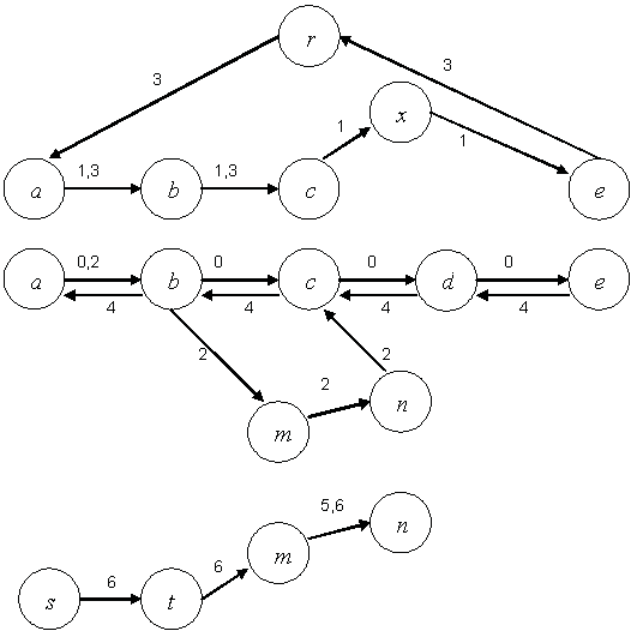


Figure 3: Clusters obtained with small α.

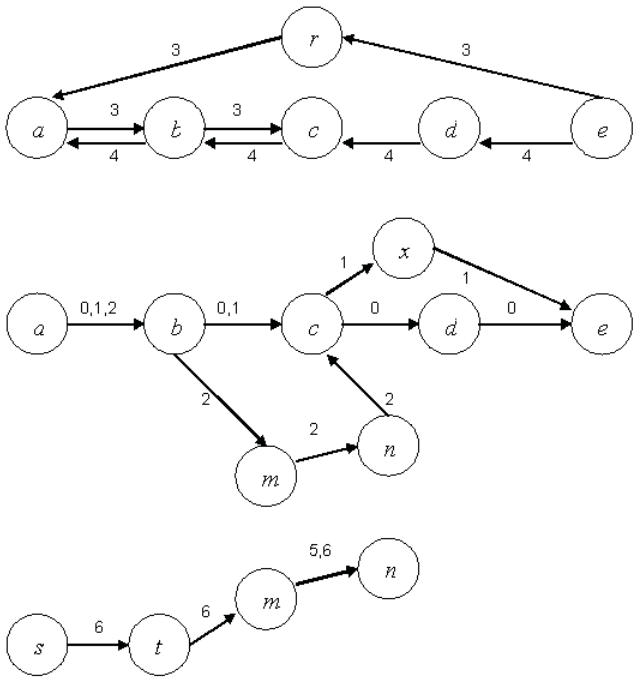


Figure 4: Clusters obtained with a medium α.

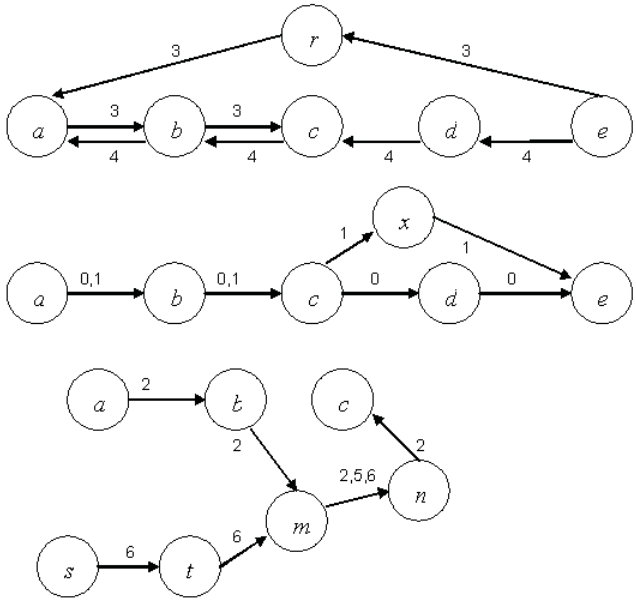


Figure 5: Clusters obtained with large α.

rows correspond to the fuzzy measure Q_1 (1) and the three last rows correspond to the fuzzy measure Q_2 (2).

The first thing to note is that the results for Q_1 and Q_2 are the same for low, medium and high α values. Using low values in the parameterization of the fuzzy measures (*e.g.* α equal to 0.4 for Q_1 and 0.1 for Q_2) generates a similarity function that stress the importance of visiting the same places, disregarding the event order. For this reason, the sequence $abcde$ is similar to the sequence $edcba$. In the same way, the sequence $abcxe$ is more similar to $erabc$ (with 4 equal elements) than to $abmnc$ (with only 3 common elements).

Oppositely, when large α values are selected (*e.g.* α equal to 4 for Q_1 and 0.8 for Q_2), the order among the common events plays a more important role in the similarity measurement. Consequently, the sequence $edcba$ is more similar to the sequence $erabc$ than to the sequence $abcde$, because the path between elements a , b and c is exactly the same. Note also that the sequences $abcxe$ and $abmnc$ also change their cluster with regards to the case of low α values.

To balance these two polices, a medium α value can be selected (*e.g.* α equal to 2 for Q_1 and 0.5 for Q_2). In this case, the groups obtained take into account both issues (common events and event order) in a similar way. Notice that in this case, the final groups are different from the ones in the other two cases.

ISBN: 978-989-95079-6-8

The clusters obtained for the different α values are graphically displayed in Figures 3, 4 and 5, for small, medium and large α, respectively. From them, it can be more easily seen that the itineraries are grouped in a different manner depending on the aggregation policy.

To sum up, it can be said that the $d_{OSS-OWA}$ function permits to model different similarity measurement polices. This fact is crucial for data mining analysis and for privacy preserving data mining. If a microaggregation method does not consider the different semantics that the similarity function can have, the partition obtained may put together individuals that should be distinguished. Moreover, if the information of those individuals is then replaced by their prototype values, the posterior analysis of these masked data will not be able to retrieve some relevant knowledge for the user.

5 Conclusions and future work

In this paper, we have introduced a new similarity function for categorical event sequences using OWA operators. We have illustrated how it is possible to introduce expert knowledge inside the similarity computation using well-known fuzzy quantifiers. We have also presented some results about the application of this similarity function to a very recent microaggregation algorithm for categorical event sequences.

As future work we include the analysis of the microaggregation with respect to information loss and disclosure risk measures when it is applied to event sequences. These measures are required to properly evaluate the performance of the microaggregation when it is applied to event sequences and compare it with other approaches. Moreover, the effect of different fuzzy measures must be studied.

Acknowledgment

Partial support by the Spanish MEC (projects ARES – CONSOLIDER INGENIO 2010 CSD2007-00004 – and eAEGIS – TSI2007-65406-C03-02), the Generalitat de Catalunya (grant 2005-SGR-00093) and Universitat Rovira i Virgili (project 2008TURISME-02). Jordi Nin wants to thank the Spanish Council for Scientific Research (CSIC) for his I3P grant.

References

[1] O. Abul, M. Atzori, F. Bonchi, and F. Giannotti. Hiding sequences. In *ICDE Workshops*, pages 147–156. IEEE Computer Society, 2007.

[2] N. Shoval, G. K. Auslander, T. Freytag, R. Landau, F. Oswald, U. Seidl, H.-W. Wahl, S. Werner, and J. Heinik. The use of advanced tracking technologies for the analysis of mobility in alzheimer's disease and related cognitive diseases. *BMC Geriatrics*, 8(7), 2008.

[3] G. Dong and J. Pei. *Sequence Data Mining*. Springer, 2007.

[4] T. G. Dietterich. Machine learning for sequential data: A review. In *Proceedings of the Joint IAPR International Workshop on Structural, Syntactic, and Statistical Pattern Recognition*, pages 15–30, London, UK, 2002. Springer-Verlag.

[5] J. Han and M. Kamber. *Data Mining: Concepts and Techniques*. The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann Publishers, 2nd edition, 2006.

[6] L. Willenborg and T. de Waal. *Elements of Statistical Diclosure Control*, volume 155 of *Lecture Notes in Statistics*. Springer, 2001.

[7] R. Agrawal and R. Srikant. Privacy-preserving data mining. In *SIGMOD '00: Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pages 439–450, 2000.

[8] L. Sweeney. Achieving k-anonymity privacy protection using generalization and suppression. *Int. J. of Unc., Fuzz. and Knowledge Based Systems*, 10(5):571–588, 2002.

[9] L. Sweeney. k-anonymity: a model for protecting privacy. *Int. J. of Unc., Fuzz. and Knowledge Based Systems*, 10(5):557–570, 2002.

[10] D. Defays and P. Nanopoulos. Panels of enterprises and confidentiality: The small aggregates method. In *Proceedings of 92th Symposium on Design and Analysis of Longitudinal Surveys*, pages 195–204. Statistics Canada, Ottawa, 1993.

[11] J. Domingo-Ferrer and V. Torra. A quantitative comparison of disclosure control methods for microdata. In *Confidentiality, disclosure, and data access: theory and practical applications for statistical agencies*, pages 111–133. Elsevier, 2001.

[12] C. Gómez-Alonso and A. Valls. *Lecture Notes on Artificial Intelligence: Modeling Decisions for Artificial Intelligence (MDAI 2008)*, chapter A similarity measure for sequences of categorical data based on the ordering of common elements, pages 134–145. Springer, 2008.

[13] V. Torra and Y. Narukawa. *Modeling decisions: information fusion and aggregation operators*. Springer, 2007.

[14] R. R. Yager. On ordered weighted averaging aggregation operators in multi-criteria decision making. *IEEE Trans. on Systems, Man and Cybernetics*, 18:183–190, 1988.

[15] R. R. Yager. Families of owa operators. *Fuzzy Sets and Systems*, 59:125–148, 1993.

[16] F. Felsö, J. Theeuwes, and G. Wagner. Disclosure limitation in use: Results of a survey. In *Confidentiality, disclosure, and data access: theory and practical applications for statistical agencies*, pages 17–42, 2001.

[17] S. Hansen and S. Mukherjee. A polynomial algorithm for optimal univariate microaggregation. *IEEE Trans. on Kwnoledge and Data Engineering*, 15(4):1043–1044, 2003.

[18] A. Oganian and J. Domingo-Ferrer. On the complexity of optimal microaggregation for statistical disclosure control. *Statistical Journal United Nations Economic Commission for Europe*, 18(4):345–354, 2000.

[19] J. Domingo-Ferrer and J. M. Mateo-Sanz. Practical data-oriented microaggregation for statistical disclosure control. *IEEE Trans. on Kwnoledge and Data Engineering*, 14(1):189–201, 2002.

[20] A. K. Jain, M. N. Murty, and P.J. Flynn. Data clustering: a review. *ACM Computing Surveys*, 31(3):264–323, 1999.

[21] A. Valls and C. Gómez-Alonso. The k-sized hierarchical clustering method for microaggregation. Technical Report DEIM-RR-09-01, Universitat Rovira i Virgili, 2009.

[22] A. Valls, C. Gómez-Alonso, and V. Torra. Generation of prototypes for masking sequences of events. In *3rd Int. Workshop on Advances in Information Security, 4th Int. Conf. on Availability, Reliability and Security (ARES)*, pages 947–952. IEEE Computer Society, 2009.

[23] R. W. Hamming. Error detecting and error correcting codes. *Bell System Technical Journal*, 26(2):147–160, 1950.

[24] V. I. Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, 10(8):707–710, 1966.

ISBN: 978-989-95079-6-8

Application of Fuzzy Vectors of Normalized Weights in Decision Making Models

Ondřej Pavlačka, Jana Talašová

Department of Mathematical Analysis and Applications of Mathematics,
Faculty of Science, Palacký University Olomouc
tř. Svobody 26, 771 46 Olomouc, Czech Republic
Email: {pavlacko,talasova}@inf.upol.cz

Abstract— *In this paper, the notion of a fuzzy vector of normalized weights is introduced, and its application in decision making models based on the weighted average operation is studied. The fuzzy vector of normalized weights represents a generalization of normalized fuzzy weights that authors formerly applied for modeling uncertain weights of criteria and/or uncertain probabilities of states of the world. Illustrative examples show that fuzzy vectors of normalized weights extend the possibilities of utilizing the vague expert information concerning weights of criteria. For three particular forms of the fuzzy vectors of normalized weights, fuzzy weighted averages of given values are calculated and the obtained results are compared. Finally, the fuzzy models of multiple criteria decision making and decision making under risk based on the fuzzy weighted average operation are described.*

Keywords— Decision making under risk; Fuzzy vectors; Fuzzy weighted average; Multiple criteria decision making; Normalized fuzzy weights.

1 Introduction

In decision making models, the weighted average is the most often used aggregation operator. In multiple criteria decision making, overall evaluations of alternatives are mostly calculated as weighted averages of their evaluations with respect to particular criteria. The weighted average operator is also used in discrete stochastic models of decision making under risk where the alternatives are usually ranked according to their expected evaluations. The expected evaluations are in fact weighted averages of evaluations of alternatives under possible states of the world; probabilities of the states of the world play the role of normalized weights.

Decision making models based on the weighted average operation can be fuzzified in two aspects: First, uncertain weighted values, i.e. uncertain evaluations with respect to particular criteria in the models of multiple criteria decision making and/or uncertain evaluations under possible states of the world in decision making under risk, can be considered. Such fuzzification is useful because it enables to process correctly e.g. the vague expert evaluations of alternatives with respect to qualitative criteria. This approach is described in detail e.g. in [1]. Second, normalized weights, i.e. weights of criteria and/or probabilities of states of the worlds, can be fuzzified. It makes the mathematical models substantially more complicated. But as the values of normalized weights are commonly set expertly, such fuzzification is very eligible.

The fuzzy extension of the weighted average operation for the case when both the weighted values and the weights are fuzzy has been studied since the second half of 70's. In [2, 3, 4], the fuzzy weights are supposed to be non-negative fuzzy numbers that express the importance of particular criteria; they are normalized only in the process of the fuzzy weighted average calculation. Later, it was shown in [5, 6] that uncertain weights expressing shares of the partial evaluation in the overall one or uncertain probabilities of states of the world have to be modeled by means of a special structure of fuzzy numbers which is called a tuple of normalized fuzzy weights. The tuple of normalized fuzzy weights is formally identical with a so called feasible tuple of fuzzy probabilities mentioned in [7]. The fuzzy weights normalization, i.e. the transformation of fuzzy weights into the normalized fuzzy weights, was studied in [8, 9].

This paper describes a new approach to modeling the uncertain normalized weights of criteria and/or the uncertain probabilities of states of the world - a notion of a fuzzy vector of normalized weights is introduced here. It will be shown that, in comparison with normalized fuzzy weights, fuzzy vectors of normalized weights extend the possibilities of utilizing the vague expert information concerning the weights.

2 Preliminaries

A fuzzy set A on a non-empty set X is characterized by its membership function $A : X \to [0,1]$. By $Ker\,A$ and $Supp\,A$, we denote a kernel of A, $Ker\,A = \{x \in X \mid A(x) = 1\}$, and a support of A, $Supp\,A = \{x \in X \mid A(x) > 0\}$, respectively. For any $\alpha \in [0,1]$, A_α means an α-cut of A, $A_\alpha = \{x \in X \mid A(x) \geq \alpha\}$. The family of all fuzzy sets on X will be denoted by $\mathcal{F}(X)$.

The intersection of two fuzzy sets $A, B \in \mathcal{F}(X)$ is defined as a fuzzy set $A \cap B \in \mathcal{F}(X)$ whose membership function is for all $x \in X$ given by $(A \cap B)(x) = \min\{A(x), B(x)\}$. Since the minimum operation is used, $(A \cap B)_\alpha = A_\alpha \cap B_\alpha$ holds for all $\alpha \in [0,1]$.

Let $X_1, X_2, \ldots, X_m$ be non-empty sets. An m-ary fuzzy relation on $X_1 \times X_2 \times \ldots \times X_m$ is defined as a fuzzy set $R \in \mathcal{F}(X_1 \times X_2 \times \ldots \times X_m)$. A particular case of an m-ary fuzzy relation is the Cartesian product of fuzzy sets, $R^C = R_1 \times R_2 \times \ldots \times R_m$, where $R_i \in \mathcal{F}(X_i)$, $i = 1, 2, \ldots, m$, and $R^C(x_1, x_2, \ldots, x_m) = \min\{R_1(x_1), R_2(x_2), \ldots, R_m(x_m)\}$ for all $(x_1, x_2, \ldots, x_m) \in X_1 \times X_2 \times \ldots \times X_m$. Since the minimum operation is used, $R^C_\alpha = R_{1\alpha} \times R_{2\alpha} \times \ldots \times R_{m\alpha}$ holds for any $\alpha \in [0,1]$.

Let us denote the index set $\{1, 2, \ldots, m\}$ by N_m throughout the paper. For any $i \in N_m$, the i-th projection of an m-ary

ISBN: 978-989-95079-6-8

fuzzy relation $R \in \mathcal{F}(X_1 \times X_2 \times \ldots \times X_m)$ is the fuzzy set $[R]_i \in \mathcal{F}(X_i)$ whose membership function is for all $x_i \in X_i$ given by

$$[R]_i(x_i) = \sup_{\substack{y_j \in X_j,\ j \in N_m \\ y_i = x_i}} R(y_1, y_2, \ldots, y_m). \qquad (1)$$

A fuzzy number is a fuzzy set C on the set of all real numbers $\mathbb{R}$ that fulfils the following conditions: a) $Ker\ C \neq \emptyset$, b) for all $\alpha \in (0,1]$, C_α are closed intervals, c) $Supp\ C$ is bounded. If $Supp\ C \subseteq [a,b]$, then C is referred to as a fuzzy number on $[a,b]$. The family of all fuzzy numbers will be denoted by $\mathcal{F}_N(\mathbb{R})$, and the family of all fuzzy numbers on $[a,b]$ will be denoted by $\mathcal{F}_N([a,b])$.

In [9, 10], it was shown that any fuzzy number C can be also determined by a couple of functions $\underline{c} : [0,1] \to \mathbb{R}$ and $\overline{c} : [0,1] \to \mathbb{R}$ that describe the minimal and maximal values of the α-cuts of C and of the closure of the support of C. The functions $\underline{c}$ and $\overline{c}$ are left-continuous on $(0,1]$, right-continuous at 0 and satisfy $\underline{c}(\alpha) \leq \underline{c}(\beta) \leq \overline{c}(\beta) \leq \overline{c}(\alpha)$ for all $0 \leq \alpha < \beta \leq 1$. In the sequel, the notation $C = \{[\underline{c}(\alpha), \overline{c}(\alpha)]\}_{\alpha \in [0,1]}$ will be used for the fuzzy number C such that $C_\alpha = [\underline{c}(\alpha), \overline{c}(\alpha)]$ for all $\alpha \in (0,1]$, and $Cl(SuppC) = [\underline{c}(0), \overline{c}(0)]$ where $Cl(SuppC)$ means the closure of the support of C. For the membership function of the fuzzy number $C = \{[\underline{c}(\alpha), \overline{c}(\alpha)]\}_{\alpha \in [0,1]}$, the following relation holds

$$C(x) = \begin{cases} \max\limits_{\alpha \in [0,1]:\ x \in [\underline{c}(\alpha), \overline{c}(\alpha)]} \alpha, & \text{for } x \in [\underline{c}(0), \overline{c}(0)], \\ 0, & \text{elsewhere.} \end{cases} \qquad (2)$$

Let us note that if $\underline{c}(\alpha) = \overline{c}(\alpha) = c$ for all $\alpha \in [0,1]$, then C represents a real number c. Furthermore, if $\underline{c}(\alpha) = a$ and $\overline{c}(\alpha) = b$ for all $\alpha \in [0,1]$, where $a < b$, then C represents a closed interval $[a,b]$.

In this paper, a fuzzy number $C = \{[\underline{c}(\alpha), \overline{c}(\alpha)]\}_{\alpha \in [0,1]}$ is called linear if both the functions $\underline{c}$ and $\overline{c}$ are linear. Any linear fuzzy number C can be fully characterized by a quadruple of real numbers $c^1 \leq c^2 \leq c^3 \leq c^4$, where $Cl(Supp\ C) = [c^1, c^4]$ and $Ker\, C = [c^2, c^3]$. In this case, the functions $\underline{c}$ and $\overline{c}$ are given for all $\alpha \in [0,1]$ as follows

$$\underline{c}(\alpha) = c^1 + \alpha(c^2 - c^1) \quad \text{and} \quad \overline{c}(\alpha) = c^4 - \alpha(c^4 - c^3). \qquad (3)$$

The notation $C = \langle c^1, c^2, c^3, c^4 \rangle$ will be used for such a linear fuzzy number C. For $c^2 \neq c^3$, the linear fuzzy number C is usually referred to as a trapezoidal fuzzy number; for $c^2 = c^3$, as a triangular fuzzy number.

The extension from functions having crisp arguments to functions with fuzzy set arguments is done according to the extension principle. Let X and Y be non-empty sets and let $f : X \to Y$ be a mapping. Then a fuzzy extension of f is the mapping $f_F : \mathcal{F}(X) \to \mathcal{F}(Y)$ such that for any $A \in \mathcal{F}(X)$, the membership function of $f_F(A)$ is defined for all $y \in Y$ as follows

$$f_F(A)(y) = \begin{cases} \sup\limits_{x \in X:\ f(x) = y} A(x), & \text{if } f^{-1}(y) \neq \emptyset, \\ 0, & \text{otherwise.} \end{cases} \qquad (4)$$

3 Fuzzy Vectors

Since an uncertain value of one continuous variable is expressed by a fuzzy number, then, by means of analogy, a fuzzy vector is used for expressing uncertain values of m continuous variables. An m-dimensional fuzzy vector $\mathbf{V}$ is a fuzzy set on $\mathbb{R}^m$ that fulfils the following conditions: a) $Ker\ \mathbf{V} \neq \emptyset$, b) for all $\alpha \in (0,1]$, $\mathbf{V}_\alpha$ are closed and convex subsets of $\mathbb{R}^m$, c) $Supp\ \mathbf{V}$ is bounded. If $Cl(Supp\ \mathbf{V}) \subseteq Q \subset \mathbb{R}^m$, then $\mathbf{V}$ is called a fuzzy vector on Q. Let us remark that the α-cuts $\mathbf{V}_\alpha$ are compact subsets of $\mathbb{R}^m$ for all $\alpha \in (0,1]$. The family of all m-dimensional fuzzy vectors will be denoted by $\mathcal{F}_V(\mathbb{R}^m)$, and the family of all m-dimensional fuzzy vectors on $Q \subset \mathbb{R}^m$ will be denoted by $\mathcal{F}_V(Q)$. Obviously, since a closed and convex subset of $\mathbb{R}$ is a closed interval, for $m = 1$ a fuzzy number is obtained, i.e. $\mathcal{F}_V(\mathbb{R}) = \mathcal{F}_N(\mathbb{R})$.

In literature, the requirement of convexity of $\mathbf{V}_\alpha$ is often weakened, e.g. in [11], the α-cuts of fuzzy vectors are supposed to be closed and simply connected subsets of $\mathbb{R}^m$. As convexity is the characteristic property of expertly given uncertain values, the convexity of α-cuts of fuzzy vectors will be considered in the paper.

For any $i \in N_m$, the i-th projection of the m-dimensional fuzzy vector $\mathbf{V}$ will be denoted by $[\mathbf{V}]_i$. It was proved in [9, 10] that $[\mathbf{V}]_i$ is a fuzzy number and its membership function is defined for all $y \in \mathbb{R}$ by the following formula

$$[\mathbf{V}]_i(y) = \max_{\substack{x_j \in \mathbb{R},\ j \in N_m \\ x_i = y}} \mathbf{V}(x_1, x_2, \ldots, x_m). \qquad (5)$$

In the fuzzy models of decision making discussed in this paper, the expert setting of uncertain input data by fuzzy vectors is based on the following results that were proved in [9, 10].

If $X_1, X_2, \ldots, X_m$ are fuzzy numbers, then the Cartesian product $\mathbf{X} = X_1 \times X_2 \times \ldots \times X_m$ is an m-dimensional fuzzy vector, and $[\mathbf{X}]_i = X_i$ for all $i \in N_m$. Similarly, if $\mathbf{X}_i \in \mathcal{F}_V(\mathbb{R}^{n_i})$ for all $i \in N_m$, then the Cartesian product $\mathbf{X} = \mathbf{X}_1 \times \mathbf{X}_2 \times \ldots \times \mathbf{X}_m$ is an n-dimensional fuzzy vector where $n = \sum_{i=1}^m n_i$.

If the uncertain values of m variables are expressed by fuzzy numbers $X_1, X_2, \ldots, X_m$, and the set of all admissible combinations of their values is given by a crisp nonseparable m-ary relation $Q \subset \mathbb{R}^m$, then the maximal fuzzy relation (in the sense of ordering given by inclusion) expressing available information concerning the values of the m-tuple of variables is given by $\mathbf{X}_Q = (X_1 \times X_2 \times \ldots \times X_m) \cap Q$. If Q is closed and convex, then the fuzzy set $\mathbf{X}_Q$ is an m-dimensional fuzzy vector on Q if and only if

$$(Ker\ X_1 \times Ker\ X_2 \times \ldots \times Ker\ X_m) \cap Q \neq \emptyset. \qquad (6)$$

Moreover, $[\mathbf{X}_Q]_i \subseteq X_i$ for all $i \in N_m$. The equality $[\mathbf{X}_Q]_i = X_i$ holds for all $i \in N_m$ if and only if for all $i \in N_m$ and for all $\alpha \in (0,1]$ the following holds:

$$\text{For any } x_i \in X_{i\alpha},\ \text{there exist } x_j \in X_{j\alpha},\ j \in N_m,\ j \neq i, \\ \text{such that } (x_1, \ldots, x_i, \ldots, x_m) \in Q. \qquad (7)$$

Calculations with fuzzy vectors are based on the following well-known result, the proof of which can be found e.g. in [11]. Let $D \subseteq \mathbb{R}^m$, $m \geq 1$, let $f : D \to \mathbb{R}$ be a continuous function, and let $\mathbf{X} \in \mathcal{F}_V(D)$. Then $f_F(\mathbf{X})$ is a fuzzy number whose membership function is given by

$$f_F(\mathbf{X})(y) = \begin{cases} \max\limits_{\mathbf{x} \in D:\ f(\mathbf{x}) = y} \mathbf{X}(\mathbf{x}), & \text{if } f^{-1}(y) \neq \emptyset, \\ 0, & \text{otherwise,} \end{cases} \qquad (8)$$

ISBN: 978-989-95079-6-8

and for all $\alpha \in (0,1]$ the following holds

$$f_F(\mathbf{X})_\alpha = f(\mathbf{X}_\alpha). \tag{9}$$

Employing the notation $f_F(\mathbf{X}) = \{[\underline{y}(\alpha), \overline{y}(\alpha)]\}_{\alpha\in[0,1]}$, then it follows from (9) that for any $\alpha \in (0,1]$, the values $\underline{y}(\alpha)$ and $\overline{y}(\alpha)$ can be obtained by solving the following problems of mathematical programming

$$\underline{y}(\alpha) = \min_{\mathbf{x}\in\mathbf{X}_\alpha} f(\mathbf{x}), \tag{10}$$

$$\overline{y}(\alpha) = \max_{\mathbf{x}\in\mathbf{X}_\alpha} f(\mathbf{x}). \tag{11}$$

Furthermore, since the functions $\underline{y}(\alpha)$ and $\overline{y}(\alpha)$ are right-continuous at 0, the values $\underline{y}(0)$ and $\overline{y}(0)$ are given as follows

$$\underline{y}(0) = \lim_{\alpha\to 0+} \underline{y}(\alpha) \quad \text{and} \quad \overline{y}(0) = \lim_{\alpha\to 0+} \overline{y}(\alpha). \tag{12}$$

If $\mathbf{X} = X_1 \times X_2 \times \ldots \times X_m$ where $X_i \in \mathcal{F}_N(\mathbb{R})$ for any $i \in N_m$, $X_i = \{[\underline{x}_i(\alpha), \overline{x}_i(\alpha)]\}_{\alpha\in[0,1]}$, then for all $\alpha \in [0,1]$ the following holds

$$\underline{y}(\alpha) = \min_{x_i\in[\underline{x}_i(\alpha),\overline{x}_i(\alpha)],\ i\in N_m} f(x_1, x_2, \ldots, x_m), \tag{13}$$

$$\overline{y}(\alpha) = \max_{x_i\in[\underline{x}_i(\alpha),\overline{x}_i(\alpha)],\ i\in N_m} f(x_1, x_2, \ldots, x_m). \tag{14}$$

If $\mathbf{X}_Q = (X_1 \times X_2 \times \ldots \times X_m) \cap Q$, where the crisp relation $Q \subset \mathbb{R}^m$ and fuzzy numbers $X_i = \{[\underline{x}_i(\alpha), \overline{x}_i(\alpha)]\}_{\alpha\in[0,1]}$, $i \in N_m$, fulfil the condition (6), then the values $\underline{y}(\alpha)$ and $\overline{y}(\alpha)$ can be obtained for all $\alpha \in [0,1]$ as follows

$$\underline{y}(\alpha) = \min_{\substack{x_i\in[\underline{x}_i(\alpha),\overline{x}_i(\alpha)],\ i\in N_m \\ (x_1,x_2,\ldots,x_m)\in Q}} f(x_1, x_2, \ldots, x_m), \tag{15}$$

$$\overline{y}(\alpha) = \max_{\substack{x_i\in[\underline{x}_i(\alpha),\overline{x}_i(\alpha)],\ i\in N_m \\ (x_1,x_2,\ldots,x_m)\in Q}} f(x_1, x_2, \ldots, x_m). \tag{16}$$

The formulas (15) and (16) correspond to the concept of constrained fuzzy arithmetics that was studied e.g. in [12].

4 Fuzzy Vectors of Normalized Weights

An m-dimensional fuzzy vector of normalized weights is an arbitrary m-dimensional fuzzy vector $\mathbf{W}$, whose support $Supp\ \mathbf{W}$ is a subset of the m-dimensional probability simplex $\mathcal{S}_m = \big\{(w_1, w_2, \ldots, w_m) \in \mathbb{R}^m \mid w_i \geq 0 \text{ for all } i \in N_m,\ \sum_{i=1}^m w_i = 1\big\}$. The family of all m-dimensional fuzzy vectors of normalized weights will be denoted by $\mathcal{F}_V(\mathcal{S}_m)$. Figure 1 illustrates an example of a 3-dimensional fuzzy vector of normalized weights.

From the general point of view, an m-dimensional fuzzy vector of normalized weights expresses an uncertain division of a unit into m fraction. In this paper, we will apply the fuzzy vectors of normalized weights to model the uncertain shares of partial objectives of evaluation in the overall one in fuzzy models of multiple criteria decision making, and to model the uncertain probabilities of states of the world in fuzzy-stochastic models of decision making under risk.

In the decision making models, direct expert setting of a fuzzy vector of normalized weights could be quite complicated. Therefore, three procedures of setting of fuzzy vectors of normalized weights will be proposed now.

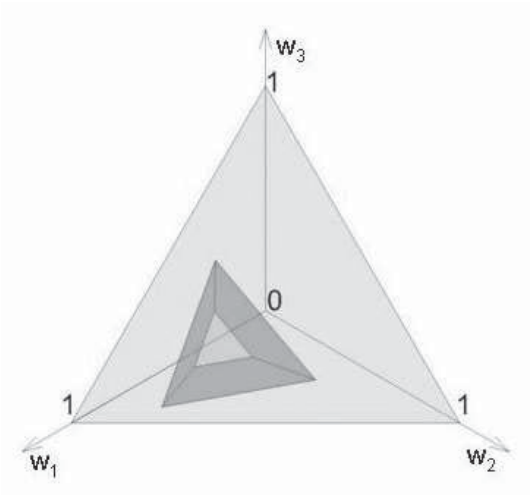


Figure 1: A 3-dimensional fuzzy vector of normalized weights.

Firstly, an expert can describe the uncertain values of normalized weights by a special structure of fuzzy numbers called normalized fuzzy weights. The normalized fuzzy weights are defined as fuzzy numbers $W_1, W_2, \ldots, W_m \in \mathcal{F}_N([0,1])$ that satisfy for all $\alpha \in (0,1]$ and for all $i \in N_m$ the following condition:

$$\text{For any } w_i \in W_{i\alpha},\ \text{there exist } w_j \in W_{j\alpha},\ j \in N_m,\ j \neq i,\ \text{such that } w_i + \sum_{j=1,\ j\neq i}^{m} w_j = 1. \tag{17}$$

The condition (17) represents the special case of (7) for $Q = \mathcal{S}_m$. Since the probability simplex $\mathcal{S}_m$ is closed and convex, the fuzzy set $\mathbf{W} = (W_1 \times W_2 \times \ldots \times W_m) \cap \mathcal{S}_m$ is a fuzzy vector of normalized weights, and $[\mathbf{W}]_i = W_i$ for all $i \in N_m$. The procedures of practical setting of normalized fuzzy weights were described in [6, 9].

Example 1 Let $m = 3$. An expert describes the uncertain values of normalized weights by the triple of linear normalized fuzzy weights $W_1 = \langle 0.4, 0.5, 0.65, 0.8\rangle$, $W_2 = \langle 0.15, 0.2, 0.35, 0.55\rangle$ and $W_3 = \langle 0.05, 0.15, 0.3, 0.45\rangle$. The corresponding fuzzy vector of normalized weights is the fuzzy set $\mathbf{W} = (W_1 \times W_2 \times W_3) \cap \mathcal{S}_3$. For all $\alpha \in (0,1]$, the α-cuts $\mathbf{W}_\alpha$ are given as follows

$$\begin{aligned} \mathbf{W}_\alpha = \big\{&(w_1, w_2, w_3) \in \mathcal{S}_3 \mid \\ &w_1 \in [0.4 + 0.1\alpha, 0.8 - 0.15\alpha], \\ &w_2 \in [0.15 + 0.05\alpha, 0.55 - 0.2\alpha], \\ &w_3 \in [0.05 + 0.1\alpha, 0.45 - 0.15\alpha]\big\}. \end{aligned} \tag{18}$$

The normalized fuzzy weights W_1, W_2, W_3 and the fuzzy vector of normalized weights $\mathbf{W} = (W_1 \times W_2 \times W_3) \cap \mathcal{S}_3$ are depicted in Figure 2. □

Secondly, an expert can give the information about normalized weights by a crisp relation $Q \subseteq \mathcal{S}_m$. If Q is non-empty, closed and convex, then it represents a special type of the fuzzy vector of normalized weights. For any $i \in N_m$, the

ISBN: 978-989-95079-6-8

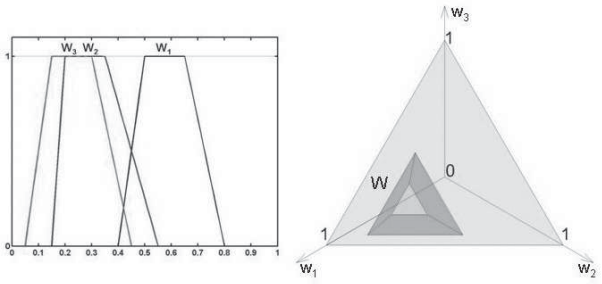


Figure 2: Normalized fuzzy weights W_1, W_2 and W_3 and the fuzzy vector of normalized weights $\mathbf{W} = (W_1 \times W_2 \times W_3) \cap \mathcal{S}_3$.

uncertain value of the i-th weight is expressed by

$$[Q]_i = \big\{w_i \in [0,1] \mid w_i \text{ is the } i\text{-th component of at least one vector } \mathbf{w} \in Q\big\}. \tag{19}$$

Example 2 Let $m = 3$. An expert gives only the ordinal information about the normalized weights by the crisp relation $T = \{(w_1, w_2, w_3) \in \mathcal{S}_3 \mid w_1 \geq w_2 \geq w_3\}$. Since T is closed and convex, T forms a 3-dimensional fuzzy vector of normalized weights $\mathbf{T}$ (see Fig. 3). Obviously, it follows from (19) that $[\mathbf{T}]_1 = [\frac{1}{3}, 1]$, $[\mathbf{T}]_2 = [0, \frac{1}{2}]$ and $[\mathbf{T}]_3 = [0, \frac{1}{3}]$. □

Thirdly, an expert can give both information about the uncertain values of weights, i.e. normalized fuzzy weights $W_1, W_2, \ldots, W_m$, and information about the crisp relations among normalized weights, i.e. the crisp relation $Q \subset \mathcal{S}_m$. If Q is non-empty, closed and convex and if $(Ker\ W_1 \times Ker\ W_2 \times \ldots \times Ker\ W_m) \cap Q \neq \emptyset$, then $\mathbf{W}_Q = (W_1 \times W_2 \times \ldots \times W_m) \cap Q$ is the corresponding fuzzy vector of normalized weights. For any $i \in N_m$, the true uncertain value of the i-th weight is expressed by $[\mathbf{W}_Q]_i \subseteq W_i$; the α-cuts of $[\mathbf{W}_Q]_i$ can be obtained for all $\alpha \in (0,1]$ in the following way

$$[\mathbf{W}_Q]_{i\alpha} = \big\{w_i \in W_{i\alpha} \mid (W_{1\alpha} \times \ldots \times W_{i-1\alpha} \times \{w_i\} \times W_{i+1\alpha} \times \ldots \times W_{m\alpha}) \cap Q \neq \emptyset\big\}. \tag{20}$$

Example 3 Let $m = 3$. First, an expert gives the ordinal information about the normalized weights by a crisp relation $T = \{(w_1, w_2, w_3) \in \mathcal{S}_3 \mid w_1 \geq w_2 \geq w_3\}$. Later on, he/she adds an uncertain cardinal information by a triple of linear normalized fuzzy weights $W_1 = \langle 0.4, 0.5, 0.65, 0.8\rangle$, $W_2 = \langle 0.15, 0.2, 0.35, 0.55\rangle$ and $W_3 = \langle 0.05, 0.15, 0.3, 0.45\rangle$. As $W_1 \geq W_2 \geq W_3$, i.e. $W_{1\alpha} \geq W_{2\alpha} \geq W_{3\alpha}$ for all $\alpha \in (0,1]$, it holds that $(KerW_1 \times KerW_2 \times KerW_3) \cap T \neq \emptyset$. The fuzzy set $\mathbf{W}_T = (W_1 \times W_2 \times W_3) \cap T$ represents the corresponding fuzzy vector of normalized weights (see Fig. 3). For all $\alpha \in (0,1]$, the following holds

$$\begin{aligned}\mathbf{W}_{T\alpha} = \big\{&(w_1, w_2, w_3) \in \mathcal{S}_3 \mid w_1 \geq w_2 \geq w_3,\\ &w_1 \in [0.4 + 0.1\alpha, 0.8 - 0.15\alpha],\\ &w_2 \in [0.15 + 0.05\alpha, 0.55 - 0.2\alpha],\\ &w_3 \in [0.05 + 0.1\alpha, 0.45 - 0.15\alpha]\big\}.\end{aligned} \tag{21}$$

According to (20), the projections of $\mathbf{W}_T$ are given as follows: $[\mathbf{W}_T]_1 = W_1$, $[\mathbf{W}_T]_2 = \{[\underline{w}_{T2}(\alpha), \overline{w}_{T2}(\alpha)]\}_{\alpha\in[0,1]}$, where $\underline{w}_{T2}(\alpha) = 0.15 + 0.05\alpha$, $\overline{w}_{T2}(\alpha) = 0.475 - 0.05\alpha$ for $0 \leq \alpha \leq 0.5$ and $\overline{w}_{T2}(\alpha) = 0.55 - 0.2\alpha$ for $0.5 < \alpha \leq 1$, and $[\mathbf{W}_T]_3 = \langle 0.05, 0.15, 0.25, 0.3\rangle$.

Let us note that although the fuzzy numbers W_1, W_2 and W_3 satisfy $W_1 \geq W_2 \geq W_3$, the crisp relation T represents an additional information for the model. Since the normalized fuzzy weights W_1, W_2 and W_3 satisfy with respect to T only the condition (6), but not (7), the projections $[\mathbf{W}_T]_1, [\mathbf{W}_T]_2$ and $[\mathbf{W}_T]_3$ are less uncertain then the normalized fuzzy weights W_1, W_2 and W_3. □

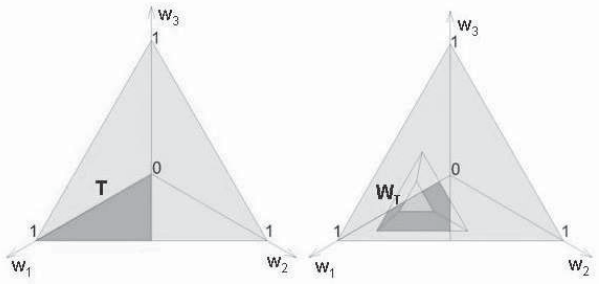


Figure 3: Fuzzy vectors of normalized weights $\mathbf{T} = \{(w_1, w_2, w_3) \in \mathcal{S}_3 \mid w_1 \geq w_2 \geq w_3\}$ and $\mathbf{W}_T = (W_1 \times W_2 \times W_3) \cap T$.

5 Fuzzy Weighted Average of Fuzzy Numbers with a Fuzzy Vector of Normalized Weights

A weighted average of real numbers $u_1, u_2, \ldots, u_m$ with a vector of normalized weights $\mathbf{w} = (w_1, w_2, \ldots, w_m)$ is defined by the following formula

$$a(\mathbf{w}, u_1, u_2, \ldots, u_m) = \sum_{i=1}^{m} w_i u_i. \tag{22}$$

Since the weighted average operation is a continuous real function defined on $\mathcal{S}_m \times \mathbb{R}^m$, its fuzzy extension a_F has the property that $a_F(\mathbf{X})$ is a fuzzy number for all $\mathbf{X} \in \mathcal{F}_V(\mathcal{S}_m \times \mathbb{R}^m)$. In decision making models, the vector of normalized weights is usually supposed to be independent of the weighted values, and the weighted values to be mutually independent as well. Therefore, only the fuzzy vectors $\mathbf{X} = \mathbf{W} \times U_1 \times U_2 \times \ldots \times U_m$, where $\mathbf{W} \in \mathcal{F}_V(\mathcal{S}_m)$ and $U_i \in \mathcal{F}_N(\mathbb{R})$ for all $i \in N_m$, will be considered in this paper.

According to (8), the fuzzy weighted average of fuzzy numbers $U_1, U_2, \ldots, U_m$ with a fuzzy vector of normalized weights $\mathbf{W}$ is a fuzzy number $U = a_F(\mathbf{W}, U_1, U_2, \ldots, U_m)$ whose membership function is for all $u \in \mathbb{R}$ given by the formula

$$U(u) = \max_{\substack{\mathbf{w}\in\mathcal{S}_m \\ u_i\in\mathbb{R},\, i\in N_m \\ \sum_{i=1}^m w_i\cdot u_i = u}} \min\big\{\mathbf{W}(\mathbf{w}), U_1(u_1), \ldots, U_m(u_m)\big\}. \tag{23}$$

Since the weighted average operation is an increasing function in variables $u_1, u_2, \ldots, u_m$, it was proved in [9] that the general mathematical programming problems (10) and (11) can be simplified. Let us denote $U_i = \{[\underline{u}_i(\alpha), \overline{u}_i(\alpha)]\}_{\alpha\in[0,1]}$ for any $i \in N_m$. For the fuzzy weighted average U, $U =$

ISBN: 978-989-95079-6-8

$\{[\underline{u}(\alpha), \overline{u}(\alpha)]\}_{\alpha\in[0,1]}$, the values of the functions $\underline{u}$ and $\overline{u}$ can be calculated for all $\alpha \in (0, 1]$ as follows

$$\underline{u}(\alpha) = \min_{(w_1,w_2,\ldots,w_m)\in \mathbf{W}_\alpha} \sum_{i=1}^{m} w_i \cdot \underline{u}_i(\alpha), \qquad (24)$$

$$\overline{u}(\alpha) = \max_{(w_1,w_2,\ldots,w_m)\in \mathbf{W}_\alpha} \sum_{i=1}^{m} w_i \cdot \overline{u}_i(\alpha). \qquad (25)$$

Hence, complexity of the calculation of U depends only on the form of the α-cuts $\mathbf{W}_\alpha$.

It is worth noting that in the case of fuzzy weighted average of fuzzy numbers with normalized fuzzy weights $W_1, W_2, \ldots, W_m$, i.e. when $\mathbf{W} = (W_1 \times W_2 \times \ldots \times W_m) \cap \mathcal{S}_m$, it is not necessary to solve the mathematical programming problems (24) and (25). For such a case, the following effective algorithm for computing the values $\underline{u}(\alpha)$ and $\overline{u}(\alpha)$, $\alpha \in [0, 1]$, was described in [6]. Let us denote $W_i = \{[\underline{w}_i(\alpha), \overline{w}_i(\alpha)]\}_{\alpha\in[0,1]}$ for any $i \in N_m$. For each $\alpha \in [0, 1]$, let $\{i_k\}_{k=1}^m$ be such a permutation on an index set N_m that $\underline{u}_{i_1}(\alpha) \le \underline{u}_{i_2}(\alpha) \le \ldots \le \underline{u}_{i_m}(\alpha)$. For $k \in N_m$, let us denote

$$w_{i_k}(\alpha) = 1 - \sum_{j=1}^{k-1} \overline{w}_{i_j}(\alpha) - \sum_{j=k+1}^{m} \underline{w}_{i_j}(\alpha). \qquad (26)$$

Let $k^* \in N_m$ be such an index that $\underline{w}_{i_{k^*}}(\alpha) \le w_{i_{k^*}}(\alpha) \le \overline{w}_{i_{k^*}}(\alpha)$. Then

$$\underline{u}^N(\alpha) = \sum_{j=1}^{k^*-1} \overline{w}_{i_j}(\alpha) \cdot \underline{u}_{i_j}(\alpha) + w_{i_{k^*}}(\alpha) \cdot \underline{u}_{i_{k^*}}(\alpha) + \sum_{j=k^*+1}^{m} \underline{w}_{i_j}(\alpha) \cdot \underline{u}_{i_j}(\alpha). \qquad (27)$$

Let $\{i_h\}_{h=1}^m$ be such a permutation on an index set N_m that $\overline{u}_{i_1}(\alpha) \ge \overline{u}_{i_2}(\alpha) \ge \ldots \ge \overline{u}_{i_m}(\alpha)$. For $h \in N_m$, let us denote

$$w_{i_h}(\alpha) = 1 - \sum_{j=1}^{h-1} \overline{w}_{i_j}(\alpha) - \sum_{j=h+1}^{m} \underline{w}_{i_j}(\alpha). \qquad (28)$$

Let $h^* \in N_m$ be such an index that $\underline{w}_{i_{h^*}}(\alpha) \le w_{i_{h^*}}(\alpha) \le \overline{w}_{i_{h^*}}(\alpha)$. Then

$$\overline{u}^N(\alpha) = \sum_{j=1}^{h^*-1} \overline{w}_{i_j}(\alpha) \cdot \overline{u}_{i_j}(\alpha) + w_{i_{h^*}}(\alpha) \cdot \overline{u}_{i_{h^*}}(\alpha) + \sum_{j=h^*+1}^{m} \underline{w}_{i_j}(\alpha) \cdot \overline{u}_{i_j}(\alpha). \qquad (29)$$

Example 4 Let us consider the fuzzy vectors of normalized weights $\mathbf{W}$ from Example 1, $\mathbf{T}$ from Example 2 and $\mathbf{W}_T$ from Example 3. Let the weighted values be, for simplicity, given by real numbers $u_1 = 0.3$, $u_2 = 0.1$ and $u_3 = 0.9$. Let us denote $U_W = a_F(\mathbf{W}, u_1, u_2, u_3)$, $U_T = a_F(\mathbf{T}, u_1, u_2, u_3)$ and $U_{W_T} = a_F(\mathbf{W}_T, u_1, u_2, u_3)$.

Since $\mathbf{W} = (W_1 \times W_2 \times W_3) \cap \mathcal{S}_3$, where W_1, W_2, W_3 are normalized fuzzy weights, by the special algorithm mentioned above we obtain that $U_W = \langle 0.22, 0.32, 0.44, 0.54\rangle$.

As $\mathbf{T} = \{(w_1, w_2, w_3) \in \mathcal{S}_3 \mid w_1 \ge w_2 \ge w_3\}$ is a crisp relation on $\mathcal{S}_3$ and weighted values are real numbers, the mathematical programming problems (24) and (25) does not depend on α. Therefore, the fuzzy weighted average U_T is equal to a closed interval, $U_T = \{[\underline{u}_T, \overline{u}_T]\}_{\alpha\in[0,1]}$. The values $\underline{u}_T = 0.2$ and $\overline{u}_T = 0.433$ were obtained by solving the following linear programming problems

$$\underline{u}_T = \min_{\substack{\sum_{i=1}^3 w_i=1 \\ w_1-w_2\ge 0 \\ w_2-w_3\ge 0 \\ w_3\ge 0}} 0.3w_1 + 0.1w_2 + 0.9w_3, \qquad (30)$$

$$\overline{u}_T = \max_{\substack{\sum_{i=1}^3 w_i=1 \\ w_1-w_2\ge 0 \\ w_2-w_3\ge 0 \\ w_3\ge 0}} 0.3w_1 + 0.1w_2 + 0.9w_3. \qquad (31)$$

The fuzzy weighted average of u_1, u_2 and u_3 with the fuzzy vector of normalized weights $\mathbf{W}_T$, $\mathbf{W}_T = (W_1 \times W_2 \times W_3) \cap T$, is the fuzzy number $U_{W_T} = \{[\underline{u}_{W_T}(\alpha), \overline{u}_{W_T}(\alpha)]\}_{\alpha\in[0,1]}$ where

$$\underline{u}_{W_T} = \min_{\substack{w_i\in W_{i\alpha},\ i\in N_3 \\ \sum_{i=1}^3 w_i=1 \\ w_1-w_2\ge 0 \\ w_2-w_3\ge 0 \\ w_3\ge 0}} 0.3w_1 + 0.1w_2 + 0.9w_3, \qquad (32)$$

$$\overline{u}_{W_T} = \max_{\substack{w_i\in W_{i\alpha},\ i\in N_3 \\ \sum_{i=1}^3 w_i=1 \\ w_1-w_2\ge 0 \\ w_2-w_3\ge 0 \\ w_3\ge 0}} 0.3w_1 + 0.1w_2 + 0.9w_3. \qquad (33)$$

From (32) and (33), we obtain that $\underline{u}_{W_T}(\alpha) = 0.235 + 0.07\alpha$ for $0 \le \alpha < 0.5$, $\underline{u}_{W_T}(\alpha) = 0.22 + 0.1\alpha$ for $0.5 \le \alpha \le 1$, and $\overline{u}_{W_T}(\alpha) = 0.42 - 0.02\alpha$.

Fig. 4 shows the fuzzy weighted averages U_W, U_T and U_{W_T}. □

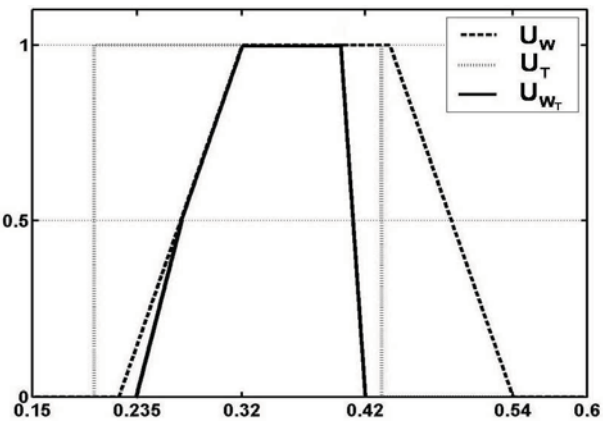


Figure 4: Fuzzy weighted averages U_W, U_T and U_{W_T}.

6 Fuzzy Models of Decision Making Based on the Fuzzy Weighted Average

In this section, a multiple criteria decision making model and a model of decision making under risk, where the fuzzy weighted average of fuzzy numbers with fuzzy vector of normalized weights is applied, will be described.

ISBN: 978-989-95079-6-8

First, let us consider a problem of multiple criteria decision making, where the best of alternatives $x_1, x_2, \ldots, x_n$ is to be chosen. The alternatives are being evaluated with respect to a given overall objective that is partitioned into m partial objectives associated with criteria $C_1, C_2, \ldots, C_m$. Let the uncertain information about the shares of the partial objectives in the overall one be given by an m-dimensional fuzzy vector of normalized weights $\mathbf{W}$. Let uncertain partial fuzzy evaluations of alternatives x_i, $i \in N_n$, with respect to the criteria C_j, $j \in N_m$, be expressed by fuzzy numbers $U_{i,j}$ defined on $[0,1]$ that represent the fuzzy degrees of satisfaction of the corresponding partial objectives of evaluation.

Then the overall fuzzy evaluation U_i of the alternative x_i, $i \in N_n$, can be expressed as follows

$$U_i = a_F(\mathbf{W}, U_{i,1}, U_{i,2}, \ldots, U_{i,m}). \tag{34}$$

The overall fuzzy evaluations $U_1, U_2, \ldots, U_n$ express the fuzzy degrees of satisfaction of the overall objective of evaluation. The best alternative is the first alternative in an ordering of the fuzzy numbers $U_1, U_2, \ldots, U_n$ or the closest to the ideal alternative whose evaluation is equal to 1. For more details on metrics and ordering of fuzzy numbers see e.g. [13]. The overall fuzzy evaluations of alternatives can also be approximated linguistically by linearly ordered elements of a proper linguistic evaluation scale defined on $[0,1]$ (see [1]).

Second, let us consider a problem of decision making under risk that is described by the following fuzzy decision matrix (see Table 1), where $x_1, x_2, \ldots, x_n$ are alternatives, $S_1, S_2, \ldots, S_r$ states of the world, $\mathbf{P}$ is an r-dimensional fuzzy vector of normalized weights that expresses the uncertain probabilities of the states of the world, and for any $i \in N_n$ and $k \in N_r$, $U_{i,k} \in \mathcal{F}_N([0,1])$ denotes the fuzzy degree in which the alternative x_i satisfies a given decision objective if the state S_k occurs.

Table 1: Fuzzy decision matrix.

	S_1	S_2	...	S_r	$\mathbf{P}$
x_1	$U_{1,1}$	$U_{1,2}$	...	$U_{1,r}$	FEU_1
x_2	$U_{2,1}$	$U_{2,2}$	...	$U_{2,r}$	FEU_2
...	...	...	...	...	...
x_n	$U_{n,1}$	$U_{n,2}$	...	$U_{n,r}$	FEU_n

It was shown in [5] that the fuzzy expected value of evaluations of alternatives $x_1, x_2, \ldots, x_n$ are for all $i \in N_n$ expressed by

$$FEU_i = a_F(\mathbf{P}, U_{i,1}, U_{i,2}, \ldots, U_{i,r}). \tag{35}$$

The best alternative is usually determined by the rule of the maximum fuzzy expected value of evaluation. The maximum fuzzy expected value of evaluation is selected from $FEU_1, FEU_2, \ldots, FEU_n$ analogously as it was shown in the case of multiple criteria decision making.

A similar approach can be applied also to multiple criteria decision making under risk (see [14]).

7 Conclusion

In this paper, a new fuzzification of the weighted average operation where uncertain normalized weights are modeled by a general fuzzy vector defined on the probability simplex was presented. Several ways of expert setting the fuzzy vector of normalized weights in fuzzy models were proposed. It was shown that, in comparison with a tuple of normalized fuzzy weights, the fuzzy vector of normalized weights extends the possibilities of modeling the expert's knowledge concerning the weights. Besides the formal definition of the fuzzy weighted average of fuzzy numbers with a fuzzy vector of normalized weights, also a general algorithm for computing the fuzzy weighted average was presented; its three particular forms were illustrated by examples. Finally, applications of the fuzzy weighted average operation in models of multiple criteria decision making and decision making under risk were described.

References

[1] J. Talašová. NEFRIT - multicriteria decision making based on fuzzy approach. *Central European Journal of Operations Research*, 8(4):297–319, 2000.

[2] S. Baas and H. Kwakernaak. Rating and ranking of multiple-aspect alternatives using fuzzy sets. *Automatica*, 13:47–58, 1977.

[3] W. M. Dong and F. S. Wong. Fuzzy weighted averages and implementation of the extension principle. *Fuzzy Sets and Systems*, 21:183–199, 1987.

[4] Y.-Y. Guh, Ch.-Ch. Hon, and E. S. Lee. Fuzzy weighted average: The linear programming approach via charness and cooper's rule. *Fuzzy Sets and Systems*, 117:157–160, 2001.

[5] J. Talašová and O. Pavlačka. Fuzzy probability spaces and their applications in decision making. *Austrian Journal of Statistics*, 35(2&3):347–356, 2006.

[6] O. Pavlačka and J. Talašová. Application of the fuzzy weighted average of fuzzy numbers in decision making models. *New Dimensions in Fuzzy Logic and Related Technologies (Eds.: Štěpnička, M., Novák, V., Bodenhofer, U.). Proceedings of the 5th EUSFLAT Conference, Ostrava, Czech Republic, September 11-14, 2007*, *II*:455–462, 2007.

[7] Y. Pan and B. Yuan. Bayesian inference of fuzzy probabilities. *Int. J. General Systems*, 26(1–2):73–90, 1997.

[8] Y.-M. Wang and T. M. S. Elhag. On the normalization of interval and fuzzy weights. *Fuzzy Sets and Systems*, 157:2456–2471, 2006.

[9] O. Pavlačka. *Fuzzy methods of decision making (in Czech)*. Dissertation thesis, Palacký University, Olomouc, 2007.

[10] O. Pavlačka and J. Talašová. Fuzzy vectors as a tool for modeling uncertain multidimensional quantities. *Submited to Fuzzy Sets and Systems in December 2008*.

[11] R. Viertl. *Statistical Methods for Non-Precise Data*. CRC Press, Boca Raton, Florida, 1996.

[12] G. J. Klir and Y. Pan. Constrained fuzzy arithmetic: basic questions and some answers. *Soft Computing*, 2:100–108, 1998.

[13] D. Dubois and H. Prade. *Fundamentals of fuzzy sets*. Kluwer Academic Publishers, Boston-London-Dordrecht, 2000.

[14] J. Talašová. Fuzzy sets in decision-making under risk (in Czech). *Proceedings of the 4th Conference Aplimat (Eds. M. Kováčová), STU, Bratislava*, pages 532–544, 2005.

ISBN: 978-989-95079-6-8

Integrating Complex User Preferences into a Route Planner: A Fuzzy-Set-Based Approach

Amine Mokhtari Olivier Pivert Allel Hadjali

IRISA/ENSSAT, Université de Rennes 1
6, Rue de Kerampont, BP 80518, 22305 Lannion Cedex, France
Email: {mokhtari, pivert, hadjali}@enssat.fr

Abstract— Route planners are systems which help users select a route between two locations. In such a context, personalization mechanisms notably aim at taking into account user preferences so as to identify the best route(s) among a set of possible answers. In this paper, we present a framework based on fuzzy set theory for modeling complex user preferences. We provide a typology of preferences which make sense in the domain considered and outline a query language integrating such flexible features. The paper also includes a discussion about query evaluation.

Keywords— Route planning, preferences, fuzzy sets.

1 Introduction

During the last century, and particularly the last fifty years, the impact of transportation means on our daily lives has constantly increased. However, it is only recently, with the tremendous development of the Internet, that an alternative solution to the use of paper maps has appeared to help users choose a route. Route planners are systems which aim at computing the "best" route between a location A and a location B, using different types of information about a given road network (structural information, of course, but also, in some cases, real-time data about traffic jams, road conditions, etc).

Even though there now exists a relatively large set of commercially available route planners, these systems have very limited capabilities — if any — when it comes to taking into account sophisticated user preferences.

The need for personalization has been felt for several years in domains such as databases or information retrieval, but not much research effort has been devoted to customizing route planners so far [20]. Currently, in the most "intelligent" commercial systems and research prototypes, only a small set of predefined preferences (e.g., avoid turnpikes, prefer freeways) are made available to the user who can sometimes attach weights to them.

In this paper, we describe a fuzzy-set-based approach to the modeling of sophisticated user preferences concerning a route planning task. Fuzzy set theory has already been used to deal with some transportations problems; in particular route choice[1], where several models based on fuzzy sets have been proposed in order to take into account uncertainty and the subjectivity of the user [17, 12].

[1]For a given origin-destination pair and a given transport mode, the route choice problem (also called traffic assignment) deals with identifying which route a given traveler *would take*. It represents the fourth step in the conventional transportation forecasting model, that aims at estimating the number of vehicles or travelers that will use a specific transportation facility in the future.

The choice of this theoretical framework is motivated by several reasons, in particular:

- the fact that it is very well suited to the interpretation of linguistics terms, which constitute a convenient way for a user to express his/her preferences,
- the fact that fuzzy set theory relies on a commensurability assumption which makes it possible to aggregate several preferences regarding different attributes, thus leading to a complete pre-order.

The remainder of the paper is organized as follows. After a survey of related work in Section 2, we deal with data representation issues in Section 3. In Section 4, we present a typology of user preferences in the context of unimodal point-to-point route planning. The main features of a query language involving such user preferences are presented in Section 5. Query evaluation is discussed in Section 6. The conclusion recalls the main contributions and outlines some perspectives for future work.

2 Related work

Building a personalized route planner is not an easy task, for several reasons. The large size and complexity of modern road networks, and the number of possible preferences present many a difficulty to overcome.

In the last decade, a few propositions have been made to personalize route planners. Some of them exploit a driver's route history to learn individual preferences that can then be used in future planning tasks. Most of the approaches assume that preferences are not given explicitly due to the difficulty of their modeling and elicitation [13].

Liu [14] proposes a route planning system which combines knowledge about the road network with case-based reasoning and brute-force search. He describes how geographical knowledge isolates the search for useful route segments to a local map region. The approach makes the strong assumption that users prefer routes that follow major roads, and the planning algorithm explicitly seeks out major roads to form the plan of a target route.

Rogers and Langley [18] propose a route planning system which can learn preferences from user feedback. During each interactive session, the user is asked to express his/her preferences among recommended routes. The feedback resulting from this interaction is used as the training data for a perceptron-style training algorithm. The authors assume a fixed user preference model which only concern route length,

ISBN: 978-989-95079-6-8

driving time and turn angles. A numeric weight can be associated to the corresponding preferences.

Let us also mention the work by McGinty and Smyth [15] about a case-based route planning approach. The system they describe generates routes which reflect implicit preferences of individual users. The main aspect that distinguishes this system from that proposed by Rogers and Langley [18] is the fact that it does not assume a fixed preference model. Every user preferences are represented as a collection of previous route cases that the user considered satisfactory. Thus, new routes are generated by reusing and combining relevant sections of multiple cases.

Letchner et al. [13] propose a route planner, named TRIP, which produces route plans that more closely match the routes chosen by people who have extensive experience travelling within a region. To do so, it incorporates time-variant road speeds learned from large amounts of driver-collected GPS data. It also exploits a driver's past GPS logs when responding to future route queries in order to provide routes which are more suited to the driver's individual driving preferences.

The work most similar to our proposal is that by Balke et al. [4, 2, 3]. These authors propose a route planning system integrating user preferences over four characteristics of a route: length, traffic jams, road works and weather conditions. To aggregate the scores related to these preferences into an overall degree, a weighted function F is used. The weights are user-defined and express the importance of each of the route characteristics. The user may use a five-level linguistic scale for asserting the importance, and the levels are automatically mapped onto numerical weights w_i. In response to a route planning query, the top k results are delivered to the user.

The main aspect which distinguishes our approach from the works [14, 15, 18] is that we choose — like Balke et al. [4, 2, 3] — to *explicitly* model user preferences. On the other hand, unlike Balke et al., we aim at taking into account a wide range of atomic user preferences, and fuzzy-set theory provides us with a very rich set of connectors for combining such preferences, leading to a highly expressive query language.

3 Representation issues

Modeling user preferences requires a clear representation of concepts, relations and geographic entities which can be involved in a route planning query. Different spatial data models have been proposed in the literature, and road networks can be represented in several ways. This section provides some elements on these topics.

3.1 Road network modeling

A road network is generally modeled as a (directed) graph. Several granularity levels can be considered, according to what the components of the graph (vertices and edges) represent. In the case of an interurban road network, a vertex represents a city, and an edge describes an interurban link (freeway, national roads, etc). However, in a urban road network, a vertex represents an intersection between two or more routes, and an edge corresponds to a street.

In our work, we consider the finest granularity: a vertex represents any intersection/bifurcation between routes or dead ends, and an edge corresponds to a road of any kind.

3.2 Geographic Data Files (GDF)

We assume a *a priori* that any spatial feature may be concerned by a user preference. Thus, interpreting and processing queries involving such a wide range of potential preferences imply to have available a model suited to the representation of these features. Geographic Data Files (GDF) [10] is international standard that specifies the conceptual and logical data model for geographic databases for Intelligent Transportation Systems (ITS) applications. It includes a specification of potential contents of such databases (Features, Attributes and Relationships), a specification of how these contents shall be represented, and of how relevant information about the database itself can be specified (meta data).

3.3 Time and its representation

We consider the following model of time: time is linear (precedence) and dense (for any two time elements there is always a third element between them). We assume time representation based on time points. A time point is represented by a real number.

4 Typology of user preferences

In the context of unimodal, point-to-point, route planning query, we distinguish three families of user preferences:

spatial preferences p_s they express preferences about roads, places or parts of the road network;

spatio-temporal preferences p_{st} they are spatial preferences involving a time component which expresses the moment or period when the spatial preference is relevant.

intrinsic preferences p_{int} they concern some global properties of a route, such as comfort, length, duration or safety.

In the next subsection, we present each family of user preferences in more detail.

4.1 Spatial preferences

Spatial preferences represent the largest class of preference in terms of diversity. Every entity of the road network may be concerned by such a preference.

Definition (Spatial preferences): a preference p is said to be spatial if it concerns a geographic entity, i.e., an entity which has spatial coordinates.

Two categories of spatial preferences may be identified: *i)* preferences on one (or several) specific element(s) of a road network, *ii)* preferences on an induced part of a road network (induced subgraph). The first category corresponds to preferences on one or several vertices/edges of a road network, i.e., on intersection(s) (resp. section(s)) of a road network. This category is used by a user to favour (respectively, disfavour) some specific elements of the network. An example is "avoid turnpikes". The second category corrresponds to preferences which concern an induced part of a road network, that we call a zone in the following. A zone is a geographic area, for instance an administrative area. Preferences on a zone are about the environment that surrounds a route. An example is "prefer a route which passes across Brittany". Two aspects of spatial entities are worthy of discussion. The first one concerns the boundary of a zone. While some zones have clear-cut (crisp)

ISBN: 978-989-95079-6-8

boundaries, as administrative areas, states or regions, others have fuzzy boundaries, such as "a polluted zone". The second aspect concerns the durability of some spatial entities. Depending on what it represents, the definition of a zone can evolve with time. It is the case, for instance, of zones defined in terms of weather conditions such as "a foggy area".

4.1.1 Spatial relationships

Two approaches exist for designating road network elements concerned by spatial user preferences. The explicit approach consists in designating the different entities concerned by the preference by means of their inherent, individual, characteristics. On the other hand, the implicit approach uses references to other spatial entities, by means of spatial relations. An example is "prefer routes which have gas stations along them".

Definition (Spatial relationship): A spatial relationship is relationship between two or more spatial objects.

Several primitive relationships exist. They can be classified into the following non-exhaustive subcategories [16]:

Topological relationships such as *adjacent*, *inside*, *disjoint*, they are invariant under topological transformations like translation, scaling or rotation.

Spatial (strict) order relationships they are based on the definition of an order, either large or strict: *left to/right to*, *above/below*, *behind/in front of*, *near/far*, *inside/outside* or *surround* [11].

Metric relationships they are based on the concept of distance in the considered space.

Directional relationships these relationships are based on an (absolute/relative) frame of reference. Cardinal direction relationships that describe orientation in space (e.g. *north*, *northeast*) are the most commonly used such relationships.

Several mathematical frameworks have been proposed for each type of spatial relations, independently or in association with one another, notably based on fuzzy set theory [5].

In a route planning query with preferences, different spatial relationships may be used. Many of them may be interpreted in a fuzzy manner while some are intrinsically crisp [21].

4.2 Spatio-temporal preferences

Time has an important place in route planning since the duration of a trip is in general a major factor as to the satisfaction of the user. Thus, it is essential to enable the user to express his/her preferences regarding this aspect.

The time dimension may appear with many semantics according to the problem to be represented. As mentioned before, in a spatio-temporal preference, time is used to express the validity period of a spatial preference. An example is "avoid the city center around noon". In other words, a spatio-temporal preference involves both spatial and temporal entities. Thus, a spatio-temporal preference p_{st} can be seen as a complex preference, formed of a spatial preference p_s and a temporal preference p_t. In the following subsections we concentrate on the temporal part of spatio-temporal preferences. We discuss the different types of temporal entities and relations that they may involve.

4.2.1 Temporal entities and representation

Three types of temporal statements may be involved in a route planning query:

- statements referring to a specific point in time, called instant, denoted by i.
- statements which make reference to a continuous time period with duration, called interval, denoted by I.
- statements which make reference to a duration (i.e., the absolute distance between two instants), denoted by d.

A classical numeric representation of time is the set of real numbers $\mathbb{R}$ [1]. Thus, an instant i is considered to be a singleton representing a date/hour (for instance: 8pm). A (time/duration) interval is described by a lower and an upper bound (beginning instant, ending instant) $[i^-, i^+]$ or a beginning instant and a duration) $[i^-, i^- + d]$.

In many cases, users tend to employ linguistic terms such as *around noon*, *early morning*, to express their temporal preferences. Fuzzy set theory provides a suitable symbolic/numeric interface to the representation of such preferences.

4.2.2 Qualitative/Quantitative temporal preferences

A route planning query may contain two types of temporal preferences: quantitative and qualitative ones.

Quantitative preferences express absolute bounds or restrict the temporal distance between two instants [6]. In other terms, they express preferences on the duration of events or their timing. Two types of quantitative preferences can be distinguished:

- unary preferences, which express a constraint on an instant i by means of a set of intervals, and are expressed as: $(i \in I_1) \vee (i \in I_2) \vee \ldots \vee (i \in I_n)$;
- binary preferences, which concern two instants i_1 and i_2 and constrain the distance $i_2 - i_1$: $(i_2 - i_1) \in I_1 \vee (i_2 - i_1) \in I_2 \vee \ldots \vee (i_2 - i_1) \in I_n$.

Example 1. Find a route from Paris to Rennes such that the duration of the trip is less than three hours.□

Qualitative preferences provide a means to specify the relative position of a pair of temporal entities (instants or intervals) t_1 and t_2 [1]. A qualitative temporal preference is expressed as $t_1 \, R \, t_2$, such that: $R \in \Re$ where $\Re$ is a finite set of basic temporal relations. The set $\Re$ contains the following types of relations:

- point-to-point relations: $<$, $\leq$, $=$, $>$, $\geq$.
- interval-to-interval relations between $I = [i^-, i^+]$ and $J = [j^-, j^+]$: *after/before* $(i^+ < j^-)$, *overlaps* $(i^- < j^- < i^+ < j^+)$, *equals* $(i^- = j^- \wedge i^+ = j^+)$.
- point-to-interval and interval-to-point relations between a point i and an interval $I = [i^-, i^+]$: *after/before* $(i < i^-)$, *during/contains* $(i^- < i < i^+)$.

ISBN: 978-989-95079-6-8

Example 2. Find a route from Paris to Berlin, with the preference "to arrive before night". Processing such a query requires to compare the arrival time with the linguistic term *night*.□

Notice that in the framework considered, temporal criteria intervene in spatio-temporal preferences, i.e., are related to a given spatial entity. In the example above, the criterion about the arrival time is related to the spatial entity "Berlin".

Let us also mention the existence of some theoretical work about the fuzzification of temporal relations [7].

4.3 Intrinsic preferences

A route can be seen as a particular geographic entity within the framework we propose. Contrary to spatial and spatio-temporal preferences which can be called *local* inasmuch as they concern a part of a route, intrinsic preferences characterize the route *as a whole*. They involve qualitative criteria such as expensive, fast, safe, and so on. Among the most intuitive properties of a route, let us mention: *rapidity*, *length*, *safety*, *cost* and *comfort*. Other properties may be of interest in some specific contexts, for example: robustness (for military applications or rescue services) or pollution produced.

5 Outline of a query language

In the previous section, we have presented the different kinds of preferences, elementary entities and possible relations involved in a route planning query with preferences. These elements constitute the basic components of a query language dedicated to route planning. In the sequel, we outline the syntax of this language as well as the different semantics that a route query can convey.

5.1 Query formalization

Let $Q^{\delta_d}_{\delta_a}(P)$ be a unimodal point-to-point route planning query with user preferences where: $\delta_d = (s_d,\ t_d)$ represents departure parameters (departure place s_d and departure time t_d), $\delta_a = (s_a,\ t_a)$ represents arrival parameters, and $P = p_1 \otimes p_2 \otimes \ldots \otimes p_n$ is a compound preference where $\otimes$ stands for a fuzzy connector (conjunction, mean, etc) and p_i is an atomic user preference. Two types of atomic preferences may be distinguished: those which concern the route as a whole, and those which concern the segment of a route. The first type takes the forms:

- $attribute\ \theta\ constant$, where θ is a crisp or fuzzy comparator, e.g., duration $\approx$ 5 hours.
- $attribute\ is\ fuzzy_term$, e.g., cost is *high*.

The second type may take the following forms:

- $[\exists \mid \forall]$ RoadElement $R\ Entity\ \{temporal\ constraint\}$ where R is a spatial relation. An example is: prefer a route passing next to at least one gas station: $\exists$ RoadElement near gas_station. Another example is: avoid the city center between 11am and 2pm: $\forall$ RoadElement outside city_center when $t \in$ [11am, 2pm].
- $[\exists \mid \forall]$ RoadElement.attribute $\theta\ value$ where θ is a comparator and $value$ is a (linguistic or numeric) constant. For instance, avoid highways: $\forall$ RoadElement.type $\neq$ 'highway'. Another example is: prefer a route with no portion with a speed limit under 60mph: $\forall$ RoadElement.speedlimit ≥ 60.

A first query representation that could be envisaged is:

GO	**FROM**	δ_d
	TO	δ_a
	PREFERENCES	P

Example 3. Find a fast, comfortable, and inexpensive route from Rennes to Paris, with a departure at 16:00, and with the constraint that the cost of the trip must be low. This query may be formulated as follows:

GO	**FROM**	Rennes, 16:00
	TO	Paris
	PREFERENCES	duration is short **and** comfort is high **and** cost is low□

However, this representation does not distinguish between constraints and wishes (in the spirit of the concept of bipolarity [8]). Consequently, in the example above, a route can be selected even though its cost is almost not low at all, as soon as it has a very short duration and a very high level of comfort. To avoid this, user preferences may be partitioned into two sets representing two distinct components of user requirements. The first one, introduced by the keyword *preferring*, gathers the preferences, denoted by P^+, which describe the wishes of the user. The second component, introduced by the keyword *avoiding*, denoted by P^-, specifies the constraints, i.e., the characteristics of routes considered as unacceptable or undesirable by the user. The following formulation makes explicit the bipolar nature of route planning queries:

GO	**FROM**	δ_d
	TO	δ_a
	PREFERRING	P^+
	AVOIDING	P^-

Thus, the previous route query can be expressed as:

GO	**FROM**	Rennes, 16:00
	TO	Paris
	PREFERRING	duration is short **and** comfort is high
	AVOIDING	cost is high.

5.2 Route query semantics

Let $Q^{\delta_d}_{\delta_a}(P)$ be a route query. The parameter t_d (resp. t_a) may be an instant or an interval. Depending on the value of t_d (resp. t_a), such a query may take two different interpretations: static or dynamic.

Static queries (SQ): consist in computing the relevant routes between s_d and s_a for a set of preferences P, at a given instant t_d (resp. t_a). Such a query involves either t_d or t_a or none of them, but not both. If the query includes t_a, an answer will provide t_d, and reciprocally.

Dynamic queries (DQ): the parameter t_d (resp. t_a) is an interval. Such a query specifies either t_a or t_d. The objective is not only to return the relevant routes, but also to determine the best moment for the trip. An answer is thus a triple (departure time, arrival time, route). Here, the satisfaction degree associated with an answer depends on the state of the network at the moment considered (traffic jams, construction works, etc) thus takes into account dynamic data.

From a processing point of view, a dynamic query $Q^{\delta_d}_{\delta_a}(P)$ boils down to a set of static ones. Assuming a discrete time

ISBN: 978-989-95079-6-8

representation, an interval t_d (resp. t_a) corresponds to a finite set of instants $\{i_1, \ldots, i_n\}$ (resp. $\{i'_1, \ldots, i'_m\}$). Evaluating Q with $\delta_d = (s_d,\ t_d)$ (resp. $\delta_a = (s_a,\ t_a)$) comes down to evaluate $Q^{\delta'_d}_{\delta'_a}(P)$ with $\delta'_d = (s_d, t_i)$ for $i = 1..n$, (resp. with $\delta'_a = (s_a, t_{i'})$ for $i' = 1..m$), and to keep the best answer(s).

6 Route query processing

In this section, we describe the evaluation procedure of a route query with preferences. Two aspects are tackled: *i)* how to define the semantics (in terms of trapezoidal membership functions, t.m.f.[2]) of the fuzzy sets which model preferences; *ii)* how to proceed for finding the best k answers. For the former, two cases are considered:

Absolute predicates The semantics of such predicates is explicitly given by the user. It is assumed that a route planning system is endowed with an interface that allows the user to define the t.m.f. of such fuzzy terms in a convenient way. Examples of conditions involving absolute predicates are: departure time is $around_2pm$, RoadElement is $close_to$ seashore.

Relative predicates Examples of such predicates are: *as fast as possible*, *as short as possible*. The definition of relative predicates (i.e., their t.m.f.) is automatically computed by the system on the basis of a set of values (that we call the *context*) returned by another query (for example, the context of the relative predicates present in a user query $Q = (\delta_d,\ \delta_a,\ P)$ may be defined as the result of the query which returns the top-k' shortest routes between δ_d and δ_a without taking into account the preferences P). The idea is to define the t.m.f. of a relative predicate using the minimum, average and maximum values of the context, as illustrated in the example given further.

As to query evaluation, finding a best path in a graph w.r.t. a set of preferences is known as a multi-objective shortest path problem, where preferences represent the set of objectives to optimize. Depending on the preference model used, the goal can then be either to compute the set of non dominated routes (preference model based on Pareto order where the preferences are not commensurable) or the k best answers (as in the fuzzy-set-based preference model used here). In any case, computing the answer $\mathcal{R}$ to such a query is an NP-hard problem [19]. In the following, we propose a method which leads to an approximation of $\mathcal{R}$ in an acceptable time.

Let us consider a route planning query $Q^{\delta_d}_{\delta_a}(P)$ and assume that the user is interested in the k best route plans. The evaluation procedure proposed involves four steps:

Step 1 compute the k' shortest paths ($k' > k$) from departure place δ_d to arrival place δ_a. This can be done using one of the classical algorithms from the literature [9]. Let $\mathcal{R}_{k'}$ be the set of resulting routes;

Step 2 build the t.m.f. of the relative predicates involved in $Q^{\delta_d}_{\delta_a}(P)$ using $\mathcal{R}_{k'}$ as their context; it is assumed that "length is short" is an implicit constraint present in every query. So, the list of relative predicates always includes $short_d$ (related to the length of the route);

[2] A t.m.f is expressed by a quadruplet (A, B, a, b), where $[A, B]$ defines the core of the fuzzy set and $[A - a, B + b]$ its support.

Step 3 evaluate each route $r \in S_{k'}$ w.r.t. each atomic user preference present in $Q^{\delta_d}_{\delta_a}(P)$;

Step 4 aggregate the preference degrees into a single score (or two: one for the constraints, one for the wishes, depending on how bipolarity is handled), sort $\mathcal{R}_{k'}$ according to this score (or these two scores) and return the k best answers.

Example 4. Let us consider a person who wants to go from city A to city B, with a departure at 16:00 and an arrival around 21:00 (wish), preferably by a route which is fast (wish) and not expensive (constraint). The route query Q is:

GO	**FROM**	A, 16:00
	TO	B, around_21h
	PREFERRING	duration is short
	AVOIDING	cost is high

The fuzzy predicates involved in this query are:

- $around_21h$: It is a user-defined (absolute) predicate. Its semantics may be given by the following trapezoidal membership function (t.m.f.): (20:30, 21:30, 1:00, 1:00).
- $short_t$: It is a relative predicate. Its t.m.f. is obtained the following way: *i)* one computes the duration of each route plan in $\mathcal{R}_{k'}$; *ii)* denoting by d_{min} and d_{avg} the minimum and the average duration respectively, then the t.m.f. associated with the predicate $short_t$ is $(d_{min}, d_{min}, 0, d_{avg} - d_{min})$.
- $high$: It is also a relative predicate. Its t.m.f. is built as follows: *i)* for each route $r \in \mathcal{R}_{k'}$, one computes the cost of the trip; *ii)* denoting by c_{avg} and c_{max} the average and the maximum of the computed costs, then the t.m.f. associated to the predicate "*high*" is $(c_{max}, \infty, c_{max} - c_{avg}, 0)$.
- implicit relative predicate $short_d$ (about length) is defined the same way as $short_t$ using the length instead of the duration.□

So as to obtain a unique scalar index for rank-ordering the routes from $\mathcal{R}_{k'}$ according to their overall satisfaction degrees w.r.t. Q, one may merge the satisfaction degree μ_{Pos} coming from the positive preferences (i.e., the wishes) with the degree μ_{Neg} related to the negative preferences (i.e., the constraints) the following way [8]:

$$\mu_Q(r) = min(\mu_{Neg}(r),\ \lambda.\mu_{Neg}(r) + (1 - \lambda).\mu_{Pos}(r))$$

where $\lambda \in]0, 1]$ is a parameter expressing some trade-off between μ_{Pos} and μ_{Neg}. In the case of our example, we have:

$$\mu_{Neg}(r) = min(1 - \mu_{high}(r.cost),\ \mu_{short_d}(r.length)),$$
$$\mu_{Pos}(r) = max(\mu_{short_t}(r.duration),\ \mu_{around_21h}(r.t_a)).$$

In case two degrees are kept, one may rank the answers using the lexicographic order (with a priority given to the constraints, which means that the wishes are only used to break ties).

One may think that in general the value k specified by the user will be ≤ 5. A choice that we think *a priori* reasonable for k' is $4 \times k$, but experimentations will have to be performed

ISBN: 978-989-95079-6-8

so as to determine which is the "best" choice for k'. The value of k' should be big enough to include the most interesting routes, and small enough to have a reasonable computational cost.

7 Conclusion

In this paper, we have outlined a fuzzy-set-based approach to the modeling and handling of complex user preferences in the context of route planning. A typology of preferences covering a variety of intuitive preferences has been proposed and discussed. We have also provided the basis of a query language dedicated to route planning on the one hand, and described the main steps of the evaluation of a route query, on the other hand.

This is still a preliminary work, and many perspectives exist for future research. First, we intend to provide a more detailed specification of the formal query language which was just outlined here. An aspect strongly connected with this aspect, and which is worthy of investigation, is that of a user-friendly interface, which is all the more crucial as fuzzy predicates are to be handled. Also, a user study is necessary to determine the preferences which really make sense in practice (i.e the most relevant ones). As to query evaluation, it would be of interest to devise optimization mechanisms aimed at reducing the search space further. In particular, it is worth investigating whether some pruning criteria could be infered from non-intrinsic preferences, which would enable computing $\mathcal{R}_{k'}$ more efficiently. Our final objective, of course, is to implement a prototype based on the principles described here and to assess its performances by means of experimentations.

References

[1] J.F. Allen. Maintaining knowledge about temporal intervals. *Commun. ACM*, 26(11):832–843, 1983.

[2] W.-T. Balke, W. Kießling, and C. Unbehend. Performance and quality evaluation of a personalized route planning system. In *Proc. of the Brazilian Symposium on Databases (SBBD)*, pages 328–340, 2003.

[3] W.-T. Balke, W. Kießling, and C. Unbehend. Personalized services for mobile route planning: a demonstration. In *Proc. of the 19th International Conference on Data Engineering*, pages 771–773, 2003.

[4] W.-T. Balke, W. Kießling, and C. Unbehend. A situation-aware mobile traffic information system. In *Proc. of the 36th Annual Hawaii International Conference on System Sciences (HICSS'03)*, Washington, DC, USA, 2003.

[5] I. Bloch. Fuzzy spatial relationships for image processing and interpretation: a review. *Image and Vision Computing*, 23(2):89–110, 2005.

[6] R. Dechter, I. Meiri, and J. Pearl. Temporal constraint networks. *Artif. Intell.*, 49(1-3):61–95, 1991.

[7] D. Dubois, A. Hadjali, and H. Prade. A possibility theory-based approach to the handling of uncertain relations between temporal points. *Int. J. Intell. Syst.*, 22(2):157–179, 2007.

[8] D. Dubois and H. Prade. *Bipolarity in Flexible Querying*, volume 2522/2002 of *Lecture Notes in Computer Science*, pages 174–182. Springer Berlin / Heidelberg, 2002.

[9] D. Eppstein. Finding the k Shortest Paths. In *Proc. of the Annual Symposium on Foundations of Computer Science*, volume 35, pages 154–154, 1994.

[10] International Organization for Standardization. Intelligent transport systems – geographic data files (gdf), 2007.

[11] J. Freeman. The modelling of spatial relations. *Computer Graphics and Image Processing*, 4(2):156–171, 1975.

[12] E. Frejinger. *Route choice analysis: data, models, algorithms and applications*. PhD thesis, Lausanne, 2008.

[13] J. Letchner, J. Krumm, and E. Horvitz. Trip router with individualized preferences (trip): Incorporating personalization into route planning. In *Proc. of the 18th Innovative Applications of Artificial Intelligence Conference (IAAI 2006)*, 2006.

[14] B. Liu. Intelligent route finding: Combining knowledge, cases and an efficient search algorithm. In *Proc. of the 12th European Conference on Artificial Intelligence*, pages 380–384, 1996.

[15] L. McGinty and B. Smyth. Personalized route planning: A case-based approach. In *Proc. of the 5th European Workshop on Advances in Case-Based Reasoning*, pages 431–442. Springer-Verlag, 2000.

[16] D. Pullar and M.J. Egenhofer. Toward formal definitions of topological relations among spatial objects. In *Proc. of the 3rd International Symposium on Spatial Data Handling*, pages 225–241, 1988.

[17] M. Ridwan. Fuzzy preference based traffic assignment problem. *Transportation Research Part C: Emerging Technologies*, 12(3-4):209 – 233, 2004.

[18] S. Rogers and P. Langley. Personalized driving route recommendations. In *Proc. of the American Association of Artificial Intelligence Workshop on Recommender Systems*, pages 96–100, 1998.

[19] P. Serafini. Some considerations about computational complexity for multiobjective combinatorial problems. In J. Jahn and W. Krabs, editors, *Recent advances and historical development of vector optimization*, volume 294 of *Lecture Notes in Economics and Mathematical Systems*, Berlin, 1986. Springer-Verlag.

[20] T.C. Son and E. Pontelli. Planning with preferences using logic programming. *Theory and Practice of Logic Programming*, 6(5):559–607, 2006.

[21] H. Tveite. *Data Modelling and Database Requirements for Geographical Data*. Phd thesis, Norwegian Institute of Technology, 1997.

ISBN: 978-989-95079-6-8

Basic Properties of L-fuzzy Quantifiers of the Type $\langle 1 \rangle$ Determined by Fuzzy Measures

Antonín Dvořák, Michal Holčapek

Institute for Research and Applications of Fuzzy Modeling, University of Ostrava
Ostrava, Czech Republic
Email: {antonin.dvorak,michal.holcapek}@osu.cz

Abstract— The aim of this paper is to study monadic **L**-fuzzy quantifiers of the type $\langle 1 \rangle$ determined by fuzzy measures. These fuzzy quantifiers are defined using a novel notion of $\otimes$-fuzzy integral. Several semantic properties of these **L**-fuzzy quantifiers are studied.

Keywords— fuzzy integral, fuzzy logic, fuzzy measure, fuzzy quantifier, generalized quantifier

1 Introduction

This paper studies semantic properties of one class of monadic **L**-fuzzy quantifiers [1, 2] by studying one specific but important class of them, namely fuzzy quantifiers of the type $\langle 1 \rangle$ determined by fuzzy measures.

Quantifiers of the type $\langle 1 \rangle$ are denotations of important noun phrases of natural language, e.g. "something" in "Something is broken.", "everyone" in "Everyone likes Bob.", "nobody" in "Nobody knows everything.", etc. Moreover, classical logical quantifiers "for all" and "there exists" also belong to this type. It is claimed (e.g. in [3]) that from the point of view of natural language semantics, quantifiers of the type $\langle 1, 1 \rangle$ (e.g. "every" in "Every book has leaves.", "most" in "Most birds fly.") are more basic and more important. However, it is usual and advantageous to start with the type $\langle 1 \rangle$ quantifiers, because they are simpler and there are important relationships between them and quantifiers of the type $\langle 1, 1 \rangle$.

Generalized quantifiers evolved, from pioneering works of Mostowski [4], Lindström [5], Barwise and Cooper [6], into quite large research field with deep results. For overview as well as new results see a recent monograph [3]. A quantifier of the type $\langle 1 \rangle$ is usually modeled, given a universe M, as a mapping $Q_M : \mathcal{P}_M \longrightarrow \{true, false\}$ (or, equivalently, as subsets of the power set $\mathcal{P}_M$). It is possible to introduce many properties of (models of) quantifiers, characterizing their behavior from various points of view, for example permutation invariance (PI), isomorphism invariance (ISOM), extension (EXT), and others, see Section 4.

When we think about a definition and properties of generalized quantifiers (like e.g. *many*, *a few* and others), we feel that their truth values should not change abruptly if we gradually change cardinalities of corresponding sets of objects. Consider for example a sentence "Many people read books." If the number of people who read books increase by 1, it would be very strange if the truth value of this sentence changes from false to true. Therefore, it was inevitable that researchers started to consider more than two truth values in this context, and so-called *fuzzy quantifiers* emerged, starting from a generalization of definition from the previous paragraph, where instead of $\{true, false\}$ we consider some other structure of truth values, notably the interval $[0, 1]$.

Research in the field of fuzzy quantifiers started with works of Zadeh [7], Thiele [8], Ralescu [9] and others, see also [10, 11, 12]. An important contribution was made by Hájek in [13]. A comprehensive study of fuzzy quantifiers was undertaken by Glöckner [1] (see also [14]). In the recent paper [15], Novák studies so-called intermediate quantifiers, mainly from the syntactic point of view in the frame of fuzzy type theory [16]. An attempt to model linguistic quantifiers by fuzzy (Sugeno) integral was presented by Ying in [17].

The semantic interpretation of many generalized quantifiers is connected to measurement of "size" of sets in concern. Consider e.g. quantifier "many". The truth value of a proposition "many books have red cover" clearly depends on the "size" of the set of red books. Therefore, it is natural to consider measures (and integrals) of (fuzzy) sets as natural tools for the modeling of important classes of monotonically non-decreasing and monotonically non-increasing generalized quantifiers.

Fuzzy measures and integrals ([18], see also [19, 20]) are important tools allowing us to compare sets with respect to their size. Standardly, fuzzy measures are set functions defined on some algebra of sets which are monotone with respect to inclusion and they assign zero to the empty set. In our approach, fuzzy measures are defined on algebras of fuzzy sets (fuzzy measure spaces) and, generally, they attain values from a complete residuated lattice **L**. Details can be found in [21] in these proceedings, here we present only basic ideas.

Two types of fuzzy integral, namely the $\otimes$-fuzzy integral and the $\rightarrow$-fuzzy integral, are defined on an arbitrary fuzzy measure space. Integrals of $\otimes$ type will be used as models of quantifiers like *all* and *some*, while integrals of $\rightarrow$ type as models of *no* and *not all*, etc. However, in this contribution we will concentrate only on $\otimes$-fuzzy integrals and quantifiers defined by means of them. Nevertheless, if the structure of truth values **L** is a complete MV-algebra, then it is possible to define the $\rightarrow$-fuzzy integral from the $\otimes$-fuzzy integral [21].

2 Preliminaries

For details we refer to our contribution [21] in this volume, here we review only a few necessary notions.

2.1 Structures of truth values

In this paper, we suppose that the structure of truth values is a *complete residuated lattice* (see e.g. [22]), i.e., an algebra $\mathbf{L} = \langle L, \wedge, \vee, \rightarrow, \otimes, \bot, \top \rangle$ with four binary operations and

ISBN: 978-989-95079-6-8

two constants such that $\langle L, \wedge, \vee, \bot, \top\rangle$ is a complete lattice, where $\bot$ is the least element and $\top$ is the greatest element of L, respectively, $\langle L, \otimes, \top\rangle$ is a commutative monoid (i.e., $\otimes$ is associative, commutative and the identity $a \otimes \top = a$ holds for any $a \in L$) and the adjointness property is satisfied, i.e.,

$$a \leq b \to c \quad \text{iff} \quad a \otimes b \leq c \tag{1}$$

holds for each $a, b, c \in L$, where $\leq$ denotes the corresponding lattice ordering. The operations $\otimes$ and $\to$ are usually called multiplication and residuum, respectively.

2.2 L-fuzzy sets

Let $\mathbf{L} = \langle L, \wedge, \vee, \to, \otimes, \bot, \top\rangle$ be a complete residuated lattice and M be a universe of discourse (possibly empty). A mapping $A : M \to L$ is called an **L**-*fuzzy set on* M. A value $A(m)$ is called a *membership degree of* m *in the* **L**-*fuzzy set* A. The set of all **L**-fuzzy sets on M is denoted by $\mathcal{F}_{\mathbf{L}}(M)$. Obviously, if $M = \emptyset$, then the empty mapping $\emptyset$ is the unique **L**-fuzzy set on $\emptyset$ and thus $\mathcal{F}(\emptyset) = \{\emptyset\}$. An **L**-fuzzy set A on M is called *crisp*, if there is a subset X of M such that $A = 1_X$, where 1_X denotes the characteristic function of X. Particularly, $1_\emptyset$ denotes the empty **L**-fuzzy set on M, i.e., $1_\emptyset(m) = \bot$ for any $m \in M$. This convention will be also kept for $M = \emptyset$. The set of all crisp **L**-fuzzy sets on M is denoted by $\mathcal{P}_{\mathbf{L}}(M)$. An **L**-fuzzy set A is *constant*, if there is $c \in L$ such that $A(m) = c$ for any $m \in M$. For simplicity, a constant **L**-fuzzy set is denoted by the corresponding element of L, e.g., a, b, c.[1]

Let A be an **L**-fuzzy set on M. The *complement* of A is an **L**-fuzzy set $\overline{A}$ on M defined by $\overline{A}(m) = \neg A(m)$ for any $m \in M$. Finally, an extension of the operations $\otimes$ and $\to$ on L to the operations on $\mathcal{F}_{\mathbf{L}}(M)$ is given by

$$(A \otimes B)(m) = A(m) \otimes B(m) \tag{2}$$

$$(A \to B)(m) = A(m) \to B(m) \tag{3}$$

for any $A, B \in \mathcal{F}_{\mathbf{L}}(M)$ and $m \in M$, respectively. A mapping $f^{\to} : \mathcal{F}_{\mathbf{L}}(M) \to \mathcal{F}_{\mathbf{L}}(M')$ defined by $f^{\to}(A)(m) = \bigvee_{m' \in f^{-1}(m)} A(m')$ is called a *fuzzy extension* of the mapping f. Obviously, if f is a bijective mapping, then $f^{\to}(A)(f(m)) = A(m)$ for any $m \in M$.

3 Fuzzy measures and integrals

In this section, we will review a notion of a fuzzy measure of **L**-fuzzy sets and of a $\otimes$-fuzzy integral that will be used to define **L**-fuzzy quantifiers. For details and examples see [21], where also definitions of a complementary fuzzy measure and a $\to$-fuzzy integral can be found. For more information about fuzzy integrals, we refer to [18, 19].

3.1 Fuzzy measures of L-fuzzy sets

For our purposes we will consider algebras of **L**-fuzzy sets as a base for defining fuzzy measures of **L**-fuzzy sets.

Definition 3.1 ([18])**.** Let M be a non-empty universe of discourse. A subset $\mathcal{M}$ of $\mathcal{F}_{\mathbf{L}}(M)$ is an *algebra of* **L**-*fuzzy sets on* M, if the following conditions are satisfied

(i) $1_\emptyset, 1_M \in \mathcal{M}$,

(ii) if $A \in \mathcal{M}$, then $\overline{A} \in \mathcal{M}$,

(iii) if $A, B \in \mathcal{M}$, then $A \cup B \in \mathcal{M}$.

A couple $(M, \mathcal{M})$ is called a *fuzzy measurable space*, if $\mathcal{M}$ is an algebra of **L**-fuzzy sets on M.

Let us introduce the concept of fuzzy measure as follows. The first definition is a modification of the definition of a normed measure with respect to truth values (see e.g. [19, 20]).

Definition 3.2. Let $(M, \mathcal{M})$ be a fuzzy measurable space. A mapping $\mu : \mathcal{M} \to L$ is called a *fuzzy measure* on $(M, \mathcal{M})$, if

(i) $\mu(1_\emptyset) = \bot$ and $\mu(1_M) = \top$,

(ii) if $A, B \in \mathcal{M}$ such that $A \subseteq B$, then $\mu(A) \leq \mu(B)$.

A triplet $(M, \mathcal{M}, \mu)$ is called the *fuzzy measure space*, if $(M, \mathcal{M})$ is a fuzzy measurable space and μ is a fuzzy measure on $(M, \mathcal{M})$.

Example 3.1. Let **L** be a complete residuated lattice with the support $[0, 1]$ and $\mathbb{N}$ be the set of natural numbers with 0. For any non-empty countable (finite or denumerable) universe M, injective mapping $f : M \to \mathbb{N}$, $n \in \mathbb{N}$ and $A \in \mathcal{F}_{\mathbf{L}}(M)$, denote

$$A_{f,n}(m) = \begin{cases} A(m), & \text{if } f(m) \leq n; \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Further, for any injective mapping $f : M \to \mathbb{N}$ and $n \in \mathbb{N}$, define $\mu_{f,n} : \mathcal{F}_{\mathbf{L}}(M) \to [0, 1]$ as follows

$$\mu_{f,n}(A) = \frac{\sum_{m \in \mathrm{Supp}(A_{f,n})} A_{f,n}(m)}{|\mathrm{Supp}(1_{M_{f,n}})|} \tag{5}$$

and, finally, define $\underline{\mu}_f, \overline{\mu}_f : \mathcal{F}_{\mathbf{L}}(M) \to [0, 1]$ as follows

$$\underline{\mu}_f = \liminf_{n \to \infty} \mu_{f,n}(A), \tag{6}$$

$$\overline{\mu}_f = \limsup_{n \to \infty} \mu_{f,n}(A). \tag{7}$$

It is easy to see that $\mu_{f,n}$, $\underline{\mu}_f$ and $\overline{\mu}_f$ are fuzzy measures on $(M, \mathcal{F}_{\mathbf{L}}(M))$ determined by an injective mapping f. If, for example, $M = \mathbb{N}$ and $f = \mathrm{id}$, then $\underline{\mu}_f(A) = \overline{\mu}_f(A) = \bot$ for any **L**-fuzzy set on a finite universe. For the set of all even or odd numbers, both fuzzy measures give $\frac{1}{2}$ and, for the set of all prime numbers, we obtain 0.

If M is finite, then $\underline{\mu}_f = \underline{\mu}_g = \overline{\mu}_f = \overline{\mu}_g$ for any injective mappings $f, g : M \to \mathbb{N}$ and

$$\underline{\mu}_f(A) = \overline{\mu}_f(A) = \frac{\sum_{m \in M} A(m)}{|M|}. \tag{8}$$

Let $(M, \mathcal{M})$ be a fuzzy measurable space and $X \in \mathcal{F}_{\mathbf{L}}(M)$. Denote $\mathcal{M}_X$ the set of all $\mathcal{M}$-measurable sets which are contained in X, i.e.,

$$\mathcal{M}_X = \{A \mid A \in \mathcal{M} \text{ and } A \subseteq X\}. \tag{9}$$

Note that $1_\emptyset \in \mathcal{M}_X$ for each $X \in \mathcal{F}_{\mathbf{L}}(M)$ and if X is $\mathcal{M}$-measurable set, then also $X \in \mathcal{M}_X$. If $X = M$, then we will write only $\mathcal{M}$ instead of $\mathcal{M}_M$.

In the following part we will define an isomorphism between fuzzy measure spaces.

[1] We suppose that the meaning of this symbol will be unmistakable from the context, that is, it should be clear when an element of L is considered and when a constant **L**-fuzzy set is assumed.

ISBN: 978-989-95079-6-8

Definition 3.3. Let $(M, \mathcal{M})$ and $(M', \mathcal{M}')$ be fuzzy measurable spaces. We say that a mapping $g : \mathcal{M} \to \mathcal{M}'$ is an *isomorphism between* $(M, \mathcal{M})$ *and* $(M', \mathcal{M}')$, if

(i) g is a bijective mapping with $g(1_\emptyset) = 1_\emptyset$,

(ii) $g(A \cup B) = g(A) \cup g(B)$ and $g(\overline{A}) = \overline{g(A)}$ hold for any $A, B \in \mathcal{M}$,

(iii) there exists a bijective mapping $f : M \to M'$ with $A(m) = g(A)(f(m))$ for any $A \in \mathcal{M}$ and $m \in M$.

Definition 3.4. Let $(M, \mathcal{M})$ and $(M', \mathcal{M}')$ be fuzzy measurable spaces. We say that a mapping $g : \mathcal{M} \to \mathcal{M}'$ is an *isomorphism between* $(M, \mathcal{M}, \mu)$ *and* $(M', \mathcal{M}', \mu')$, if

(i) g is an isomorphism between $(M, \mathcal{M})$ and $(M', \mathcal{M}')$,

(ii) $\mu(A) = \mu'(g(A))$.

If g is an isomorphism between fuzzy measure spaces $(M, \mathcal{M}, \mu)$ and $(M', \mathcal{M}', \mu')$, then we write $g(M, \mathcal{M}, \mu) = (M', \mathcal{M}', \mu')$.

Let $[(M, \mathcal{M}, \mu)]$ denote the class of all fuzzy measure spaces defined on M that are isomorphic with $(M, \mathcal{M}, \mu)$. Obviously, we can write

$$[(M, \mathcal{M}, \mu)] = \{(M, f^{\rightarrow}(\mathcal{M}), \mu_{f^{\rightarrow}}) \mid \\ \mid f : M \to M \text{ is a bijective mapping}\}.$$

3.2 $\otimes$-fuzzy integral

Definition 3.5. Let $(M, \mathcal{M}, \mu)$ be a fuzzy measure space, $A \in \mathcal{F}_{\mathbf{L}}(M)$ and $X \in \mathcal{M}$. The *$\otimes$-fuzzy integral of A on X* is given by

$$\int_X^{\otimes} A \, d\mu = \bigvee_{Y \in \mathcal{M}_X \setminus \{1_\emptyset\}} \bigwedge_{m \in \mathrm{Supp}(Y)} (A(m) \otimes \mu(Y)). \quad (10)$$

If $X = 1_M$, then we write $\int^{\otimes} A \, d\mu$.

Theorem 3.1. *Let $(M, \mathcal{M}, \mu)$ be a fuzzy measure space. If $X \in \mathcal{M}$ is such that $1_{\mathrm{Supp}(Y)} \in \mathcal{M}_X$ for any $Y \in \mathcal{M}_X$, then for any $A \in \mathcal{F}_{\mathbf{L}}(M)$*

$$\int_X^{\otimes} A \, d\mu = \bigvee_{1_Y \in \mathcal{P}_X \setminus \{1_\emptyset\}} \bigwedge_{m \in Y} (A(m) \otimes \mu(1_Y)), \quad (11)$$

where $\mathcal{P}_X = \{1_{\mathrm{Supp}(Z)} \mid Z \in \mathcal{M}_X\}$.

Theorem 3.2. *Let $\mathbf{L}$ be a complete MV-algebra, $(M, \mathcal{M}, \mu)$ be a fuzzy measure space, $A \in \mathcal{F}_{\mathbf{L}}(M)$ and $X \in \mathcal{M}$. Then*

$$\int_X^{\otimes} A \, d\mu = \bigvee_{Y \in \mathcal{M}_X \setminus \{1_\emptyset\}} \big(\mu(Y) \otimes \bigwedge_{m \in \mathrm{Supp}(Y)} A(m)\big). \quad (12)$$

Moreover,

$$\int_X^{\otimes} (c \otimes A) \, d\mu = c \otimes \int_X^{\otimes} A \, d\mu \quad (13)$$

for any $c \in L$.

Theorem 3.3. *Let g be an isomorphism between fuzzy measure spaces $(M, \mathcal{M}, \mu)$ and $(M', \mathcal{M}', \mu')$ and $X \in \mathcal{M}$. Then*

$$\int_X^{\otimes} A \, d\mu = \int_{g(X)}^{\otimes} g(A) \, d\mu' \quad (14)$$

for any $A \in \mathcal{F}_{\mathbf{L}}(M)$.

In the end of this part we will show that the Sugeno integral is a special case of our proposed integral. Let $\mathbf{L}$ be a complete residuated lattice and $(M, \mathcal{M})$ be a fuzzy measurable space such that $A \cap B \in \mathcal{M}$ for any $A, B \in \mathcal{M}$. Denote $A_a = \{m \mid m \in M \ \& \ A(m) \geq a\}$. We say that an $\mathbf{L}$-fuzzy set A is *$\mathcal{M}$-Sugeno measurable*, if $1_{A_a} \in \mathcal{M}$ for any $a \in L$. The *Sugeno integral* is given, for any fuzzy measure space $(M, \mathcal{M}, \mu)$ with $A \cap B \in \mathcal{M}$ for any $A, B \in \mathcal{M}$, for any $\mathcal{M}$-Sugeno measurable $\mathbf{L}$-fuzzy set A and for any $X \in \mathcal{M}$, by

$$\int_X A \, d\mu = \bigvee_{a \in L} (a \wedge \mu(1_{A_a} \cap X)). \quad (15)$$

Theorem 3.4. *Let $\mathbf{L}$ be a complete Heyting algebra, $(M, \mathcal{M}, \mu)$ be a fuzzy measure space with $A \cap B \in \mathcal{M}$ for any $A, B \in \mathcal{M}$, A be an $\mathcal{M}$-Sugeno measurable $\mathbf{L}$-fuzzy set and $X \in \mathcal{M}$. Then $\int_X A \, d\mu = \int_X^{\otimes} A \, d\mu$.*

For a definition and properties of $\rightarrow$-fuzzy integral we refer to [21].

4 L-fuzzy quantifiers of the type $\langle 1 \rangle$

In [2], we defined the monadic $\mathbf{L}$-fuzzy quantifiers of the type $\langle 1^n, 1 \rangle$. Here, we restrict ourselves to their special subclass, namely, to the monadic $\mathbf{L}$-fuzzy quantifiers of the type $\langle 1 \rangle$ that can be defined as follows.

Definition 4.1. Let $\mathbf{L}$ be a complete residuated lattice, M be a universe (possibly empty[2]). A mapping $Q_M : \mathcal{F}_{\mathbf{L}}(M) \to L$ is called a *monadic $\mathbf{L}$-fuzzy quantifier of the type $\langle 1 \rangle$ limited to M*.

Definition 4.2. An *unlimited (finite, countable) monadic $\mathbf{L}$-fuzzy quantifier of the type $\langle 1 \rangle$* is a functional Q assigning to each (finite, countable) universe M a monadic $\mathbf{L}$-fuzzy quantifier Q_M of the type $\langle 1 \rangle$ limited to M.

In the following text, we will usually omit the terms "unlimited", "monadic" and "of the type $\langle 1 \rangle$" and we will say only "$\mathbf{L}$-fuzzy quantifier". Let us demonstrate several examples of unlimited $\mathbf{L}$-fuzzy quantifiers that are interpretations of well-known quantifiers in natural language (see [2]). We will use expressions all, some, not all and no as generic expressions which stand for natural language quantifiers of the type $\langle 1 \rangle$, e.g. "everything", "someone", "not everyone" and "nothing", respectively.

[2] To define the behavior of generalized quantifiers for the empty universe is important in some situations. It happens, for example, when we study type $\langle 1, 1 \rangle$ quantifiers which are obtained from type $\langle 1 \rangle$ quantifiers by means of *relativization*. Then it is vital to have values of e.g. $\mathsf{some}_\emptyset(1_\emptyset)$ defined, see discussion in [3], p. 137.

ISBN: 978-989-95079-6-8

Example 4.1. Let $\mathbf{L}$ be a complete residuated lattice. Then

$$(\mathsf{all})_M(A) = \bigwedge_{m\in M} A(m),$$
$$(\mathsf{some})_M(A) = \bigvee_{m\in M} A(m),$$
$$(\mathsf{not\ all})_M(A) = \bigvee_{m\in M} \neg A(m),$$
$$(\mathsf{no})_M(A) = \bigwedge_{m\in M} \neg A(m),$$

where M is an arbitrary universe and $A \in \mathcal{F}_{\mathbf{L}}(M)$, define unlimited $\mathbf{L}$-fuzzy quantifiers of the type $\langle 1\rangle$. Obviously, the definitions of all and some (interpretations of quantifiers *all* and *some*) are the same as the interpretations of $\forall$ and $\exists$ in fuzzy logic, respectively. The others are negations of the previous ones. Notice that $(\mathsf{all})_\emptyset(\emptyset) = (\mathsf{no})_\emptyset(\emptyset) = \top$ and $(\mathsf{some})_\emptyset(\emptyset) = (\mathsf{not\ all})_\emptyset(\emptyset) = \bot$.

Now, let us recall some well-known semantics properties that are usually investigated in the case of the $\mathbf{L}$-fuzzy quantifiers of the type $\langle 1\rangle$. For more information as well as examples we refer to [1, 2].

Definition 4.3. Let Q, P be $\mathbf{L}$-fuzzy quantifiers. Then we say that Q *is less than or equal to* P and denote it by $Q \le P$, if, for any non-empty universe M and $A \in \mathcal{F}_{\mathbf{L}}(M)$,

$$Q_M(A) \le P_M(A). \tag{16}$$

Further, we say that Q *is equal to* P and denote it by $Q = P$, if $Q \le P$ and $P \le Q$.

Definition 4.4. Let Q, P be $\mathbf{L}$-fuzzy quantifiers. We say that Q *is identical to* P and denote it by $Q \equiv P$, if for any (possibly empty) universe M and $A \in \mathcal{F}_{\mathbf{L}}(M)$,

$$Q_M(A) = P_M(A). \tag{17}$$

Remark 4.2. Note that the behavior of $\mathbf{L}$-fuzzy quantifiers for the empty universe is often unpredictable (e.g., $\mathsf{all}_M(A) \le \mathsf{some}_M(A)$ for all $M \neq \emptyset$, but $\mathsf{some}_\emptyset(1_\emptyset) \le \mathsf{all}_\emptyset(1_\emptyset)$), therefore, we require only non-empty universes for their comparison in the first definition. Moreover, this restriction seems to be insignificant from the practical point of view. The second definition of identity of $\mathbf{L}$-fuzzy quantifiers gives useful denotation.

Definition 4.5. An $\mathbf{L}$-fuzzy quantifier Q is *permutation-invariant*, if for arbitrary universe M, bijective mapping $f : M \to M$ and $A \in \mathcal{F}_{\mathbf{L}}(M)$

$$Q_M(A) = Q_M(f^{\rightarrow}(A)). \tag{18}$$

The set of all permutation-invariant $\mathbf{L}$-fuzzy quantifiers is denoted by PI.

Definition 4.6. An $\mathbf{L}$-fuzzy quantifier Q is *isomorphism-invariant*, if for arbitrary universe M, bijective mapping $f : M \to M'$ and $A \in \mathcal{F}_{\mathbf{L}}(M)$,

$$Q_M(A) = Q_{M'}(f^{\rightarrow}(A)). \tag{19}$$

The set of all isomorphism-invariant $\mathbf{L}$-fuzzy quantifiers is denoted by ISOM.

Definition 4.7. An $\mathbf{L}$-fuzzy quantifier Q satisfies *extension*, if for arbitrary universes M, M' with $M \subseteq M'$ and $A \in \mathcal{F}_{\mathbf{L}}(M)$,

$$Q_M(A) = Q_{M'}(A). \tag{20}$$

The set of all $\mathbf{L}$-fuzzy quantifiers satisfying extension is denoted by EXT.

Definition 4.8. Let Q be an $\mathbf{L}$-fuzzy quantifier. We say that Q is *monotonically non-decreasing*, if for arbitrary universe M and $A \in \mathcal{F}_{\mathbf{L}}(M)$ and $A' \in \mathcal{F}_{\mathbf{L}}(M)$ with $A \subseteq A'$,

$$Q_M(A) \le Q_M(A') \tag{21}$$

and Q is *monotonically non-increasing*, if for arbitrary universe M and $A \in \mathcal{F}_{\mathbf{L}}(M)$ and $A' \in \mathcal{F}_{\mathbf{L}}(M)$ with $A' \subseteq A$,

$$Q_M(A) \le Q_M(A'). \tag{22}$$

For our purpose we will consider the following stronger definition of $\mathbf{L}$-similarity of fuzzy sets. Recall that a mapping $R : \mathcal{F}_{\mathbf{L}}(M) \times \mathcal{F}_{\mathbf{L}}(M) \to L$ is called an $\mathbf{L}$-*fuzzy relation on* $\mathcal{F}_{\mathbf{L}}(M)$. Let $[A\ R\ B]$ denote the degree in which $\mathbf{L}$-fuzzy sets A and B belongs to $\mathbf{L}$-fuzzy relation, i.e., $[A\ R\ B] = R(A, B)$. Let us define an $\mathbf{L}$-fuzzy relation $\equiv_M$: $\mathcal{F}_{\mathbf{L}}(M) \times \mathcal{F}_{\mathbf{L}}(M) \to L$ by $[A \equiv_M B] = \top$, if there is a bijective mapping f of M onto M such that $f^{\rightarrow}(A) = B$, and $[A \equiv_M B] = \bot$, otherwise. The following definition generalizes $\equiv_M$.

Definition 4.9. An $\mathbf{L}$-fuzzy relation $\approx_M$: $\mathcal{F}_{\mathbf{L}}(M) \times \mathcal{F}_{\mathbf{L}}(M) \to L$ is called an $\mathbf{L}$-*permutation equivalence* on $\mathcal{F}_{\mathbf{L}}(M)$, if

$$[A \approx_M B] \ge [A \equiv_M B] \tag{23}$$
$$[A \approx_M B] = [B \approx_M A] \tag{24}$$
$$[A \approx_M B] \le [\overline{A} \approx_M \overline{B}] \tag{25}$$
$$[A \approx_M B] \otimes [B \approx_M C] \le [A \approx_M C] \tag{26}$$

hold for arbitrary $A, B, C \in \mathcal{F}_{\mathbf{L}}(M)$.

Obviously, (24) and (26) are the common axioms of symmetry and transitivity, respectively. Let $A, B \subseteq M$ are $\mathbf{L}$-fuzzy sets which are similar. Then one could wish that the complements of A and B are also similar (at least in the degree in which A and B are $\mathbf{L}$-equivalent). This idea is expressed in (25).

Example 4.3 (see [2])**.** Let $\mathbf{L}$ be a complete residuated lattice, M be any universe and $\mathrm{Perm}(M)$ denote the set of all bijective mappings of M onto M. Then

$$[A \approx^{\wedge}_M B] = \bigvee_{f\in \mathrm{Perm}(M)} \bigwedge_{m\in M} (A(m) \leftrightarrow B(m)) \tag{27}$$

defines the $\mathbf{L}$-permutation equivalence $\approx^{\wedge}_M$ on $\mathcal{F}_{\mathbf{L}}(M)$.

Definition 4.10. Let $\approx$ be a class of $\mathbf{L}$-permutation equivalences such that for each (finite, countable) universe M there is a unique $\approx_M$ from $\approx$ defined on $\mathcal{F}_{\mathbf{L}}(M)$. A (finite, countable) $\mathbf{L}$-fuzzy quantifier Q of the type $\langle 1\rangle$ is *extensional with respect to* $\approx$, if

$$[A \approx_M A'] \le Q_M(A) \leftrightarrow Q_M(A') \tag{28}$$

for each (finite, countable) universe M and $A, A' \in \mathcal{F}_{\mathbf{L}}(M)$. The set of all extensional $\mathbf{L}$-fuzzy quantifiers with respect to $\approx$ is denoted by $\mathrm{EXTENS}(\approx)$.

ISBN: 978-989-95079-6-8

5 L-fuzzy quantifiers of the type $\langle 1\rangle$ determined by fuzzy measures

Let $\mathcal{S}(M)$ denote a set of fuzzy measure spaces defined on M. For better readability, we will denote by

$$\int_{(M,\mathcal{M})}^{\otimes} A\, d\mu \tag{29}$$

the $\otimes$-fuzzy integral $\int^{\otimes} A\, d\mu$ defined over a fuzzy measure space $(M,\mathcal{M},\mu)$. Now we can define $\mathbf{L}$-fuzzy quantifiers limited to M using fuzzy measure spaces from a set $\mathcal{S}(M)$ as follows.

Definition 5.1. Let $\mathcal{S}(M)$ be a (possibly empty) set of fuzzy measure spaces defined on a non-empty universe M. An $\mathbf{L}$-*fuzzy quantifier of the type* $\langle 1\rangle$ *limited to* M *determined by the fuzzy measure spaces from* $\mathcal{S}(M)$ is a mapping $Q_{\mathcal{S}(M)} : \mathcal{F}_{\mathbf{L}}(M) \to L$ defined by

$$Q_{\mathcal{S}(M)}(A) = \bigvee_{(M,\mathcal{M},\mu)\in\mathcal{S}(M)} \int_{(M,\mathcal{M})}^{\otimes} A\, d\mu. \tag{30}$$

Remark 5.1. It is easy to see that if $\mathcal{S}(M) = \emptyset$ for some (possibly non-empty) set M, then $Q_{\mathcal{S}(M)}(A) = \bot$ for any $A \in \mathcal{F}_{\mathbf{L}}(M)$.

It is easy to see that $\mathcal{S}(\emptyset) = \emptyset$ (there is no fuzzy measure space with $M = \emptyset$). Hence, each unlimited fuzzy quantifier Q based only on the formula (30) has $Q_\emptyset(1_\emptyset) = Q_{\mathcal{S}(\emptyset)}(1_\emptyset) = \bot$. However, for example, it holds that $(\mathsf{all})_\emptyset(1_\emptyset) = \top$. This motivates us to exclude the determination of $Q_\emptyset(1_\emptyset)$ by (30) in the following definition of unlimited $\mathbf{L}$-fuzzy quantifier.

Definition 5.2. Let $\mathcal{S}$ be a functional assigning to each universe M a set $\mathcal{S}(M)$ of fuzzy measure spaces defined on M. An *unlimited* $\mathbf{L}$-*fuzzy quantifier of the type* $\langle 1\rangle$ *determined by fuzzy measures over* $\mathcal{S}$ is an unlimited $\mathbf{L}$-fuzzy quantifier of the type $\langle 1\rangle$ assigning an $\mathbf{L}$-fuzzy quantifier $Q_{\mathcal{S}(M)}$ determined by the fuzzy measure spaces from $\mathcal{S}(M)$ to each *non-empty* universe M.

Example 5.2. Let M be a non-empty universe and $\mathcal{S}(M)_i = \{(M,\mathcal{F}_{\mathbf{L}}(M),\mu_i)\}$, where, for $i = 1, 2$,

$$\mu_1(A) = \begin{cases} \bot, & \text{if } A = 1_\emptyset, \\ \top, & \text{otherwise} \end{cases} \tag{31}$$

and

$$\mu_2(A) = \begin{cases} \top, & \text{if } A = 1_M, \\ \bot, & \text{otherwise.} \end{cases} \tag{32}$$

If Q is determined by fuzzy measures from $Q_{\mathcal{S}(M)_1}$ for all $M \neq \emptyset$, then $Q = \mathsf{some}$. In fact, if $M \neq \emptyset$ and $A \in \mathcal{F}_{\mathbf{L}}(M)$, then

$$Q_M(A) = Q_{\mathcal{S}(M)_1}(A) = \int_{(M,\mathcal{F}_{\mathbf{L}}(M))}^{\otimes} A\, d\mu_1 =$$
$$= \bigvee_{m\in M} \mu_1(\{m\}) \otimes A(m) =$$
$$= \bigvee_{m\in M} \top \otimes A(m) = \bigvee_{m\in M} A(m) = (\mathsf{some})_M(A).$$

According to Definition 4.3, $Q = \mathsf{some}$. One checks easily that Q determined by $Q_{\mathcal{S}(M)_2}$ for all $M \neq \emptyset$ is equal to all.

Example 5.3. Let $\mathbf{L}$ be the standard Gödel algebra and

$$\mathcal{S}(M) = \{(M,\mathcal{F}_{\mathbf{L}}(M),\mu)\}.$$

One checks easily, using Theorem 3.1, that

$$Q_{\mathcal{S}(M)}(A) = \bigvee_{1_Y\in\mathcal{P}_{\mathbf{L}}(M)\setminus\{1_\emptyset\}} \Big(\mu(1_Y) \wedge \bigwedge_{m\in Y} A(m)\Big),$$

because $\otimes$ coincides with $\wedge$ in Gödel algebras. Notice that because $\mathbf{L}$ is the Gödel algebra, it is also a Heyting algebra and, according to Theorem 3.4, the $\otimes$-fuzzy integral coincides with the Sugeno integral in this case.

Let M be a non-empty countable universe and μ_f denote one of the fuzzy measures on $(M,\mathcal{F}_{\mathbf{L}}(M))$ defined by (6) and (7) in Example 3.1. Putting

$$Q_{\mathcal{S}(M)}(A) = \bigvee_{1_Y\in\mathcal{P}_{\mathbf{L}}(M)\setminus\{1_\emptyset\}} \Big(\mu_f(1_Y) \wedge \bigwedge_{m\in Y} A(m)\Big)$$

for any non-empty countable universe M and $Q_\emptyset(1_\emptyset) = \top$, we obtain a countable $\mathbf{L}$-fuzzy quantifier which is an interpretation of the quantifier *many things*. Define

$$\mu_f^{1/2}(A) = \begin{cases} \top, & \text{if } \mu_f(A) \geq \frac{1}{2}, \\ \bot, & \text{otherwise,} \end{cases}$$

for any $A \in \mathcal{F}_{\mathbf{L}}(M)$. Then putting

$$Q_{\mathcal{S}(M)}(A) = \bigvee_{1_Y\in\mathcal{P}_{\mathbf{L}}(M)\setminus\{1_\emptyset\}} \Big(\mu_f^{1/2}(1_Y) \wedge \bigwedge_{m\in Y} A(m)\Big)$$

for any non-empty countable universe M and $Q_\emptyset(1_\emptyset) = \top$, we obtain a countable $\mathbf{L}$-fuzzy quantifier which is an interpretation of the quantifier *at least half things*. If we restrict ourselves to the class of all finite $\mathbf{L}$-fuzzy quantifiers, then one checks easily (using the equality $\mu_f(A) = \mu_f(h^{\rightarrow}(A))$ from Example 3.1) that both defined quantifiers are PI and ISOM. Moreover, they are $\mathrm{EXTENS}(\approx^{\wedge})$ (see Theorem 5.6).

Intuitively, it is obvious that all is the smallest and some the greatest $\mathbf{L}$-fuzzy quantifier Q determined by fuzzy measures with respect to the ordering from Definition 4.3.

Theorem 5.1. *For each* $\mathbf{L}$*-fuzzy quantifier* Q *determined by fuzzy measures over* $\mathcal{S}$*, it holds that*

$$\mathsf{all} \leq Q \leq \mathsf{some}. \tag{33}$$

In the following part, we will present some results on the semantic properties of $\mathbf{L}$-fuzzy quantifiers determined by fuzzy measures. The following theorem states a sufficient condition for $\mathbf{L}$-fuzzy quantifiers to be permutation invariant.

Theorem 5.2. *Let* Q *be an unlimited* $\mathbf{L}$*-fuzzy quantifier of the type* $\langle 1\rangle$ *determined by fuzzy measures over* $\mathcal{S}$ *such that for each non-empty universe* M *it holds that* $\mathcal{S}(M) = [(M,\mathcal{M},\mu)]$*. Then* $Q \in \mathrm{PI}$.

Note that the specification of a necessary condition for $\mathbf{L}$-fuzzy quantifiers being permutation invariant seems to be immensely complicated and it is still an open problem. In the following theorem, let us denote fuzzy measure spaces $(M,\mathcal{M},\mu)$ and $(M',\mathcal{M}',\mu')$ by $\mathbf{M}$ and $\mathbf{M}'$, respectively.

ISBN: 978-989-95079-6-8

Theorem 5.3. *Let Q be an unlimited* $\mathbf{L}$*-fuzzy quantifier of the type $\langle 1\rangle$ determined by fuzzy measures over $\mathcal{S}$ such that, for any universes M, M' with the same cardinality,*

(i) if $\mathbf{M} \in \mathcal{S}(M)$ and $f : M \to M'$ is a bijection, then $f^{\to}(\mathbf{M}) \in \mathcal{S}(M')$,

(ii) if $\mathbf{M} \in \mathcal{S}(M)$ and $\mathbf{M}' \in \mathcal{S}(M')$, then $\mathbf{M}$ and $\mathbf{M}'$ are isomorphic.

Then $Q \in$ ISOM.

From the definition of $\mathbf{L}$-fuzzy quantifiers of the type $\langle 1\rangle$ determined by fuzzy measures it should be clear that they possess non-decreasing behavior, which is expressed by the following theorem. Naturally, if we base $\mathbf{L}$-fuzzy quantifiers of the type $\langle 1\rangle$ on complementary fuzzy measures and $\to$-fuzzy integrals [21], they will be non-increasing.

Theorem 5.4. *Let Q be an unlimited* $\mathbf{L}$*-fuzzy quantifier of the type $\langle 1\rangle$ determined by fuzzy measures over $\mathcal{S}$. Then Q is a non-decreasing* $\mathbf{L}$*-fuzzy quantifier.*

The following theorem shows that some is the only type $\langle 1\rangle$ quantifier with the extension (EXT) property which can be successfully modeled by our $\mathbf{L}$-fuzzy quantifiers of the type $\langle 1\rangle$ determined by fuzzy measures. However, the quantifier all and quantifiers which are interesting from the point of view of fuzzy logic, for example "at least half things", "many things", "most things" do not possess the extension property, and we can model them, see Examples 5.2 and 5.3. Nevertheless, quantifiers which refer to absolute cardinalities, e.g. "at least three things", possess the extension property, therefore they cannot be successfully modeled by our quantifiers.

Theorem 5.5. *Let Q be an unlimited* $\mathbf{L}$*-fuzzy quantifier of the type $\langle 1\rangle$ determined by fuzzy measures over $\mathcal{S}$. Then $Q \in$* EXT *if and only if $Q \equiv$* some.

Finally, we show that fuzzy quantifiers of type $\langle 1\rangle$ determined by fuzzy measures are behaving well with respect to extensionality (see Definition 4.10).

Theorem 5.6. *Let Q be an unlimited* $\mathbf{L}$*-fuzzy quantifier of the type $\langle 1\rangle$ determined by fuzzy measures over $\mathcal{S}$ such that $\mathcal{S}(M) = [(M, \mathcal{M}, \mu)]$ for each non-empty universe M. Then $Q \in$* EXTENS($\approx^{\wedge}$).

6 Conclusion

In the future we will concentrate on studying of $\mathbf{L}$-fuzzy quantifiers of the type $\langle 1, 1\rangle$ (and possibly also of the type $\langle 1^n, 1\rangle$) generated by fuzzy measures. Quantifiers of the type $\langle 1, 1\rangle$ serve as models of very important class of natural language determiners (cf. e.g. [3]), for example "a few X are Y", "almost all X are Y", etc. Our definitions of $\otimes$-fuzzy integral and $\to$-fuzzy integral allow us to define these quantifiers, and we believe that they provide an important class of models with interesting properties.

Acknowledgment

This paper has been supported by the Grant IAA108270901 of the GA AV ČR.

References

[1] I. Glöckner. *Fuzzy Quantifiers in Natural Language: Semantics and Computational Models*. Der Andere Verlag, Osnabrück, Germany, 2004.

[2] M. Holčapek. Monadic $\mathbf{L}$-fuzzy quantifiers of the type $\langle 1^n, 1\rangle$. *Fuzzy Sets and Systems*, 159:1811–1835, 2008.

[3] S. Peters and D. Westerståhl. *Quantifiers in Language and Logic*. Oxford University Press, New York, 2006.

[4] A. Mostowski. On a generalization of quantifiers. *Fund. Math.*, 44:12–36, 1957.

[5] P. Lindström. First ordered predicate logic with generalized quantifiers. *Theoria*, 32:186–195, 1966.

[6] J. Barwise and R. Cooper. Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4:159–219, 1981.

[7] L.A. Zadeh. A computational approach to fuzzy quantifiers in natural languages. *Comp. Math. with Applications*, 9:149–184, 1983.

[8] H. Thiele. On fuzzy quantifiers. In Z. Bien and K.C. Min, editors, *Fuzzy Logic and its Applications, Information Sciences, and Intelligent Systems*, pages 343–352. Kluwer Academic Publisher, Dordrecht, 1996.

[9] D. Ralescu. Cardinality, quantifiers and the aggregation of fuzzy criteria. *Fuzzy Sets and Systems*, 69:355–365, 1995.

[10] Y. Liu and E.E. Kerre. An overview of fuzzy quantfiers. part I: Interpretation. *Fuzzy Sets and Systems*, 95(1):1–21, 1998.

[11] Y. Liu and E.E. Kerre. An overview of fuzzy quantfiers. part II: Reasoning and application. *Fuzzy Sets and Systems*, 95(2):135–146, 1998.

[12] V. Novák. Antonyms and linguistic quantifiers in fuzzy logic. *Fuzzy Sets and Systems*, 124(3):335–351, 2001.

[13] P. Hájek. *Metamathematics of Fuzzy Logic*. Kluwer Academic Publishers, Dordrecht, 1998.

[14] I. Glöckner. *Fuzzy Quantifiers: A Computational Theory*. Springer-Verlag, Berlin, 2006.

[15] V. Novák. A formal theory of intermediate quantifiers. *Fuzzy Sets and Systems*, 159(10):1229 1246, 2008.

[16] V. Novák. On fuzzy type theory. *Fuzzy Sets and Systems*, 149:235–273, 2005.

[17] M. Ying. Linguistic quantifiers modeled by Sugeno integrals. *Artificial Intelligence*, 170:581–606, 2006.

[18] Z. Wang and G.J. Klir. *Fuzzy measure theory*. Plenum Press, New York, 1992.

[19] M. Grabisch, T. Murofushi, and M. Sugeno, editors. *Fuzzy Measures and Integrals. Theory and Applications*. Studies in Fuzziness and Soft Computing. Physica Verlag, Heidelberg, 2000.

[20] E.P. Klement, R. Mesiar, and E. Pap. A universal integral based on measures of level sets. *IEEE Transactions on Fuzzy Systems*, to appear.

[21] A. Dvořák and M Holčapek. Fuzzy integral over complete residuated lattices. In *IFSA/EUSFLAT 2009*, Lisboa, 2009.

[22] R. Bělohlávek. *Fuzzy Relational Systems: Foundations and Principles*. Kluwer Academic Publisher, New York, 2002.

ISBN: 978-989-95079-6-8

On the Definition of Extended Norms and Co-norms to Aggregate Fuzzy Bipolar Conditions

Ludovic Liétard[1] Daniel Rocacher[2]

1 IRISA/IUT, rue Edouard Branly, BP 30219,
22302 Lannion Cedex, France
2 IRISA/ENSSAT, rue de Kérampont, BP 447,
22305 Lannion Cedex, France
Email: ludovic.lietard@univ-rennes1.fr, rocacher@enssat.fr

Abstract— *Fuzzy queries addressed to a relational database management system are based on the expression of user preferences which leads to obtain more appropriate answers. Fuzzy set theory provides powerful tools to define such conditions (fuzzy conditions). In this context, it is possible to consider fuzzy bipolar conditions which are made of two fuzzy sets to describe user's preferences: i) a fuzzy constraint to represent a mandatory requirement, ii) a fuzzy wish to represent an optional requirement. A Fuzzy constraint is mandatory in the sense that it* ***must*** *be fulfilled (and it can be used to discard elements). A fuzzy wish is optional since it cannot be used to discard elements (but its satisfaction brings a bonification). This paper is devoted to the aggregation of fuzzy bipolar conditions and to the definition of extended norms and co-norms.*

Keywords— *flexible querying, fuzzy bipolar condition, relational database.*

1 Introduction

When querying a relational database management system (DBMS), the expression of user preferences is a manner to integrate the personalized needs of the user. As a consequence, it leads to obtain more appropriate answers, since they can be ranked from the more relevant (the more preferred) to the less relevant (a set of discriminated answers being obtained). The problem of expressing and managing user preferences in the relational model of data has received more and more attention in the last few years [1, 2, 3, 4, 5, 6, 7] and we consider the study of bipolar conditions [8, 9, 10, 11, 12]. Such conditions are made of two components: a mandatory condition (also called a constraint) and an optional condition (also called a wish). More precisely, we consider fuzzy bipolar conditions where the constraint and the wish express user preferences and are defined by fuzzy sets.

The main interrogation systems proposed to integrate user preferences do not take into account such bipolar conditions. The SQLf language considers preferences defined by fuzzy sets and proposes a commensurable approach without bipolarity. It does not distinguish between constraints and wishes, each atomic fuzzy condition being a commensurable constraint. On the contrary, PreferenceSQL deals with a particular case of bipolarity where constraints are Boolean while wishes are not commensurable preferences. The preferences being not commensurable in PreferenceSQL, this proposition leads to a partial order of answers while SQLf language delivers a total order (the preferences being commensurable). One of the main differences [10] between these two approaches is that SQLf is set in the algebraic framework while it is not the case for PreferenceSQL. As a consequence, this article is a contribution to the extension of the relational algebra (and SQL) to fuzzy bipolar conditions. Our final aim is to propose an SQL-like query language where fuzzy bipolar conditions can be defined as an extension of fuzzy conditions. As a consequence, a total order of answers is expected.

Fuzzy bipolar conditions are introduced in section 2 and it is shown that a fuzzy condition defined by a fuzzy set is a particular case of a fuzzy bipolar condition. Section 3 is interested in extending logical operators to define more complex fuzzy bipolar conditions. These operators are soundness from an user's point of view but they cannot be considered as extended norms and co-norms. Section 4 proposes definition for an extended norm and an extended co-norm to aggregate fuzzy bipolar conditions.

2 Fuzzy bipolar conditions

Bipolarity can be defined as the human capability to evaluate situations in terms of pros and cons or positive and negative aspects. Depending of the relationship between these two aspects, Dubois and Prade [8, 9] distinguish between three categories of bipolarity, called types I, II and III.

The type I is called *symmetric univariate bipolarity*. It corresponds to the case where the scale to evaluate situations is bipolar which means defined by an ordered set (total order) with a neutral elements to separate positive and negative information (*symmetric bipolarity*). As a consequence a single value (*univariate*) from this scale expresses, at the same time the positive and negative aspects since the opposite evaluation is obtained by symmetry. An example of type I bipolarity is provided by probability measures, the neutral element being 0.5. The type II is called *symmetric bivariate unipolarity*. In this case, the scale to evaluate situations is unipolar which means that the neutral element is at one end of the scale and two dual values (symmetric bivariate) is necessary to judge a situation. Unlike the symmetric univariate bipolarity, it is possible for a situation to be evaluated positive and negative at the same time. An example of type II bipolarity is provided by possibility theory where a couple of dual unipolar indexes (possibility and possibility of opposite event) is used to evaluate a situation. The type III is called *asymmetric bipolarity*. In this case two bipolar scales are used to evaluate a situation, one expressing the positive, the other one the negative aspects. As a consequence, positive and negative evaluations are of different nature and can be provided by different kind of sources.

ISBN: 978-989-95079-6-8

A bipolar condition is made of two components: a *constraint* (a condition denoted C) whose negation defines not acceptable values (negative evaluation) and a *wish* (a condition denoted W) to define desired values (positive evaluation). The satisfaction to the constraint and the wish are based on two different bipolar scales, the two extremum being the total satisfaction 1 and the total non satisfaction 0. In addition, constraints and wishes are neither dual nor symmetric. As a consequence, a bipolar condition expresses a type III bipolarity.

This paper considers that a bipolar condition denoted (C, W) means "to satisfy C and, if possible, to satisfy W" and this meaning has two important consequences. Firstly, it implies that the satisfaction with respect to the constraint is mandatory while the satisfaction with respect to the wish is optional. Secondly, the set of wished values must be a subset of that of mandatory values (W $\subseteq$ C). This property states that it is incoherent to wish non-acceptable values.

In the context of database querying with such conditions, tuples which do not satisfy the constraint are discarded (since this condition is mandatory), while tuples which do not satisfy the wish can be proposed to the user (the wish being optional). When C and W are Boolean conditions, the satisfaction with respect to the bipolar condition (C,W) is a couple (c, w) where c and w are values in {0, 1}. Tuples satisfying the constraint and the wish are returned in priority to the user. If such answers do not exist, the tuples satisfying only the constraint are delivered. When the two components express preferences and are defined by fuzzy sets, their satisfactions are sets in [0,1]. As a consequence, the satisfaction of a given element x with respect to a fuzzy bipolar condition is a couple of degrees ($\mu_C(x)$, $\mu_W(x)$), and their processing is more tricky. These two degrees are not commensurable (due to the chosen meaning for the fuzzy bipolar conditions) and, consequently, they cannot be aggregated to compute an overall degree of satisfaction. In addition, the constraint being mandatory, its satisfaction is firstly used to discriminate among answers. The satisfactions with respect to the wish being not mandatory, they can only be used to discriminate among answers having same evaluation with respect to the constraint. Thus, the wishes allow to differentiate between tuples which are equal with respect to their constraints and a *total order* is obtained on C and W (with (1, 1) as greatest element and (0, 0) as least element). In other words, it is possible to rank the tuples by using a *lexicographical* order on the constraints and the wishes:

a tuple t_i is preferred to a tuple t_j
if ($\mu_C(t_i) > \mu_C(t_j)$)
or (($\mu_C(t_i) = \mu_C(t_j)$) and ($\mu_W(t_i) > \mu_W(t_j)$)),

Obviously, a tuple t_i is similar to a tuple t_j (they have same level of preference) if and only if $\mu_C(t_i) = \mu_C(t_j)$ and $\mu_W(t_i) = \mu_W(t_j)$.

For the sake of simplicity, "t_i is preferred to t_j" is denoted $t_i > t_j$ while "t_i is similar to t_j" is denoted $t_i = t_j$.

Example 1. We consider relation Sales from Table 1 where attributes #Sale, Date, Town, Benefit are respectively the identification of a sale, the date and place where it has been concluded and the percentage of benefit for the salesman.

A salesman wants to know the town where a sale has occurred with a *high* benefit and, if possible, a benefit *around 100%*. A *high* benefit is a mandatory condition while condition *around 100%* is optional. These two fuzzy conditions are given by Figure 1 and define the fuzzy bipolar condition (C, W) where C is "*high* benefit" and W is "benefit *around 100%*".

Table 1 : Relation Sales.

#Sale	Date	Town	Benefit
#1	01-05-07	Paris	82%
#2	04-06-07	Paris	85%
#3	08-09-07	Londres	59%
#4	15-06-07	Nice	94%
#5	17-07-07	Londres	97%
#6	08-09-07	Madrid	56%

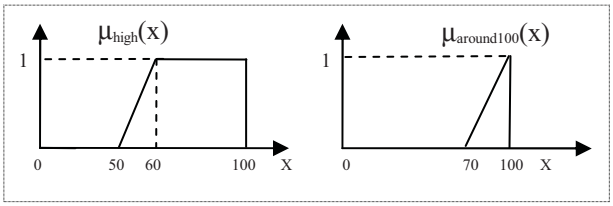


Figure 1 : *high* benefit and benefit *around 100%*.

The satisfactions with respect to the constraint C and the wish W are given by Table 2.

Table 2 : Satisfactions with respect to the fuzzy bipolar condition : "*high* benefit" and if possible "*around 100%*".

#Sale	*(high* benefit, benefit *around 100%*) (μ_C(#Sale), μ_W(#Sale)),
#1	(1, 0.4)
#2	(1, 0.5)
#3	(0.9, 0)
#4	(1, 0.8)
#5	(1, 0.9)
#6	(0.6, 0)

The lexicographical order gives :

#5 > #4 > #2 > #1 > #3 > #6.

Sale #5 is the best sale with respect to the bipolar condition (C, W) because, among the best sales according to the constraint (at degree 1), it is the best one according to the wish. ◆

The property of constraints and wishes expressed on a same universe is that the set of wished values must be a subset of that of mandatory values (W $\subseteq$ C). In case of fuzzy sets, this property of inclusion is rewritten :

for each value x in X, $\mu_W(x) \leq \mu_C(x)$,

ISBN: 978-989-95079-6-8

such that X is the domain set where the fuzzy bipolar condition applies. Such a property can be checked on the fuzzy bipolar condition provided by example 1. This property is natural and is always implicit in the formulation of fuzzy bipolar conditions. As a consequence, a condition formulated as "a *cheap* and, if possible, a *low* consuming car" should be rewritten : "a *cheap* car and, if possible, a *cheap* and *low* consuming car".
An important particular case arises when a fuzzy bipolar condition is a fuzzy predicate (a pure constraint). Since a fuzzy predicate C can be rewritten "C and, if possible, C", a pure constraint is defined by a fuzzy bipolar condition where the constraint equals the wish ((C, C)). As a consequence, a fuzzy bipolar condition is a generalization of a fuzzy condition.

3 Logical operators and fuzzy bipolar conditions

In this section, we consider several fuzzy bipolar conditions and their aggregation using conjunction and disjunction operators.
A first idea to define the conjunction of two fuzzy bipolar conditions (C_1, W_1) and (C_2, W_2) is to compute the minimum of both the satisfactions to the constraints and to the wishes. In this case "(C_1, W_1) and (C_2, W_2)" is the fuzzy bipolar condition "to satisfy (C_1 and C_2) and if possible to satisfy (W_1 and W_2)". Since the minimum is a triangular norm, this operation can be generalized to define parameterized operators $O_{op1,op2}$ using various norms and conorms :

$$(C_1,W_1)\ O_{op1,\,op2}\ (C_2,W_2) \equiv (op_1(C_1, C_2), op_2(W_1,W_2)), \quad (1)$$

where op_1 and op_2 are either a norm or a co-norm. Depending on the choice for op_1 and op_2 , different meanings can be obtained for the operator (as shown by Table 3).

Table 3 : Different meanings for $O_{op1,op2}$.

	op_1 is a norm	op_1 is a co-norm
op_2 is a norm	$(C_1,W_1)\ O_{op1,op2}\ (C_2,W_2)$ ≡ (C_1 and C_2) and if possible (W_1 and W_2)	$(C_1,W_1)\ O_{op1,\,op2}\ (C_2,W_2)$ ≡ (C_1 or C_2) and if possible (W_1 and W_2)
op_2 is a co-norm	$(C_1,W_1)\ O_{op1,op2}\ (C_2, W_2)$ ≡ (C_1 and C_2) and if possible (W_1 or W_2)	$(C_1,W_1)\ O_{op1,op2}\ (C_2,W_2)$ ≡ (C_1 or C_2) and if possible (W_1 or W_2)

Example 2. A client looks for a trip in a travel agency. His request is expressed as: "a *cheap* journey, if possible with a *small* flight time, and an hotel in a *small* town, if possible located *in the countryside*". This query is a selection R:Θ where relation R describes trips proposed by the agency and condition Θ is the conjunction between two fuzzy bipolar conditions "(C_1,W_1) and (C_2,W_2)". Constraint C_1 is the fuzzy condition "a *cheap* journey", W_1 is the fuzzy wish "a *cheap* journey with a *small* flight time" and C_2 is the fuzzy constraint "hotel in a *small* town" while W_2 is the fuzzy wish "hotel in a *small* town and *in the countryside*". Each one of these two atomic bipolar fuzzy conditions is defined on several attributes (on the Cartesian product of their respective domains). The different satisfactions with respect to the constraints and wishes are given by Table 4. Obviously, the result depends on the chosen operator to represent the conjunction. We propose the two different conjunction operators $O_{min,\,min}$ and $O_{min,\,max}$.

Table 4 : Satisfactions with respect to (C_1,W_1) and (C_2,W_2)

#Trip	C_1 "a *cheap* journey"	W_1 "a *cheap* journey and a *small* flight time"	C_2 "*small* town"	W_2 "a *small* town and *in the countryside*"
#1	0.9	0.7	1	0.1
#2	1	0.5	0.9	0.5
#3	0.9	0.3	1	1
#4	0.7	0.6	0.6	0.5
#5	0.8	0.2	0.6	0.5

When the operator $O_{min,\,min}$ is chosen, the meaning of the query becomes:

"a *cheap* journey with an hotel in a *small* town, if possible with a *small* flight time ***and*** an hotel located *in the countryside*".

The results are given by Table 5.

Table 5 : Results provided by $O_{min,\,min}$

#Trip	μ_C(#Trip)	μ_W(#Trip)
#2	0.9	0.5
#3	0.9	0.3
#1	0.9	0.1
#4	0.6	0.5
#5	0.6	0.2

We obtain the order : #2 > #3 > #1 > #4 > #5.

When the operator $O_{min,\,max}$ is chosen, the meaning of the query becomes:

"a *cheap* journey with an hotel in a *small* town, if possible with a *small* flight time ***or*** an hotel located *in the countryside*".

The results are given by Table 6:

Table 6 : Results provided by $O_{min,\,max}$

#Trip	μ_C(#Trip)	μ_W(#Trip)
#3	0.9	1
#1	0.9	0.7
#2	0.9	0.5
#4	0.6	0.6
#5	0.6	0.5

ISBN: 978-989-95079-6-8

We obtain the order : #3 > #1 > #2 > #4 > #5.

The obtained order is different from the ones provided by $O_{min, min}$ (the meanings of the operator being different).
This example shows that the two different operators considered here lead to two different results since they have two different meanings ($O_{min, min}$ means "*all* the constraints and *all* the wishes", while $O_{min, max}$ means "*all* the constraints and *at least one* wish"). ♦

Last example has shown that, when $O_{min, min}$ and $O_{min, max}$ are used, the fundamental property of constraints and wishes ($W \subseteq C$) is lost (since the result are based on ad-hoc operators). However, we think that this loss is not an issue since it reflects the user's attitude with respect to the aggregation of wishes (either all, either at least one). However, it stresses the need for other operators (as extended norms and co-norms).

4 Extended norms and co-norms

This section is interested in defining operators acting as extended norms and co-norms for fuzzy bipolar conditions. Subsection 4.1 shows that it is not possible to define an extended norm (and an extended co-norm) using any operator $O_{op1, op2}$ where op_1 and op_2 are either a norm or a co-norm (cf. section 3). Section 4.2 defines a couple of extended norm and co-norm (and their associated negation).

4.1 Extended norms, co-norms and $O_{op1, op2}$

The different properties bring by norms and co-norms are necessary conditions to keep the algebraic framework of computations (or at least to achieve computations in an human-consistent framework). The operator $O_{min,min}$ is a rather rational candidate for an extended norm but next example shows that it does not respect the monotonicity property of norms and it concludes that this loss is very damageable.

Example 3. The property of monotonicity of a given norm $\wedge$ is defined by:

$$a \geq b \text{ and } c \geq d \Rightarrow (a \wedge c) \geq (b \wedge d)$$

The following counter example shows it is not respected by $O_{min,min}$:

Let x and y be two elements and two fuzzy bipolar conditions FBC1 and FBC2. The respective satisfaction of x and y to these two fuzzy bipolar conditions are shown by Figure 2.

	FBC1	FBC2
x	(0.7, 0.7)	(0.9, 0.4)
y	(0.7, 0.6)	(0.8, 0.5)

Figure 2 : Satisfactions of x and y to FBC1, FBC2.

The property of monotonicity does not hold in this case since:

$$(0.7, 0.7) \geq (0.7, 0.6) \text{ and } (0.9 , 0.4) \geq (0.8, 0.5)$$

while :

$$(0.7, 0.7)\ O_{min,min}\ (0.9, 0.4)$$

$$< (0.7, 0.6)\ O_{min,min}\ (0.8, 0.5).$$

This behaviour is not acceptable for a conjunction since x is preferred to y when considering criteria FBC1 and when considering criteria FBC2, but x is not preferred to y when considering the conjunction "FBC1 and FBC2". ♦

More generally, an extended norm (or co-norm) cannot be defined by any operator $O_{op1,op2}$ as demonstrated hereafter (the demonstration is only provided for the norm and it can be easily adapted for a co-norm).

Demonstration. We assume an extended norm defined as $O_{op1,op2}$:

$$(C_1,W_1)\ O_{op1, op2}\ (C_2,W_2) \equiv (op_1(C_1, C_2), op_2(W_1,W_2)),$$

where op_1 and op_2 are either a norm or a co-norm. The demonstration is based on two steps. At first, two properties which are necessarily satisfied by such an operator are pointed out. Secondly, it is shown that these properties are not compatible with an extended norm since the property of monotonicity does not hold.

First step: Since a fuzzy condition (a fuzzy set) is a particular case of a fuzzy bipolar condition, an extended norm for fuzzy bipolar conditions should revert to a norm in case of a fuzzy set. A fuzzy set being defined as a fuzzy bipolar condition where the constraint equals the wish, the application of an extended norm $O_{op1,op2}$ can be rewritten (in case of fuzzy sets):

$$(C_1,C_1)\ O_{op1, op2}\ (C_2,C_2) \equiv (op_1(C_1, C_2), op_2(C_1,C_2)),$$

In addition $(op_1(C_1, C_2), op_2(C_1,C_2))$ is nothing but the expression of a norm $\wedge$ between C_1 and C_2 . Since $\wedge (C_1, C_2)$ can be represented by $(\wedge (C_1, C_2), \wedge (C_1,C_2))$ we immediately get two properties:

P1) op1 equals op2,
P2) op1 is a norm (denoted $\wedge$).

Second step: Due to the previous result, any $O_{op1,op2}$ operator defining an extended norm is rewritten $O_{\wedge,\wedge}$ where $\wedge$ is a norm. As a consequence, the property of monotonicity does not hold since:

$$(0, 1) \geq (0, 1)$$

and $(1, 0) \geq (0, 1)$

while:

$$(0, 1)\ O_{\wedge,\wedge}\ (1, 0) < (0, 1)\ O_{\wedge,\wedge}\ (0, 1)$$

(since $(0, 1)\ O_{\wedge,\wedge}\ (1, 0) = (0 \wedge 1, 1 \wedge 0) = (0, 0)$
and $(0, 1)\ O_{\wedge,\wedge}\ (0, 1) = (0 \wedge 0, 1 \wedge 1) = (0, 1)$). ♦

ISBN: 978-989-95079-6-8

This subsection has shown that it is necessary to define operators satisfying the properties of norms and co-norms (we recall that the total order considered here is the lexicographical order with (1, 1) as greatest element and (0, 0) as least element).

4.2 Definition for a couple of extended norm and co-norm

An extended norm $\wedge$ and an extended co-norm $\vee$ can be defined by the minimum and maximum with respect to the lexicographical order (which provides a total order):

$$(x, y) \wedge (x', y') = \min((x, y), (x', y')), \quad (2)$$

$$(x, y) \vee (x', y') = \max((x, y), (x', y')). \quad (3)$$

Due to the total order, it is easy to prove that $\wedge$ and $\vee$ are respectively an extended norm and an extended co-norm.

Similarly, a negation operator $\neg$ can be defined by:

$$\neg(x, y) = (1-x, 1-y). \quad (4)$$

Next proof shows that the De Morgan laws are preserved:

$$\neg(x, y) \vee \neg(x', y') = \neg[(x, y) \wedge (x', y')]$$

$$\neg(x, y) \wedge \neg(x', y') = \neg[(x, y) \vee (x', y')].$$

Proof. We consider $(x, y) \wedge (x', y') = \min((x, y), (x', y')) = (x, y)$. The properties to demonstrate obviously hold when ($x = x'$) and ($y = y'$). When ($x \neq x'$) or ($y \neq y'$), we get ($x < x'$ or (($x = x'$) and ($y < y'$)). The complementation to one gives : ($1-x > 1-x'$) or ((1- x = 1-x') and ($1-y > 1-y'$)). As a consequence $(1 - x, 1 -y) \vee (1-x', 1-y') = (1-x, 1-y)$ which can be rewritten $\neg(x, y) \vee \neg(x', y') = \neg(x, y)$. Then we get $\neg(x, y) \vee \neg(x', y') = \neg(x, y) = \neg[(x, y) \wedge (x', y')]$. The first law is then demonstrated and the second one is valid .♦

Example 4. We consider the framework and the query of example 2 ("a *cheap* journey, if possible with a *small* flight time, and an hotel in a *small* town, if possible located *in the countryside*). Constraint C_1 is the fuzzy condition "a *cheap* journey", W_1 is the fuzzy wish "a *cheap* journey with a *small* flight time" and C_2 is the fuzzy constraint "hotel in a *small* town" while W_2 is the fuzzy wish "hotel in a *small* town and *in the countryside*".
When the norm $\wedge$ is chosen, the conjunction returns the worst case between (C_1,W_1) and (C_2,W_2) and the meaning of the query is a pure conjunction (the worst evaluation being retained). We get the results provided by Table 7.

Table 7 : Results provided by norm $\wedge$

#Trip	μ_C(#Trip)	μ_W(#Trip)
#1	0.9	0.7
#2	0.9	0.5
#3	0.9	0.3
#4	0.6	0.5
#5	0.6	0.5

We obtain the order : #1 > #2 > #3 > #4 = #5.

Here again, the obtained order is different from the ones provided by $O_{min,\ max}$ and $O_{min,\ min}$ (the meanings of the operator being different, cf example 3). In addition, one may remark that the fundamental property of fuzzy wish and constraint ($W \subseteq C$) holds in this case (since $\mu_W(\#Trip) \leq \mu_C(\#Trip)$ for any trip). ♦

5 Particular case of fuzzy conditions

This subsection shows the behaviours of the operators defined in the previous sections in the particular case where the fuzzy bipolar conditions are pure constraints ($W = C$).
When dealing with pure constraints, the extended norm $\wedge$ and co-norm $\vee$ are nothing but the norm min and the co-norm max of fuzzy sets.

Proof. We consider $(x, x) \wedge (y, y) = (x, x)$. It means (x, x) is the minimum when the lexicographical order is used which can be rewritten ($x < y$) or ($x = y$) and then:

$$(x, x) \wedge (y, y) = (x, x) = (\min(x, y), \min(x, y)).$$

The expression $(\min(x, y), \min(x, y))$ is the representation (in terms of constraints and wishes) of the minimum of the two degrees x and y which demonstrate that $(x, x) \wedge (y, y)$ represents the norm min. A similar proof holds for $\vee$ and the co-norm max.♦

The extended norm $\wedge$ and co-norm $\vee$ can also be defined by the parameterized operator $O_{op1,op2}$ where op_1 equals op_2 and is either the minimum or the maximum (it is not the case in general since $\wedge$ and $\vee$ cannot be expressed by $O_{op1,op2}$, as example $(0.6, 0.8) \wedge (0.5, 1) = (0.5, 1)$ while $(0.6, 0.8)\ O_{min,\ min}\ (0.5, 1) = (0.5, 0.8)$).

Proof. We consider $(x, x) \wedge (y, y) = (x, x)$ and $(x, x) \vee (y, y) = (y, y)$. It means ($x < y$) or ($x = y$) and then :

$$(x, x) \wedge (y, y) = (x, x) = (\min(x, y), \min(x, y))$$
$$= (x, x)\ O_{min,\ min}\ (y, y),$$
$$(x, x) \vee (y, y) = (y, y) = (\max(x, y), \max(x, y))$$
$$= (x, x)\ O_{max,\ max}\ (y, y).\ ♦$$

More generally, when op_1 equals op_2 the parameterized operator $O_{op1,op2}$ is either a norm or a co-norm between two fuzzy sets (the one chosen to implement op_1 and op_2).
The main question is about the meaning of $O_{op1,op2}$ when op_1 and op_2 differs. First of all, for a sake of simplicity, we assume that op_1 is a norm. Since the lexicographical order is used, it is possible to claim that $O_{op1,op2}$ allows for a refinement of norm op_1 since results are firstly ranked on the degree provided by op1, op_2 being only used to discriminate among elements having same degrees for the norm op_1. Similarly, when op1 is a co-norm, the lexicographical order provides a refinement of this co-norm.

ISBN: 978-989-95079-6-8

Example 5. We consider four items x,y,z and t which are valuated using a conjunction between the satisfactions of two fuzzy bipolar conditions FBC1 and FBC2 (cf. Table 8).

Table 8 : Satisfactions with respect to FBC1 and FBC2

item	FBC1	FBC2
x	(0.7, 0.7)	(0.8,0.8)
y	(0.5, 0.5)	(1, 1)
z	(0.5, 0.5)	(0.8, 0.8)
t	(0.3, 0.3)	(1, 1)

The satisfactions with respect to the conjunction are given by Table 9.

Table 9 : Satisfactions with respect to the conjunction

item	FBC1 $O_{min, min}$ FBC2	FBC1 $O_{min, max}$ FBC2
x	(0.7, 0.7)	(0.7,0.8)
y	(0.5, 0.5)	(0.5, 1)
z	(0.5, 0.5)	(0.5, 0.8)
t	(0.3, 0.3)	(0.3, 1)

When the minimum ($O_{min, min}$) is considered we get $x > y = z > t$, while $O_{min, max}$ gives $x > y > z > t$. The operator $O_{min, max}$ allows to discriminate elements y and z which are undistinguishable when $O_{min, min}$ is considered. ♦

6 Conclusion

This paper has dealt with the evaluation of fuzzy bipolar conditions into fuzzy queries addressed to a regular relational database (where the stored data are precisely known). Such conditions are made of two components: a mandatory condition (also called a constraint) and an optional condition (also called a wish). More precisely, we have considered fuzzy bipolar conditions where the constraint and the wish express user preferences and are defined by fuzzy sets. In particular, we have shown that a fuzzy bipolar condition is a generalization of a fuzzy condition (defined by a fuzzy set) and several operators to aggregate fuzzy bipolar conditions have been defined.

This paper has pointed out the importance to define an extended norm and an extended co-norm for fuzzy bipolar conditions. It proposes a couple of extended norm and co-norm which is a generalization of the couple (min, max) of fuzzy sets. In the particular case where fuzzy bipolar conditions are fuzzy conditions, some proposed operators leads to a refinement of norm and conorm as it is the case for the leximin and discrimin orders [13].

References

[1] Agrawal A. and Wimmers E.L. "A Framework for Expressing and Combining Preferences". *In Proc. of the 2000 ACM SIGMOD International Conference on Management of Data,* Dallas, USA, pp. 297–306, 2000.

[2] Bordogna G., Pasi G. "Linguistic aggregation operators of selection criteria in fuzzy information retrieval" In *Int. Jour. of Intelligent Systems*, vol. 10(2), pp.233–248, 1995.

[3] Borzsnyi, S. Kossmann D. and Stoker K. "The Skyline Operator". In Proc. of the 17th International Conference on Data Engineering, (ICDE01), Heidelberg, Germany, pp. 421–430, 2001.

[4] Bosc P., Pivert O. "SQLf: a relational database language for fuzzy querying". In *IEEE Transactions on Fuzzy Systems*, vol(3) pp.1–17, 1995.

[5] Chomicki J. "Preference Formulas in Relational Queries". In ACM Transactions on Database Systems, (TODS'03), 28(4):1–39, 2003.

[6] Kießling W. "Foundations of Preferences in Database Systems". In Proc. of the 28th International Conference on Very Large Data bases, (VLDB), Hong Kong, China, pp. 331–322, 2002.

[7] Kießling W. and Kostler G. "Preference SQL - Design, Implementation, Experiences". In Proc. of the 28th International Conference on Very Large Databases, (VLDB), Hong Kong, China, pp. 990–1001, 2002.

[8] Dubois D. and Prade H. "Handling Bipolar Queries in Fuzzy Information Processing". In Handbook of Research on Fuzzy Information Processing in Databases, J. Galindo (Ed.), 97–114, 2008.

[9] Dubois D. and Prade H. "An Introduction to Bipolar Representations of Information and Preference". In International Journal of Intelligent Systems, vol. 23, 866-877, 2008.

[10] Liétard L., Rocacher D. and Tbahriti S.-E., Preferences and Bipolarity in Query Language, International Conference of the North American Fuzzy Information Processing Society (NAFIPS 2008), New-York, USA.

[11] Zadrozny S. "Bipolar Queries Revisited". In Modeling Decisions for Artificial Intelligence, V. Torra et al. (Eds.), 387-398, 2005.

[12] Zadrozny S. and Kacprzyk J. "Bipolar Queries and Queries with Preferences". In Proceedings of the 17th International Conference on Database and Expert Systems Applications (DEXA'06), 415-419, 2006.

[13] Dubois D., Fargier H. and Prade H. "Beyond min aggregation in multi-criteria decision: (ordered) weighted Min, Discrimin, Leximin". In The Ordered Weighted Averaging Operators - Theory and Applications. R.R. Yager, J. Kacprzyk (Eds.), Kluwer Academic Publ., Boston, 181–192, 1997.

ISBN: 978-989-95079-6-8

An abstract approach to fuzzy logics: implicational semilinear logics

Petr Cintula[1] Carles Noguera[2]

1. Institute of Computer Science, Czech Academy of Sciences
Prague, Czech Republic
2. Department of Mathematics and Computer Science, University of Siena
Siena, Italy
Email: cintula@cs.cas.cz, cnoguera@iiia.csic.es

Abstract— This paper presents a new abstract framework to deal in a uniform way with the increasing variety of fuzzy logics studied in the literature. By means of notions and techniques from Abstract Algebraic Logic, we perform a study of non-classical logics based on the kind of generalized implication connectives they possess. It yields the new *hierarchy of implicational logics*. In this framework the notion of *implicational semilinear logic* can be naturally introduced as a property of the implication, namely a logic L is an implicational semilinear logic iff it has an implication such that L is complete w.r.t. the matrices where the implication induces a linear order, a property which is typically satisfied by well-known systems of fuzzy logic. The hierarchy of implicational logics is then restricted to the semilinear case obtaining a classification of implicational semilinear logics that encompasses almost all the known examples of fuzzy logics and suggests new directions for research in the field.

Keywords— Abstract Algebraic Logic, Implicative logics, Leibniz Hierarchy, Mathematical Fuzzy Logic, Semilinear logics.

1 Introduction

Mathematical Fuzzy Logic is the subdiscipline of Mathematical Logic which studies the logical systems that, since the inception of the theory of fuzzy sets, have been proposed to deal with the reasoning with predicates that can be modelled by fuzzy sets. The first ones, coming from the many-valued logic tradition, were Łukasiewicz and Gödel-Dummett logics, both complete w.r.t. the semantics given by a continuous t-norm. Later, a third system with this feature was introduced: product logic. Starting from these three main examples, the area has followed a long process of increasing generalization that has led to wider and wider classes of fuzzy logics. The first step was taken by Hájek [14] when he proposed the Basic fuzzy Logic BL, which turn out to be complete w.r.t. the semantics of all continuous t-norms. Later on, to put it in Hájek's words, scholars kept *removing legs from the flea* by considering weaker notions of fuzzy logic: divisibility was removed in the logic MTL [8] which is complete w.r.t. the semantics of all left-continuous t-norms, negation was removed when considering fuzzy logics based on hoops [9], commutativity of t-norms was disregarded in [13], and t-norms were replaced by *uninorms* in [16]. On the other hand, logics with a higher expressive power were introduced by considering expanded real-valued algebras (with projection Δ, involution $\sim$, truth-constants, etc.), and in recent works fuzzy logics have started emancipating from the real-valued algebras as the only intended semantics by considering systems complete w.r.t. rational, finite or hyperreal linearly ordered algebras [5].

When dealing with this huge variety of fuzzy logics one may want to have some tools to prove general results that apply not only to a particular logic, but to a class of logics. To some extent this has been achieved by means of the notions of core and Δ-core fuzzy logics [15] and results for these classes can be already found in a number of papers. However, those classes contain roughly just expansions of MTL and $\mathrm{MTL}_\triangle$ logics, so they do not cover weaker systems such as those from [16]. This shows that general notions of fuzzy logics are very useful, but we need to look for a more abstract framework to cope with all known examples and with other new logics that may arise in the near future.

In doing so, one certainly needs some intuition about the class of objects he would like to mathematically determine, namely some intuition of what are the minimal properties that should be required for a logic to be fuzzy. The evolution outlined above shows that almost no property of these systems was essential as they were step-by-step disregarded. Nevertheless, there is one that has remained untouched so far: *completeness w.r.t. a semantics based on linearly ordered algebras*. It actually corresponds to the main thesis of [1] that defends that *fuzzy logics are the logics of chains*. Such a claim must be read as a methodological statement, pointing at a roughly defined class of logics, rather than a precise mathematical description of what fuzzy logics are, since there could be many different ways in which a logic might enjoy a complete semantics based on chains. The aim of the present paper is to use some notions and techniques from Abstract Algebraic Logic (AAL) to provide a new framework where we can develop in a natural way a particular technical notion corresponding to the intuition of fuzzy logics as the logics of chains. Namely, we will present the *hierarchy of implicational logics* as a new classification of non-classical logics extending the well-known *Leibniz hierarchy* and encompassing other important classes such as implicative logics [18] and weakly implicative logics [4]. Inside this new hierarchy we will build a very general class of fuzzy logics that we will call *implicational semilinear logics*.

The technical aspects of this paper are based on the submitted work [6] however it concentrates on presenting the justification of our new framework from the point of view of Mathematical Fuzzy Logic. Section 2 describes our general setting and Section 3 informally summarizes our arguments. The remaining sections present samples of technical arguments supporting our thesis. Finally, Appendix A recalls crucial preliminary notions from the theory of logical calculi.

ISBN: 978-989-95079-6-8

2 The hierarchy of implicational logics

Although implication in the vast majority of existing fuzzy logics is given by a single (primitive or derived) connective, we follow a long-established tradition of Abstract Algebraic Logic and consider that implication could be given by a (possibly parameterized) set of formulae. However the following convention will allow us to hide this feature of our approach, providing a high level of abstraction without any apparent increase of complexity. Let $\Rightarrow(p, q, \overrightarrow{r})$ be a set of $\mathcal{L}$-formulae in two variables and, possibly, with a sequence of parameters $\overrightarrow{r}$. Given formulae φ, ψ and a sequence of formulae $\overrightarrow{\alpha}$, $\Rightarrow(\varphi, \psi, \overrightarrow{\alpha})$ denotes the set obtained by substituting the variables in $\Rightarrow(p, q, \overrightarrow{r})$ by the corresponding formulae, and $\varphi \Rightarrow \psi$ denotes the set $\bigcup\{\Rightarrow(\varphi, \psi, \overrightarrow{\alpha}) \mid \overrightarrow{\alpha} \in \mathrm{Fm}_{\mathcal{L}}^{<\omega}\}$.

We generalize the following properties, typically satisfied by an implication, to sets of (parameterized) formulae. However a reader can always understand these conditions as if $\Rightarrow$ would be just a single binary connective.

Definition 1. *Let* L *be a logic and* $\Rightarrow(p, q, \overrightarrow{r}) \subseteq \mathrm{Fm}_{\mathcal{L}}$ *be a parameterized set of formulae. We say that* $\Rightarrow$ *is a* weak p-implication *in* L *if:*

$$
\begin{array}{ll}
\text{(R)} & \vdash_{\mathrm{L}} \varphi \Rightarrow \varphi \\
\text{(MP)} & \varphi, \varphi \Rightarrow \psi \vdash_{\mathrm{L}} \psi \\
\text{(T)} & \varphi \Rightarrow \psi, \psi \Rightarrow \chi \vdash_{\mathrm{L}} \varphi \Rightarrow \chi \\
\text{(sCng)} & \varphi \Rightarrow \psi, \psi \Rightarrow \varphi \vdash_{\mathrm{L}} c(\chi_1, \ldots, \chi_i, \varphi, \ldots, \chi_n) \\
 & \Rightarrow c(\chi_1, \ldots, \chi_i, \psi, \ldots, \chi_n) \\
 & \text{for each } \langle c, n\rangle \in \mathcal{L} \text{ and each } i < n
\end{array}
$$

We change the prefix 'weak' to 'algebraic' if there is a set $\mathcal{E}(p)$ *of equations in one variable such that*

$$\text{(Alg)} \quad p \dashv\vdash_{\mathrm{L}} E[\mathcal{E}(p)],$$

where $E(p, q, \overrightarrow{r}) = \Rightarrow(p, q, \overrightarrow{r}) \cup \Rightarrow(q, p, \overrightarrow{r})$
We change the prefix 'weak' to 'regular' if:

$$\text{(Reg)} \quad \varphi, \psi \vdash_{\mathrm{L}} \psi \Rightarrow \varphi$$

We change the prefix 'weak' to 'Rasiowa' if:

$$\text{(W)} \quad \varphi \vdash_{\mathrm{L}} \psi \Rightarrow \varphi$$

Finally, if $\Rightarrow$ *is parameter-free we drop the prefix 'p-'.*

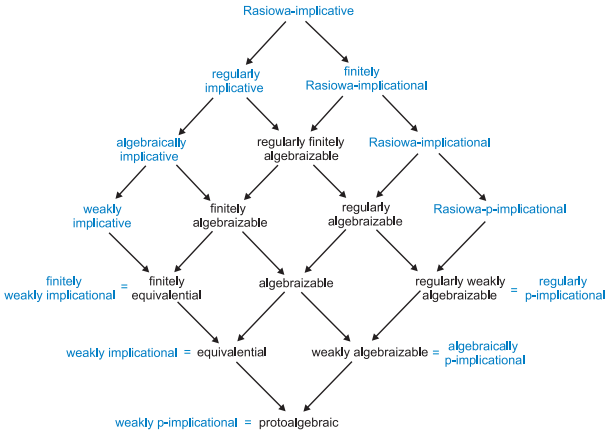
We can easily show the relative strength of defined notions: each Rasiowa p-implication is a regular p-implication and each regular p-implication is an algebraic p-implication.

Definition 2. *We say that a logic* L *is a* weakly/algebraically/regularly/Rasiowa- (p-)implicational logic *if there is a (parameterized) set of formulae* $\Rightarrow$ *which is a weak/algebraic/regular/Rasiowa (p-)implication in* L. *We add the prefix* 'finitely' *if* $\Rightarrow$ *is finite and we use the adjective* 'implicative' *instead of* 'implicational' *if* $\Rightarrow$ *is a* parameter-free singleton.

Each class of the well-known Leibniz hierarchy [7] coincides with some of our newly defined classes. In fact, our new taxonomy extends it, incorporates other already existing classes of logics,[1] and offers a more systematic way of classification: in one axis we go from p-implicational, implicational, finitely implicational to implicative (depending on the form of the implication set); in the second one we use prefixes 'weakly', 'algebraically', 'regularly', or 'Rasiowa-' (depending on extra properties fulfilled by that set). The translation table is:

Classes of Leibniz hierachy	*Our systematic names*
protoalgebraic	weakly p-implicational
(finitely) equivalential	(finitely) weakly impl.
weakly algebraizable	algebraically p-impl.
regularly weakly algebraizable	regularly p-implicational
(finitely) algebraizable	(finitely) algebraically impl.
(finitely) regularly algebraizable	(finitely) regularly impl.

Our new classification of logics, *the hierarchy of implicational logics* is depicted below (the arrows correspond to the class subsumption relation). We can show that almost all classes of logics in this hierarchy are mutually different; only the difference between Rasiowa-implicational and Rasiowa-p-implicational logics remains to be shown.



The syntactical notion of weak p-implication that we have introduced has a natural semantical counterpart: a preorder in the models that becomes an order in reduced models.

Definition 3. *Let* $\Rightarrow$ *be a parameterized set of formulae and* $\mathbf{A} = \langle \mathcal{A}, F\rangle$ *a matrix. We define a binary relation* $\leq_{\mathbf{A}}^{\Rightarrow}$ *on* A*:*

$$a \leq_{\mathbf{A}}^{\Rightarrow} b \qquad \textit{iff} \qquad a \Rightarrow^{\mathcal{A}} b \subseteq F.$$

Proposition 1. *Let* L *be a logic and* $\mathbf{A} \in \mathbf{MOD}(\mathrm{L})$*. Then a parameterized set* $\Rightarrow$ *is a weak p-implication in* L *iff* $\leq_{\mathbf{A}}^{\Rightarrow}$ *is a preorder and its symmetrization of* $\leq_{\mathbf{A}}^{\Rightarrow}$ *is the Leibniz congruence of* $\mathbf{A}$.

Clearly $\leq_{\mathbf{A}}^{\Rightarrow}$ is an order iff $\mathbf{A}$ is reduced. Thus (by virtue of Theorem 10) we can say that a L is complete w.r.t. the class of ordered matrices. Our main interest are the logics complete w.r.t. linearly ordered matrices in the following sense:

Definition 4. *Let* L *be a logic and* $\mathbf{A} = \mathbf{MOD}(\mathrm{L})$*. We say that* $\mathbf{A}$ *is a* linear model *w.r.t.* $\Rightarrow$ *if* $\leq_{\mathbf{A}}^{\Rightarrow}$ *is a linear order. The class of linear models of* L *is denoted by* $\mathbf{MOD}^{\ell}_{\Rightarrow}(\mathrm{L})$.

Observe that the class of linear models is not intrinsically defined for a given logic as it depends on the chosen implication. However, we will see later that in a reasonably wide class of logics all *semilinear implications* define the same linear models. But even in a general case we can prove:

Theorem 1. *Let* L *be a protoalgebraic logic. Then, for any weak p-implication* $\Rightarrow$*,* $\mathbf{MOD}^{\ell}_{\Rightarrow}(\mathrm{L}) \subseteq \mathbf{MOD}^{*}(\mathrm{L})_{\mathrm{RFSI}}$.

[1] Rasiowa-implicative logics were already defined in 1974 by Rasiowa [18] and weakly implicative logics in 2006 by Cintula [4].

ISBN: 978-989-95079-6-8

3 Semilinear implications and logics

Given a logic L and a weak p-implication $\Rightarrow$, we say that $\Rightarrow$ is a weak *semilinear* p-implication if the logic is complete w.r.t. the class of its corresponding linear models. Formally:

Definition 5. *Let* L *be a logic and* $\Rightarrow$ *a weak p-implication. We say that* $\Rightarrow$ *is a weak* semilinear *p-implication if*

$$\vdash_{\mathrm{L}} \;=\; \models_{\mathbf{MOD}^{\ell}_{\Rightarrow}(\mathrm{L})}.$$

Later we define *implicational semilinear logics* as those possessing some weak semilinear p-implication. Obviously, they will be fuzzy logics in the sense of [1]. However, we choose the term 'semilinear' instead of 'fuzzy' in spite of the fact that a first step towards the general definition we are offering here had been done by the first author in [4], when he defined the class of *weakly implicative fuzzy logics* (in our new framework: logics with a weak semilinear implication given a single binary connective). We have realized that the attempt of [4] of using the term 'fuzzy' to formally define a class of logics was rather futile as such word is heavily charged with many conflicting potential meanings.

Therefore, we have opted now for the new neutral name 'semilinear'. The term was first used by Olson and Raftery in [17] in the context of residuated lattices; it refers to the Universal Algebra tradition of calling a class of algebras 'semi$X$' whenever its subdirectly irreducible members have the property X (how this is related to our case will be apparent after Theorem 3). Despite using a new neutral name our intention remains the same: to formally delimit the class of fuzzy logics inside some existing abstract class of formal non-classical logics (originally, among weakly implicative ones now among the protoalgebraic ones). Of course it should include almost all the prominent examples of fuzzy logics known so far and exclude non-classical logics which are usually not recognized as fuzzy logics in the Logic community.

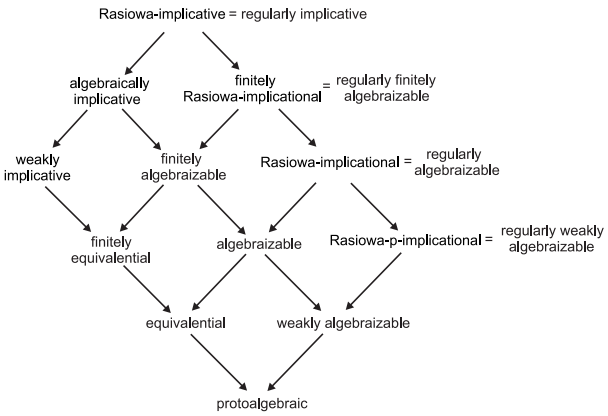
However let us stress that we *do not* expect to capture in a mathematical definition *the whole* intuitive notion of *arbitrary* fuzzy logic. Even if we would agree that linearity of semantics is crucial for a formal logic to be *fuzzy* there could be several other ways in which a logic might have a complete semantics somehow based on chains (see e.g. [2] or some recent work on modal fuzzy logics).

We formally define classes of implicational semilinear logics based on the form of semilinear implication they possess.

Definition 6. *We say that* L *is a* weakly/algebraically/ Rasiowa- (p-)implicational semilinear logic *if there is a (parameterized) set of formulae* $\Rightarrow$ *such that it is a weak/algebraic/Rasiowa semilinear (p-)implication in* L. *We add the* 'finitely' *if the set* $\Rightarrow$ *is finite and we use* 'implicative' *instead of* 'implicational' *if* $\Rightarrow$ *is a parameter-free singleton.*

We have not defined the class of *regularly* (p-)implicational (implicative) semilinear logics, because (as we will see in Corollary 2) we would obtain that each regularly p-implicational semilinear logic is a Rasiowa-p-implicational semilinear logic (and analogously for the other three Rasiowa-classes in the hierarchy of implicational logics). See all the classes and their inclusions in the next figure.

We can prove the mutual difference of many classes, but three differences remain to be seen: Rasiowa-implicational



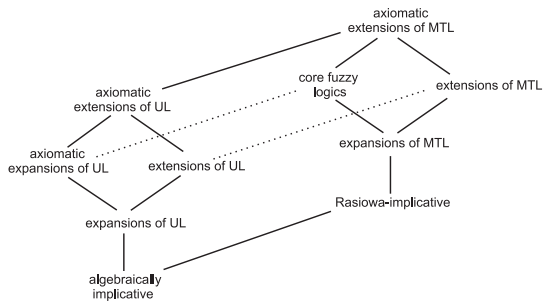
semilinear logics $\neq$ Rasiowa-p-implicational semilinear logics, algebraizable semilinear logics $\neq$ weakly algebraizable semilinear logics, and equivalential semilinear logics $\neq$ protoalgebraic semilinear logics.

Proposition 2. *Let* $\mathcal{X}$ *be any class in the hierarchy of implicational logics. Then, there is an* $\mathcal{X}$ *logic which is not an* $\mathcal{X}$ *semilinear logic.*

The three main logics based on continuous t-norms (Łukasiewicz, Gödel-Dummett, and Product logics) as well as the logic of all continuous t-norms BL are clearly Rasiowa-implicative semilinear logics. The same can be said in general as regards to left-continuous t-norm-based logics such as MTL and its t-norm based axiomatic extensions, and even for all *axiomatic extensions of MTL* (even those which are not complete w.r.t. a semantics of t-norms) as all of them are complete w.r.t. a subvariety of MTL-algebras generated by its linearly ordered members. Two incomparable superclasses of this one have been considered in the literature. On one hand, we have the so-called *core fuzzy logics* introduced in [15] as finitary logics expanding MTL or $\mathrm{MTL}_{\triangle}$, satisfying (sCng) for $\rightarrow$, and one of the following forms of Deduction Theorem: (i) $T, \varphi \vdash_{\mathrm{MTL}_{\triangle}} \psi$ iff $T \vdash_{\mathrm{MTL}_{\triangle}} \Delta\varphi \rightarrow \psi$, for expansions of $\mathrm{MTL}_{\triangle}$, or (ii) $T, \varphi \vdash_{\mathrm{MTL}} \psi$ iff there is $n \in \mathbb{N}$ such that $T \vdash_{\mathrm{MTL}} \varphi^n \rightarrow \psi$, for expansions of MTL. On the other hand, we can consider the class of all *semilinear finitary extensions of MTL*. Their equivalent quasivariety semantics are the subquasivarieties of MTL-algebras generated by chains. Since there are such quasivarieties that are not varieties, we have that this class is strictly bigger than that of axiomatic extensions of MTL. Both incomparable classes are included in the class of *semilinear expansions of MTL*, and finally this one is included in the Rasiowa-implicative semilinear logics. In the recent paper [16] the fuzzy logic UL based on uninorms instead of t-norms has been studied. It is an algebraizable logic without weakening, so it belongs to the class of algebraically implicative semilinear logics. We can consider the same structure of classes as above without weakening by replacing MTL for UL. See the resulting hierarchy of classes of semilinear logics in the next figure. We realize that all of them lie on the top of our classification, above Rasiowa-implicative or algebraically implicative semilinear logics. But if, by means of our definition of semilinear implication presented in this paper, we have succeeded in capturing an interesting way by which a logic can be fuzzy this means that fuzzy logics are a much wider class than those studied so far. Thus,

ISBN: 978-989-95079-6-8

future research in the field will probably bring new significant examples of fuzzy logics throughout the whole hierarchy of implicational semilinear logics.



4 Characterizations of semilinearity

This section provides some useful mathematical characterization of semilinear implications. First we define:

Definition 7. *Let* $\mathbf{A} = \langle \mathcal{A}, F \rangle \in \mathbf{MOD}(\mathrm{L})$. *The filter* F *is called* $\Rightarrow$-linear *if* $\leq^{\Rightarrow}_{\mathbf{A}}$ *is a total preorder.*

We say that L *has the* Linear Extension Property (LEP) *w.r.t.* $\Rightarrow$ *if for every theory* $T \in Th(\mathrm{L})$ *and every formula* $\varphi \in \mathrm{Fm}_{\mathcal{L}} \setminus T$, *there is a* $\Rightarrow$*-linear theory* $T' \supseteq T$ *s.t.* $\varphi \notin T'$.

Notice that a matrix $\langle \mathcal{A}, F \rangle \in \mathbf{MOD}(\mathrm{L})$ is in $\mathbf{MOD}^{\ell}_{\Rightarrow}(\mathrm{L})$ iff it is reduced and F is $\Rightarrow$-linear. Clearly the (LEP) says that the $\Rightarrow$-linear theories form a basis of the closure system $Th(\mathrm{L})$. Next theorem shows that an analogous statement holds for other than Lindenbaum matrices, as one of the so-called 'transfer principles' from AAL.

Theorem 2. *Let* L *be a finitary logic with* (LEP) *w.r.t.* $\Rightarrow$ *and* $\mathcal{A} \in \mathbf{ALG}^{*}(\mathrm{L})$. *Then* $\Rightarrow$*-linear filters form a basis of* $\mathcal{F}i_{\mathrm{L}}(\mathcal{A})$.

Next we generalize the 'Prelinearity property' from [4]. However, here we prefer the new name 'Semilinearity Property' following our new terminology.

Definition 8. *We say that* L *has the* Semilinearity Property (SLP) *w.r.t.* $\Rightarrow$ *if the following meta rule is valid:*

$$\frac{\Gamma, \varphi \Rightarrow \psi \vdash_{\mathrm{L}} \chi \qquad \Gamma, \psi \Rightarrow \varphi \vdash_{\mathrm{L}} \chi}{\Gamma \vdash_{\mathrm{L}} \chi}$$

Theorem 3 (Characterization of semilinear implications). *The following are equivalent:*

1. $\Rightarrow$ *is semilinear in* L,
2. L *has the* (LEP) *w.r.t.* $\Rightarrow$.

Furthermore, if L *is finitary we can add:*

3. L *has the* (SLP) *w.r.t.* $\Rightarrow$,
4. $\mathbf{MOD}^{*}(\mathrm{L})_{\mathrm{RSI}} \subseteq \mathbf{MOD}^{\ell}_{\Rightarrow}(\mathrm{L})$.

Moreover, if $\Rightarrow$ *is finite we can add:*

5. $\mathbf{MOD}^{*}(\mathrm{L})_{\mathrm{RFSI}} \subseteq \mathbf{MOD}^{\ell}_{\Rightarrow}(\mathrm{L})$.

The previous theorem has several important corollaries. Using Theorem 1 we obtain that, at least in a reasonably wide class of logics, being the class of linear models w.r.t. any finite semilinear implication is an intrinsic property of a logic.

Corollary 1. *Let* L *be a finitary protoalgebraic logic and* $\Rightarrow$ *a finite weak semilinear p-implication. Then* $\mathbf{MOD}^{*}(\mathrm{L})_{\mathrm{RFSI}} = \mathbf{MOD}^{\ell}_{\Rightarrow}(\mathrm{L})$.

Corollary 2. *Each regular semilinear p-implication is a Rasiowa p-implication.*

Corollary 3. *Let* $\Rightarrow$ *a weak semilinear p-implication in* L. *Then,* $\Rightarrow$ *is semilinear in all axiomatic extensions of* L.

This last corollary will be particulary useful for showing that some logic has *no* semilinear implication. It is quite easy to show that an implication in some logic is not semilinear, consider e.g. the normal implication of the intuitionistic logic; the well-know fact that the linear Heyting algebras do not generate the variety of Heyting algebras does the job. However using part 5. of the characterization theorem we can show much more: there is *no weak semilinear p-implication definable* in the intuitionistic logic, i.e. not only the standard nice Rasiowa implication given by a single formula is not semilinear but even using an infinite set with parameters we could never obtain an implication whose linearly ordered Heyting algebras would generate the variety of Heyting algebras.

Proposition 3. *Let* L *be the logic of a quasivariety of pointed residuated lattices containing the variety of Heyting algebras. Then,* L *is not weakly semilinear p-implicational logic.*

Many well-known logics fall under the scope of the previous proposition: Full Lambek logic (possibly extended with structural rules), multiplicative-additive fragment of (Affine) Intuitionistic Linear logic, Relevance logic R, etc.

Corollary 4. *Let* L *be a logic and* $\Rightarrow$ *a weak p-implication. Then, there is the weakest logic extending* L *where* $\Rightarrow$ *is semilinear. Let us denote this logic as* $\mathrm{L}^{\ell}_{\Rightarrow}$.

In Section 6 we will show how to axiomatize $\mathrm{L}^{\ell}_{\Rightarrow}$. However, to determine a complete semantics is simple:

Proposition 4. *Let* $\Rightarrow$ *be a weak p-implication in* L. *Then,* $\mathrm{L}^{\ell}_{\Rightarrow} = \models_{\mathbf{MOD}^{\ell}_{\Rightarrow}(\mathrm{L})}$ *and* $\mathbf{MOD}^{\ell}_{\Rightarrow}(\mathrm{L}^{\ell}_{\Rightarrow}) = \mathbf{MOD}^{\ell}_{\Rightarrow}(\mathrm{L})$.

5 Disjunctions

In order to provide additional characterizations of semilinearity (and to fill the gap between our abstract setting and real-life logics) we need to study a generalized notion of disjunction. As in the case of implication, given a parameterized set of formulae $\nabla(p, q, \overrightarrow{r})$ and formulae φ and ψ we define $\varphi \nabla \psi$.

Definition 9. *Given a logic* L *and a parameterized set of formulae* ∇, *we define the following properties:*

(PD)	$\varphi \vdash_{\mathrm{L}} \varphi \nabla \psi \quad$ *and* $\quad \psi \vdash_{\mathrm{L}} \varphi \nabla \psi$
(C)	$\varphi \nabla \psi \vdash_{\mathrm{L}} \psi \nabla \varphi$
(I)	$\varphi \nabla \varphi \vdash_{\mathrm{L}} \varphi$
(A)	$\varphi \nabla (\psi \nabla \chi) \dashv\vdash_{\mathrm{L}} (\varphi \nabla \psi) \nabla \chi$
(PCP)	*If* $\Gamma, \varphi \vdash_{\mathrm{L}} \chi$ *and* $\Gamma, \psi \vdash_{\mathrm{L}} \chi$, *then* $\Gamma, \varphi \nabla \psi \vdash_{\mathrm{L}} \chi$.

They correspond to well known usual properties of disjunction connectives. (C), (I) and (A) are respectively commutativity, idempotency and associativity, which are typically also satisfied by conjunction connectives. In contrast, the (PCP) is typically satisfied only by disjunction connectives. In [6] we also study a weaker variant of (PCP) for $\Gamma = \emptyset$.

ISBN: 978-989-95079-6-8

Definition 10. *Given a logic* L *and a parameterized set of formulae* $\nabla(p, q, \overrightarrow{r})$ *and a set of properties* $\sigma \subseteq \{(\mathrm{C}), (\mathrm{I}), (\mathrm{A})\}$*, we say that* ∇ *is a* σ-p-protodisjunction *in* L *if* (PD) *and the properties of* σ *are satisfied. Furthermore a p-protodisjunction* ∇ *is a* p-disjunction *if it satisfies* (PCP)*. Finally, if* ∇ *has no parameters we drop the prefix 'p-'.*

All these defined notions are mutually distinct and any p-disjunction is in fact a $\{(\mathrm{C}), (\mathrm{I}), (\mathrm{A})\}$-p-protodisjunction. Moreover, the notion of p-disjunction is intrinsic as any two p-disjunctions are mutually interderivable.

Definition 11. *We call a logic* (p-)disjunctional *if it has a (p-)disjunction. Furthermore, we call a logic* disjunctive *if it has a disjunction given by a single parameter-free formula.*

The classes of disjunctive and disjunctional logics are mutually different. E.g. the implication fragment of Gödel logic is not disjunctive but the set $\{(p \to q) \to q, (q \to p) \to p\}$ is its disjunction.

Definition 12. *Let* L *be a logic,* ∇ *a parameterized set of formulae,* $\mathcal{A} \in \mathbf{ALG}^*(\mathrm{L})$*, and* $F \in \mathcal{F}i_{\mathrm{L}}(\mathcal{A})$*. F is called* ∇-prime *if for every* $a, b \in A$*,* $a\nabla^{\mathcal{A}}b \subseteq F$ *iff* $a \in F$ *or* $b \in F$*.*

The *prime extension property* (PEP) is defined as the (LEP) by substituting the notion of $\Rightarrow$-linear filter for that of ∇-prime filter. Then we can prove:

Theorem 4. *Let* L *be a finitary logic and* ∇ *a p-protodisjunction. Then* L *has the* (PEP) *iff* ∇ *is p-disjunction.*

Theorem 5. *Let* L *be a finitary p-disjunctional logic. Then* ∇*-prime filters form a basis of* $\mathcal{F}i_{\mathrm{L}}(\mathcal{A})$ *for any* $\mathcal{A} \in \mathbf{ALG}^*(\mathrm{L})$*.*

Now we provide a syntactical characterization of (PCP). Let us by R^{∇} (for an $\mathcal{L}$-consecution $R = \Gamma \rhd \varphi$) denote the set $\{\Gamma\nabla\chi \rhd \delta \mid \chi \in \mathrm{Fm}_{\mathcal{L}}$ and $\delta \in \varphi\nabla\chi\}$.

Theorem 6. *Let* L *be a finitary logic with a presentation* $\mathcal{AS}$ *and* ∇ *a* $\{(\mathrm{C}), (\mathrm{I})\}$*-p-protodisjunction. Then, the following are equivalent:*

1. ∇ *is a p-disjunction,*
2. $R^{\nabla} \subseteq \mathrm{L}$ *for each (finitary)* $R \in \mathrm{L}$*,*
3. $R^{\nabla} \subseteq \mathrm{L}$ *for each* $R \in \mathcal{AS}$*.*

Corollary 5. *Let* ∇ *be a p-disjunction in a finitary logic* L_1 *and* L_2 *an expansion of* L_1 *by a set of consecutions* $\mathcal{C}$*. Then:*

- ∇ *is a p-disjunction in* L_2 *if* $R^{\nabla} \subseteq \mathrm{L}_2$ *for each* $R \in \mathcal{C}$*.*
- *If all the consecutions from* $\mathcal{C}$ *are finitary, then* $R^{\nabla} \subseteq \mathrm{L}_2$ *for each* $R \in \mathcal{C}$ *iff* ∇ *is a p-disjunction in* L_2*.*
- *If all the consecutions from* $\mathcal{C}$ *are axioms, then* ∇ *is a p-disjunction.*

Definition 13. *Let* L *be a logic and* ∇ *a parameterized set of formulae. We denote by* L^{∇} *the least logic extending* L *where* ∇ *is a p-disjunction.*

Theorem 7. *Let* L *be a finitary logic with a finitary presentation* $\mathcal{AS}$ *and* ∇ *a* $\{(\mathrm{C}), (\mathrm{I}), (\mathrm{A})\}$*-p-protodisjunction. Then,* L^{∇} *is axiomatized by* $\mathcal{AS} \cup \bigcup\{R^{\nabla} \mid R \in \mathcal{AS}\}$*.*

6 Disjunctions and semilinearity

In this section we consider the interesting relationships between the several kinds of disjunctions and implications we have defined and their corresponding properties. First, we introduce two natural syntactical conditions: a version of Modus Ponens with disjunction (DMP): $\varphi \Rightarrow \psi, \varphi\nabla\psi \vdash_{\mathrm{L}} \psi$ and $\varphi \Rightarrow \psi, \psi\nabla\varphi \vdash_{\mathrm{L}} \psi$, and a generalization of the prelinearity axiom used in fuzzy logics (P): $\vdash_{\mathrm{L}} (\varphi \Rightarrow \psi)\nabla(\psi \Rightarrow \varphi)$.

Theorem 8. *Let* L *be a logic,* ∇ *a p-protodisjunction, and* $\Rightarrow$ *a weak p-implication.*

- *If* L *fulfills* (DMP)*, we have:*
  1. *each* $\Rightarrow$*-linear theory is* ∇*-prime,*
  2. *if* $\Rightarrow$ *has the* (LEP)*, then* ∇ *has the* (PEP)*,*
  3. *if* $\Rightarrow$ *has the* (SLP)*, then* ∇ *has the* (PCP)*.*
- *If* L *fulfills* (P)*, we have:*
  4. *each* ∇*-prime theory is* $\Rightarrow$*-linear,*
  5. *if* ∇ *has the* (PEP)*, then* $\Rightarrow$ *has the* (LEP)*.*
- *If* L *fulfills* (P) *and either it is finitary or* $\Rightarrow$ *is finite and parameter-free, we have:*
  6. *if* ∇ *has the* (PCP)*, then* $\Rightarrow$ *has the* (SLP)*.*

This theorem together with known relations of the properties (SLP), (PCP), (PEP), (LEP) and semilinearity (Theorems 3 and 4) allows us to formulate numerous corollaries about their mutual relationships.

Corollary 6. *If* L *is finitary,* ∇ *is a p-protodisjunction, and* $\Rightarrow$ *a weak p-implication, the following are equivalent:*

1. L *satisfies* (DMP) *and* $\Rightarrow$ *is semilinear.*
2. L *satisfies* (DMP) *and* $\Rightarrow$ *has the* (SLP)*.*
3. L *satisfies* (DMP) *and* $\Rightarrow$ *has the* (LEP)*.*
4. L *satisfies* (P) *and* ∇ *has the* (PEP)*.*
5. L *satisfies* (P) *and* ∇ *has the* (PCP)*.*

Thus e.g. a weak p-implication in a finitary p-disjunctional logic L is semilinear iff L satisfies (P). In p-disjunctional logic we can strengthen two important results from the previous section. First, we can remove the precondition of finiteness of implication in Part 5. of Theorem 3.

Corollary 7. *Let* L *be a finitary p-disjunctional logic and* $\Rightarrow$ *a weak p-implication. Then the following are equivalent:*

1. $\Rightarrow$ *is semilinear in* L*,*
2. $\mathbf{MOD}^*(\mathrm{L})_{\mathrm{RFSI}} \subseteq \mathbf{MOD}^{\ell}_{\Rightarrow}(\mathrm{L})$*.*

Furthermore, in any finitary p-disjunctional protoalgebraic logic it holds that: $\mathbf{MOD}^*(\mathrm{L})_{\mathrm{RFSI}} = \mathbf{MOD}^{\ell}_{\Rightarrow}(\mathrm{L})$ *for any semilinear p-implication* $\Rightarrow$*.*

Theorem 9. *If* L *is finitary,* ∇ *is a* $\{(\mathrm{C}),(\mathrm{I}),(\mathrm{A})\}$*-p-protodisjunction,* $\Rightarrow$ *is a weak p-implication and* L *satisfies* (DMP)*, then* $\mathrm{L}^{\ell}_{\Rightarrow}$ *is the extension of* L^{∇} *by* (P)*.*

Corollary 8. *Let* L *be a finitary p-disjunctional logic and* $\Rightarrow$ *a weak p-implication. Then,* $\mathrm{L}^{\ell}_{\Rightarrow}$ *is extension of* L *by* (P)*.*

ISBN: 978-989-95079-6-8

A Bits of the theory of logical calculi

We recall some basic definitions and results of Abstract Algebraic Logic.[2] The notion of *propositional language* $\mathcal{L}$ is defined in the usual way (a set of connectives with finite arity). By $\mathbf{Fm}_{\mathcal{L}}$ we denote the free term algebra over a denumerable set of variables in the language $\mathcal{L}$, by $\mathrm{Fm}_{\mathcal{L}}$ we denote its universe and we call its elements $\mathcal{L}$*-formulae*.

A $\mathcal{L}$*-consecution* is a pair $\Gamma \rhd \varphi$, where $\Gamma \subseteq \mathrm{Fm}_{\mathcal{L}}$ and $\varphi \in \mathrm{Fm}_{\mathcal{L}}$. A consecution $\Gamma \rhd \varphi$ is *finitary* if Γ is finite. For a set of consecutions L we write $\Gamma \vdash_{\mathrm{L}} \varphi$ rather than $\Gamma \rhd \varphi \in \mathrm{L}$. A *propositional logic* is a pair $\mathrm{L} = \langle \mathcal{L}, \vdash_{\mathrm{L}} \rangle$ where $\vdash_{\mathrm{L}}$ is a structural consequence relation.

A logic L is *finitary* if for every $\Gamma \cup \{\varphi\} \subseteq \mathrm{Fm}_{\mathcal{L}}$ such that $\Gamma \vdash_{\mathrm{L}} \varphi$ there is a finite $\Gamma_0 \subseteq \Gamma$ such that $\Gamma_0 \vdash_{\mathrm{L}} \varphi$. We write $\Gamma \vdash_{\mathrm{L}} \Delta$ when $\Gamma \vdash_{\mathrm{L}} \varphi$ for every $\varphi \in \Delta$. A *theory* of a logic L is a set of formulae T such that if $T \vdash_{\mathrm{L}} \varphi$ then $\varphi \in T$. By $Th(\mathrm{L})$ we denote the set of all theories of L.

Given a *finitary* logic $\mathrm{L} = \langle \mathcal{L}, \vdash_{\mathrm{L}} \rangle$, we say that a set $\mathcal{AS}$ of $\mathcal{L}$-consecutions whose left member is finite is a *presentation* of L if the relation $\vdash_{\mathrm{L}}$ coincides with the provability relation given by $\mathcal{AS}$ as a Hilbert-style axiomatic system.

Given a language $\mathcal{L}$, an $\mathcal{L}$*-matrix* is a pair $\mathbf{A} = \langle \mathcal{A}, F \rangle$ where $\mathcal{A}$ is an $\mathcal{L}$-algebra and F is a subset of A called the *filter* of $\mathbf{A}$. A homomorphism from $\mathbf{Fm}_{\mathcal{L}}$ to $\mathcal{A}$ is called an $\mathbf{A}$*-evaluation*. The semantical consequence w.r.t. a class of matrices $\mathbb{K}$ is defined as: $\Gamma \models_{\mathbb{K}} \varphi$ iff for each $\mathbf{A} \in \mathbb{K}$ and each $\mathbf{A}$-evaluation e we obtain $e(\varphi) \in F$ whenever $e[\Gamma] \subseteq F$. Clearly, $\langle \mathcal{L}, \models_{\mathbb{K}} \rangle$ is a logic. We say that a matrix $\mathbf{A}$ is a *model* of L if $\vdash_{\mathrm{L}} \subseteq \models_{\mathbf{A}}$ and write $\mathbf{A} \in \mathbf{MOD}(\mathrm{L})$.

Given an $\mathcal{L}$-algebra $\mathcal{A}$, a subset $F \subseteq A$ is an L*-filter* if $\langle \mathcal{A}, F \rangle \in \mathbf{MOD}(\mathrm{L})$. Let $\mathcal{F}i_{\mathrm{L}}(\mathcal{A})$ be the set of all L-filters over $\mathcal{A}$. Observe that for every set of formulae T, we have $T \in Th(\mathrm{L})$ iff $\langle \mathbf{Fm}_{\mathcal{L}}, T \rangle \in \mathbf{MOD}(\mathrm{L})$; these models are called the *Lindenbaum matrices* for L.

It is straightforward to check that $\mathcal{F}i_{\mathrm{L}}(\mathcal{A})$ is closed under arbitrary intersections and hence it is a closure system. Recall that a family $\mathcal{B} \subseteq \mathcal{C}$ is a *basis* of a closure system $\mathcal{C}$ if for every $X \in \mathcal{C}$ there is a $\mathcal{D} \subseteq \mathcal{C}$ such that $X = \bigcap \mathcal{D}$ (which can be equivalent formulated as: for every $Y \in \mathcal{C}$ and every $a \in A \setminus Y$ there is $Z \in \mathcal{B}$ such that $Y \subseteq Z$ and $a \notin Z$).

Given a matrix $\mathbf{A} = \langle \mathcal{A}, F \rangle$, a binary relation $\Omega_{\mathcal{A}}(F)$ is defined as $\langle a, b \rangle \in \Omega_{\mathcal{A}}(F)$ if, and only if, for every sequence of parameters $\overrightarrow{z}$, $\mathcal{L}$-formula $\varphi(x, \overrightarrow{z})$, and $\overrightarrow{c} \in A^{<\omega}$ we have $\varphi^{\mathcal{A}}(a, \overrightarrow{c}) \in F$ iff $\varphi^{\mathcal{A}}(b, \overrightarrow{c}) \in F$. Inspired by the famous Leibniz's identity of indiscernibles principle, $\Omega_{\mathcal{A}}(F)$ is called the *Leibniz congruence* of $\langle \mathcal{A}, F \rangle$.

A matrix is said to be *reduced* if its Leibniz congruence is the identity relation. Given a logic L, the class of its reduced models is denoted by $\mathbf{MOD}^*(\mathrm{L})$, and the class of algebraic reducts of $\mathbf{MOD}^*(\mathrm{L})$ is denoted by $\mathbf{ALG}^*(\mathrm{L})$. They are enough to provide a complete semantics for the logic:

Theorem 10. *Let* L *be a logic. Then* $\Gamma \vdash_{\mathrm{L}} \varphi$ *if, and only if,* $\Gamma \models_{\mathbf{MOD}^*(\mathrm{L})} \varphi$*, for every set of formulae* $\Gamma \cup \{\varphi\}$.

A matrix $\mathbf{A} \in \mathbf{MOD}^*(\mathrm{L})$ is called *(finitely) subdirectly irreducible* if it is not a non-trivial (finite) subdirect product of reduced matrices. The corresponding classes of matrices are denoted as $\mathbf{MOD}^*(\mathrm{L})_{\mathrm{RSI}}$ and $\mathbf{MOD}^*(\mathrm{L})_{\mathrm{RFSI}}$ respectively.

[2] For a comprehensive survey see [7, 10, 11]. Any necessary background on Universal Algebra can be found e.g. in [3].

Acknowledgments

Petr Cintula was partly supported by ESF EUROCORES project ICC/08/E018 of the Grant Agency of the Czech Republic and partly by Institutional Research Plan AVOZ10300504. Carles Noguera acknowledges partial support from the grant 2006-BP-A-10043 of the Departament d'Educació i Universitats of the Generalitat de Catalunya and the Spanish project MULOG2 (TIN2007-68005-C04).

References

[1] Libor Běhounek and Petr Cintula. Fuzzy logics as the logics of chains. *Fuzzy Sets and Systems*, 157(5):604–610, 2006.

[2] Félix Bou, Francesc Esteva, Josep Maria Font, Àngel Gil, Lluís Godo, Antoni Torrens, and Ventura Verdú. Logics preserving degrees of truth from varieties of residuated lattices. To appear in *Journal of Logic and Computation*, 2009.

[3] Stanley Burris and H.P. Sankappanavar. *A Course in Universal Algebra*, volume 78 of *Graduate Texts in Mathematics*. Springer-Verlag, 1981.

[4] Petr Cintula. Weakly implicative (fuzzy) logics I: Basic properties. *Archive for Mathematical Logic*, 45(6):673–704, 2006.

[5] Petr Cintula, Francesc Esteva, Joan Gispert, Lluís Godo, Franco Montagna, and Carles Noguera. Distinguished algebraic semantics for t-norm based fuzzy logics: Methods and algebraic equivalencies. *Annals of Pure and Applied Logic*, 160(1):53–81, 2009

[6] Petr Cintula and Carles Noguera. A hierarchy of implicational (semilinear) logics: the propositional case. Submitted, 2009. (Preprint available at http://www.cs.cas.cz/mathfuzzlog)

[7] Janusz Czelakowski. *Protoalgebraic Logics*, volume 10 of *Trends in Logic*. Kluwer, Dordercht, 2001.

[8] Francesc Esteva and Lluís Godo. Monoidal t-norm based logic: Towards a logic for left-continuous t-norms. *Fuzzy Sets and Systems*, 124(3):271–288, 2001.

[9] Francesc Esteva, Lluís Godo, Petr Hájek, and Franco Montagna. Hoops and Fuzzy Logic, *Journal of Logic and Computation*, 13(4):531–555, 2003.

[10] Josep Maria Font and Ramon Jansana. *A General Algebraic Semantics for Sentential Logics*. Springer-Verlag, 1996.

[11] Josep Maria Font, Ramon Jansana, and Don Pigozzi. A survey of Abstract Algebraic Logic. *Studia Logica*, 74(1–2, Special Issue on Abstract Algebraic Logic II):13–97, 2003.

[12] Nikolaos Galatos, Peter Jipsen, Tomasz Kowalski, and Hiroakira Ono. *Residuated Lattices: An Algebraic Glimpse at Substructural Logics*, volume 151 of *Studies in Logic and the Foundations of Mathematics*. Elsevier, Amsterdam, 2007.

[13] Petr Hájek. Fuzzy Logics with Noncommutative Conjunctions. *Journal of Logic and Computation* 13(4):469–479, 2003.

[14] Petr Hájek. *Metamathematics of Fuzzy Logic*, volume 4 of *Trends in Logic*. Kluwer, Dordercht, 1998.

[15] Petr Hájek and Petr Cintula. On theories and models in fuzzy predicate logics. *Journal of Symbolic Logic*, 71(3):863–880, 2006.

[16] George Metcalfe and Franco Montagna. Substructural fuzzy logics. *Journal of Symbolic Logic*, 72(3):834–864, 2007.

[17] Jeffrey S. Olson and James G. Raftery. Positive Sugihara monoids. *Algebra Universalis*, 57:75–99, 2007.

[18] Helena Rasiowa. *An Algebraic Approach to Non-Classical Logics*. North-Holland, Amsterdam, 1974.

ISBN: 978-989-95079-6-8

Polynomial functions on bounded chains

Miguel Couceiro Jean-Luc Marichal

Mathematics Research Unit, University of Luxembourg
162A, avenue de la Faïencerie, L-1511 Luxembourg, Luxembourg
Email: {miguel.couceiro, jean-luc.marichal}@uni.lu

Abstract— *We are interested in representations and characterizations of lattice polynomial functions $f : L^n \to L$, where L is a given bounded distributive lattice. In an earlier paper [4, 5], we investigated certain representations and provided various characterizations of these functions both as solutions of certain functional equations and in terms of necessary and sufficient conditions. In the present paper, we investigate these representations and characterizations in the special case when L is a chain, i.e., a totally ordered lattice. More precisely, we discuss representations of lattice polynomial functions given in terms of standard simplices and we present new axiomatizations of these functions by relaxing some of the conditions given in [4, 5] and by considering further conditions, namely comonotonic minitivity and maxitivity.*

Keywords— Lattice polynomial function, discrete Sugeno integral, normal form, homogeneity, strong idempotency, median decomposability, comonotonicity.

1 Introduction

In [5], the class of (*lattice*) *polynomial functions*, i.e., functions representable by combinations of variables and constants using the lattice operations $\wedge$ and $\vee$, was considered and characterized both as solutions of certain functional equations and in terms of necessary and sufficient conditions rooted in aggregation theory.

Formally, let L be a bounded distributive lattice with operations $\wedge$ and $\vee$, and with least and greatest elements 0 and 1, respectively. An *n-ary polynomial function* on L is any function $f : L^n \to L$ which can be obtained by finitely many applications of the following rules:

(i) For each $i \in [n] = \{1, \ldots, n\}$ and each $c \in L$, the projection $\mathbf{x} \mapsto x_i$ and the constant function $\mathbf{x} \mapsto c$ are polynomial functions from L^n to L.

(ii) If f and g are polynomial functions from L^n to L, then $f \vee g$ and $f \wedge g$ are polynomial functions from L^n to L.

Polynomial functions are also called lattice functions (Goodstein [9]), algebraic functions (Burris and Sankappanavar [3]) or weighted lattice polynomial functions (Marichal [15]). Polynomial functions obtained from projections by finitely many applications of (ii) are referred to as *lattice term functions*.

The recent interest by aggregation theorists in this class of polynomial functions is partially motivated by its connection to noteworthy aggregation functions such as the (*discrete*) *Sugeno integral*, which was introduced by Sugeno [18, 19] and widely investigated in aggregation theory, due to the many applications in fuzzy set theory, data fusion, decision making, image analysis, etc. As shown in [15], the discrete Sugeno integrals are nothing other than those polynomial functions $f : L^n \to L$ which are idempotent, that is, satisfying $f(x, \ldots, x) = x$. For general background on aggregation theory, see [1, 11] and for a recent reference, see [10].

In this paper, we refine our previous results in the particular case when L is a chain, by relaxing our conditions and proposing weak analogues of those properties used in [4, 5], and then providing characterizations of polynomial functions, accordingly. Moreover, and motivated by the axiomatizations of the discrete Sugeno integrals established by de Campos and Bolaños [6] (in the case when $L = [0, 1]$ is the unit real interval), we present further and alternative characterizations of polynomial functions given in terms of comonotonic minitivity and maxitivity. As particular cases, we consider the subclass of discrete Sugeno integrals.

The current paper is organized as follows. We start in §2 by introducing the basic notions needed in this paper and presenting the characterizations of lattice polynomial functions on arbitrary (possibly infinite) bounded distributive lattices, established in [4, 5]. Those characterizations are reassembled in Theorem 1. We discuss representations of polynomial functions in normal form (such as the classical disjunctive and conjunctive normal forms) and introduce variant representations in the case of chains and given in terms of standard simplices in §3. In §4, we present characterizations of polynomial functions on chains given in terms of weak analogues of the properties used in Theorem 1 as well as in terms of comonotonic minitivity and maxitivity. Axiomatizations for the subclass of discrete Sugeno integrals are then presented in §5.

2 Basic notions and terminology

Throughout this paper, let L be a bounded chain with operations $\wedge$ and $\vee$, and with least and greatest elements 0 and 1, respectively. A subset S of a chain L is said to be *convex* if for every $a, b \in S$ and every $c \in L$ such that $a \leqslant c \leqslant b$, we have $c \in S$. For any subset $S \subseteq L$, we denote by $\overline{S}$ the convex hull of S, that is, the smallest convex subset of L containing S. For every $a, b \in S$ such that $a \leqslant b$, the *interval* $[a, b]$ is the set $[a, b] = \{c \in L : a \leqslant c \leqslant b\}$. For any integer $n \geqslant 1$, let $[n] = \{1, \ldots, n\}$.

For any bounded chain L, we regard the Cartesian product L^n, $n \geqslant 1$, as a distributive lattice endowed with the operations $\wedge$ and $\vee$ given by

$$\begin{aligned}(a_1, \ldots, a_n) \wedge (b_1, \ldots, b_n) &= (a_1 \wedge b_1, \ldots, a_n \wedge b_n),\\ (a_1, \ldots, a_n) \vee (b_1, \ldots, b_n) &= (a_1 \vee b_1, \ldots, a_n \vee b_n).\end{aligned}$$

The elements of L are denoted by lower case letters $a, b, c, \ldots$, and the elements of L^n, $n > 1$, by bold face letters $\mathbf{a}, \mathbf{b}, \mathbf{c}, \ldots$.

ISBN: 978-989-95079-6-8

We also use $\mathbf{0}$ and $\mathbf{1}$ to denote the least element and greatest element, respectively, of L^n. For $c \in L$ and $\mathbf{x} = (x_1, \ldots, x_n) \in L^n$, set

$$\mathbf{x} \wedge c = (x_1 \wedge c, \ldots, x_n \wedge c) \quad \text{and} \quad \mathbf{x} \vee c = (x_1 \vee c, \ldots, x_n \vee c).$$

The *range* of a function $f\colon L^n \to L$ is defined by $\mathcal{R}_f = \{f(\mathbf{x}) : \mathbf{x} \in L^n\}$. A function $f\colon L^n \to L$ is said to be *nondecreasing* (*in each variable*) if, for every $\mathbf{a}, \mathbf{b} \in L^n$ such that $\mathbf{a} \leqslant \mathbf{b}$, we have $f(\mathbf{a}) \leqslant f(\mathbf{b})$. Note that if f is nondecreasing, then $\overline{\mathcal{R}}_f = [f(\mathbf{0}), f(\mathbf{1})]$.

Let S be a nonempty subset of L. A function $f\colon L^n \to L$ is said to be

- *S-idempotent* if for every $c \in S$, we have $f(c, ..., c) = c$.
- *S-min homogeneous* if for every $\mathbf{x} \in L^n$ and $c \in S$,
$$f(\mathbf{x} \wedge c) = f(\mathbf{x}) \wedge c. \tag{1}$$
- *S-max homogeneous* if for every $\mathbf{x} \in L^n$ and $c \in S$,
$$f(\mathbf{x} \vee c) = f(\mathbf{x}) \vee c. \tag{2}$$
- *horizontally S-minitive* if for every $\mathbf{x} \in L^n$ and $c \in S$,
$$f(\mathbf{x}) = f(\mathbf{x} \vee c) \wedge f([\mathbf{x}]^c), \tag{3}$$
where $[\mathbf{x}]^c$ is the n-tuple whose ith component is 1, if $x_i \geqslant c$, and x_i, otherwise.
- *horizontally S-maxitive* if for every $\mathbf{x} \in L^n$ and $c \in S$,
$$f(\mathbf{x}) = f(\mathbf{x} \wedge c) \vee f([\mathbf{x}]_c), \tag{4}$$
where $[\mathbf{x}]_c$ is the n-tuple whose ith component is 0, if $x_i \leqslant c$, and x_i, otherwise.
- *median decomposable* if, for every $\mathbf{x} \in L^n$ and $k \in [n]$,
$$f(\mathbf{x}) = \operatorname{median}\big(f(\mathbf{x}_k^0), x_k, f(\mathbf{x}_k^1)\big), \tag{5}$$
where $\operatorname{median}(x, y, z) = (x \wedge y) \vee (y \wedge z) \vee (z \wedge x)$ and $\mathbf{x}_k^c = (x_1, \ldots, x_{k-1}, c, x_{k+1}, \ldots, x_n)$ for $c \in L$.
- *strongly idempotent* if, for every $\mathbf{x} \in L^n$ and $k \in [n]$,
$$f(x_1, \ldots, x_{k-1}, f(\mathbf{x}), x_{k+1}, \ldots, x_n) = f(\mathbf{x}).$$

Remark 1. In the case when $S = L$ is the real interval $[0, 1]$, the concepts of S-min and S-max homogeneity were introduced by Fodor and Roubens [8] to specify certain classes of aggregation functions, and the concept of horizontal S-maxitivity was introduced by Benvenuti et al. [2] as a general property of the Sugeno integral. The concept of median decomposability was introduced in [15] and that of strong idempotency in [4, 5] as properties of polynomial functions.

We say that a function $f\colon L^n \to L$ has a *componentwise convex range* if either $n = 1$ and f has a convex range, or $n > 1$ and for every $\mathbf{a} = (a_1, \ldots, a_n) \in L^n$ and every $k \in [n]$, the unary function $f_{\mathbf{a}}^k\colon L \to L$, given by $f_{\mathbf{a}}^k(x) = f(\mathbf{a}_k^x)$ has a convex range.

The following theorem reassembles the various characterizations of polynomial functions, established in [4], in the particular case when L is a bounded chain.

Theorem 1. *Let $f\colon L^n \to L$ be a function. The following conditions are equivalent:*

(i) *f is a polynomial function.*

(ii) *f is median decomposable.*

(iii) *f is nondecreasing, strongly idempotent, has a convex range and a componentwise convex range.*

(iv) *f is nondecreasing, $\overline{\mathcal{R}}_f$-min homogeneous, and $\overline{\mathcal{R}}_f$-max homogeneous.*

(v) *f is nondecreasing, $\overline{\mathcal{R}}_f$-min homogeneous, and horizontally $\overline{\mathcal{R}}_f$-maxitive.*

(vi) *f is nondecreasing, horizontally $\overline{\mathcal{R}}_f$-minitive, and $\overline{\mathcal{R}}_f$-max homogeneous.*

(vii) *f is nondecreasing, $\overline{\mathcal{R}}_f$-idempotent, horizontally $\overline{\mathcal{R}}_f$-minitive, and horizontally $\overline{\mathcal{R}}_f$-maxitive.*

3 Representations of polynomial functions

Polynomial functions are known to be exactly those functions which can be represented by formulas in disjunctive and conjunctive normal forms. This fact was first observed by Goodstein [9] who, in fact, showed that each polynomial function $f\colon L^n \to L$ is uniquely determined by its restriction to $\{0, 1\}^n$. For a recent reference, see Rudeanu [17].

In this section we recall and refine some known results concerning normal forms of polynomial functions and, in the special case when L is a chain, we present variant representations given in terms of standard simplices of L^n.

The following three results are due to Goodstein [9].

Corollary 2. *Every polynomial function is completely determined by its restriction to $\{0, 1\}^n$.*

Corollary 3. *A function $g\colon \{0, 1\}^n \to L$ can be extended to a polynomial function $f\colon L^n \to L$ if and only if it is nondecreasing. In this case, the extension is unique.*

Proposition 4. *Let $f\colon L^n \to L$ be a function. The following conditions are equivalent:*

(i) *f is a polynomial function.*

(ii) *There exists $\alpha\colon 2^{[n]} \to L$ such that*
$$f(\mathbf{x}) = \bigvee_{I \subseteq [n]} \big(\alpha(I) \wedge \bigwedge_{i \in I} x_i\big).$$

(iii) *There exists $\beta\colon 2^{[n]} \to L$ such that*
$$f(\mathbf{x}) = \bigwedge_{I \subseteq [n]} \big(\beta(I) \vee \bigvee_{i \in I} x_i\big).$$

We shall refer to the expressions given in (ii) and (iii) of Proposition 4 as the *disjunctive normal form* (DNF) representation and the *conjunctive normal form* (CNF) representation, respectively, of the polynomial function f.

Remark 2. By requiring α and β to be nonconstant functions from $2^{[n]}$ to $\{0, 1\}$ and satisfying $\alpha(\varnothing) = 0$ and $\beta(\varnothing) = 1$, respectively, we obtain the analogue of Proposition 4 for term functions.

ISBN: 978-989-95079-6-8

For each polynomial function $f\colon L^n \to L$, set

$$\mathrm{DNF}(f) = \Big\{\alpha \in L^{2^{[n]}} : f(\mathbf{x}) = \bigvee_{I\subseteq[n]} \big(\alpha(I) \wedge \bigwedge_{i\in I} x_i\big)\Big\},$$

$$\mathrm{CNF}(f) = \Big\{\beta \in L^{2^{[n]}} : f(\mathbf{x}) = \bigwedge_{I\subseteq[n]} \big(\beta(I) \vee \bigvee_{i\in I} x_i\big)\Big\}.$$

A complete description of the sets $\mathrm{DNF}(f)$ and $\mathrm{CNF}(f)$ can be found in [5, §3.1]. As we are concerned by the case when L is a chain, we recall the description only in this special case; see [15, §3].

For each $I \subseteq [n]$, let $\mathbf{e}_I$ be the element of L^n whose ith component is 1, if $i \in I$, and 0, otherwise. Let $\alpha_f\colon 2^{[n]} \to L$ be the function given by $\alpha_f(I) = f(\mathbf{e}_I)$ and consider the function $\alpha_f^*\colon 2^{[n]} \to L$ defined by

$$\alpha_f^*(I) = \begin{cases} \alpha_f(I), & \text{if } \bigvee_{J\subsetneq I} \alpha_f(J) < \alpha_f(I),\\ 0, & \text{otherwise.}\end{cases}$$

Dually, Let $\beta_f\colon 2^{[n]} \to L$ be the function given by $\beta_f(I) = f(\mathbf{e}_{[n]\setminus I})$ and consider the function $\beta_f^*\colon 2^{[n]} \to L$ defined by

$$\beta_f^*(I) = \begin{cases} \beta_f(I), & \text{if } \bigwedge_{J\subsetneq I} \beta_f(J) > \beta_f(I),\\ 1, & \text{otherwise.}\end{cases}$$

Proposition 5. *Let $f\colon L^n \to L$ be a polynomial function. Then*

(i) $\mathrm{DNF}(f) = [\alpha_f^*, \alpha_f]$ and f has a unique DNF representation if and only if $\bigvee_{J\subsetneq I} \alpha_f(J) < \alpha_f(I)$ for every $I \subseteq [n]$,

(ii) $\mathrm{CNF}(f) = [\beta_f, \beta_f^*]$ and f has a unique CNF representation if and only if $\bigwedge_{J\subsetneq I} \beta_f(J) > \beta_f(I)$ for every $I \subseteq [n]$.

In particular, α_f and β_f are the unique isotone and antitone, respectively, maps in $\mathrm{DNF}(f)$ and $\mathrm{CNF}(f)$, respectively.

In the case of chains, the DNF and CNF representations of polynomial functions $f\colon L^n \to L$ can be refined and given in terms of standard simplices of L^n (see Proposition 6 below). Recall that

$$\mathrm{median}(x_1, \ldots, x_{2n+1}) = \bigvee_{\substack{I\subseteq[2n+1]\\ |I|=n+1}} \bigwedge_{i\in I} x_i.$$

Let σ be a permutation on $[n]$. The *standard simplex* of L^n associated with σ is the subset $L_\sigma^n \subset L^n$ defined by

$$L_\sigma^n = \{(x_1, \ldots, x_n) \in L^n : x_{\sigma(1)} \leqslant \cdots \leqslant x_{\sigma(n)}\}.$$

For each $i \in [n]$, define $S_\sigma^{\uparrow}(i) = \{\sigma(i), \ldots, \sigma(n)\}$ and $S_\sigma^{\downarrow}(i) = \{\sigma(1), \ldots, \sigma(i)\}$. As a matter of convenience, set $S_\sigma^{\uparrow}(n+1) = S_\sigma^{\downarrow}(0) = \varnothing$.

Proposition 6. *Let $f\colon L^n \to L$ be a function. The following conditions are equivalent:*

(i) f is a polynomial function.

(ii) For any permutation σ on $[n]$ and every $\mathbf{x} \in L_\sigma^n$, we have

$$\begin{aligned} f(\mathbf{x}) &= \bigvee_{i\in[n+1]} \big(\alpha_f(S_\sigma^{\uparrow}(i)) \wedge x_{\sigma(i)}\big)\\ &= \bigwedge_{i\in[n+1]} \big(\alpha_f(S_\sigma^{\uparrow}(i)) \vee x_{\sigma(i-1)}\big)\\ &= \mathrm{median}\big(x_1, ..., x_n, \alpha_f(S_\sigma^{\uparrow}(1)), ..., \alpha_f(S_\sigma^{\uparrow}(n+1))\big),\end{aligned}$$

where $x_{\sigma(0)} = 0$ and $x_{\sigma(n+1)} = 1$.

(iii) For any permutation σ on $[n]$ and every $\mathbf{x} \in L_\sigma^n$, we have

$$\begin{aligned} f(\mathbf{x}) &= \bigvee_{i\in[n+1]} \big(\beta_f(S_\sigma^{\downarrow}(i-1)) \wedge x_{\sigma(i)}\big)\\ &= \bigwedge_{i\in[n+1]} \big(\beta_f(S_\sigma^{\downarrow}(i-1)) \wedge x_{\sigma(i-1)}\big)\\ &= \mathrm{median}\big(x_1, \ldots, x_n, \beta_f(S_\sigma^{\downarrow}(0)), \ldots, \beta_f(S_\sigma^{\downarrow}(n))\big),\end{aligned}$$

where $x_{\sigma(0)} = 0$ and $x_{\sigma(n+1)} = 1$.

Remark 3. The equivalence between (i) and (ii) of Proposition 6 was already observed in [15, §5]. Prior to this, Propositions 5 and 6 were already established in [14] for idempotent polynomial functions (discrete Sugeno integrals) in the case when L is the unit real interval $[0, 1]$; see also [13, §4.3].

4 Characterizations of polynomial functions

In this section, we propose weak analogues of the properties used in Theorem 1 and provide characterizations of polynomial functions on chains, accordingly. Moreover, we introduce further properties, namely, comonotonic minitivity and maxitivity, which we then use to provide further characterizations of polynomial functions.

For integers $0 \leqslant p \leqslant q \leqslant n$, define $L_n^{(p,q)} = \{\mathbf{x} \in L^n : |\{x_1, \ldots, x_n\} \cap \{0, 1\}| \geqslant p \text{ and } |\{x_1, \ldots, x_n\}| \leqslant q\}$. For instance, $L_n^{(0,2)}$ is the set of Boolean vectors of L^n that are two-sided trimmed by constant vectors, that is

$$L_n^{(0,2)} = \bigcup_{\substack{\mathbf{e}\in\{0,1\}^n\\ c,d\in L}} \{\mathrm{median}(c, \mathbf{e}, d)\},$$

where the median is taken componentwise.

4.1 Weak homogeneity

Let S be a nonempty subset of L. We say that a function $f\colon L^n \to L$ is *weakly S-min homogeneous* (resp. *weakly S-max homogeneous*) if (1) (resp. (2)) holds for every $\mathbf{x} \in L_n^{(0,2)}$ and every $c \in S$.

For every integer $m \geqslant 1$, every $\mathbf{x} \in L^m$, and every $f\colon L^n \to L$, we define $\langle \mathbf{x} \rangle_f \in L^m$ as the m-tuple

$$\langle \mathbf{x} \rangle_f = \mathrm{median}(f(\mathbf{0}), \mathbf{x}, f(\mathbf{1})),$$

where the right-hand side median is taken componentwise. As observed in [5], for every nonempty subset $S \subseteq L$, we have that f is S-min homogeneous and S-max homogeneous if and only if it satisfies

$$f(\mathrm{median}(r, \mathbf{x}, s)) = \mathrm{median}(r, f(\mathbf{x}), s)$$

ISBN: 978-989-95079-6-8

for every $\mathbf{x} \in L^n$ and every $r, s \in S$. In particular, if $f(\mathbf{0}), f(\mathbf{1}) \in S$ then, for any $\mathbf{x} \in L^n$ such that $f(\mathbf{0}) \leqslant f(\mathbf{x}) \leqslant f(\mathbf{1})$, we have $f(\mathbf{x}) = f(\langle \mathbf{x} \rangle_f)$.

It was also shown in [5] that, for every nonempty subset $S \subseteq L$, if f is S-min homogeneous and S-max homogeneous, then it is S-idempotent. The following lemma shows that the weak analogue also holds.

Lemma 7. *Let S be a nonempty subset of L. If $f\colon L^n \to L$ is weakly S-min homogeneous and weakly S-max homogeneous, then it is S-idempotent. Moreover, if $f(\mathbf{0}), f(\mathbf{1}) \in S$ then, for any $\mathbf{x} \in L_n^{(0,2)}$ such that $f(\mathbf{0}) \leqslant f(\mathbf{x}) \leqslant f(\mathbf{1})$, we have $f(\mathbf{x}) = f(\langle \mathbf{x} \rangle_f)$.*

As we are going to see, of particular interest is when $S = \overline{\mathcal{R}}_f$. The following result characterizing the class of polynomial functions shows that, in the case of chains, the conditions in (iv) of Theorem 1 can be replaced with their weak analogues.

Theorem 8. *A function $f\colon L^n \to L$ is a polynomial function if and only if it is nondecreasing, weakly $\overline{\mathcal{R}}_f$-min homogeneous, and weakly $\overline{\mathcal{R}}_f$-max homogeneous.*

Remark 4. Note that Theorem 8 does not generally hold in the case of bounded distributive lattices. To see this, let $L = \{0, a, b, 1\}$ where $a \wedge b = 0$ and $a \vee b = 1$, and consider the binary function $f\colon L^2 \to L$ defined by

$$f(x_1, x_2) = \begin{cases} 1, & \text{if } x_1 = 1 \text{ or } x_2 = 1, \\ 1, & \text{if } x_1 = x_2 = b, \\ a, & \text{if } (x_1 = a \text{ and } x_2 \neq 1) \\ & \text{or } (x_2 = a \text{ and } x_1 \neq 1), \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to verify that f is nondecreasing and both weakly $\overline{\mathcal{R}}_f$-min homogeneous and weakly $\overline{\mathcal{R}}_f$-max homogeneous, but it is not $\overline{\mathcal{R}}_f$-max homogeneous. For instance, $f(a \vee a, b \vee a) = f(a, 1) = 1 \neq a = f(a, b) \vee a$. Thus, f is not a polynomial function by Theorem 1.

4.2 Weak horizontal minitivity and maxitivity

Let S be a nonempty subset of L. We say that a function $f\colon L^n \to L$ is *weakly horizontally S-minitive* (resp. *weakly horizontally S-maxitive*) if (3) (resp. (4)) holds for every $\mathbf{x} \in L_n^{(0,2)}$ and every $c \in S$.

The following result provides characterizations of the n-ary polynomial functions on a chain L, given in terms of weak homogeneity and weak horizontal minitivity and maxitivity.

Theorem 9. *Let $f\colon L^n \to L$ be a function. The following conditions are equivalent:*

(i) *f is a polynomial function.*

(ii) *f is nondecreasing, weakly $\overline{\mathcal{R}}_f$-min homogeneous, and weakly $\overline{\mathcal{R}}_f$-max homogeneous.*

(iii) *f is nondecreasing, weakly $\overline{\mathcal{R}}_f$-min homogeneous, and weakly horizontally $\overline{\mathcal{R}}_f$-maxitive.*

(iv) *f is nondecreasing, weakly horizontally $\overline{\mathcal{R}}_f$-minitive, and weakly $\overline{\mathcal{R}}_f$-max homogeneous.*

(v) *f is nondecreasing, $\overline{\mathcal{R}}_f$-idempotent, weakly horizontally $\overline{\mathcal{R}}_f$-minitive, and weakly horizontally $\overline{\mathcal{R}}_f$-maxitive.*

4.3 Weak median decomposability

As we saw in Theorem 1, in the case of distributive lattices L, the n-ary polynomial functions on L are exactly those which satisfy the median decomposition formula (5). As we are going to see, in the case of chains, this condition can be relaxed by restricting the satisfaction of (5) by a function $f : L^n \to L$ to the vectors of $L_n^{(0,2)} \cup L_n^{(1,3)}$. In the latter case, we say that $f : L^n \to L$ is *weakly median decomposable*.

Lemma 10. *Let $f\colon L^n \to L$ be a nondecreasing function. If f is weakly median decomposable, then it is $\overline{\mathcal{R}}_f$-idempotent.*

Proposition 11. *Let $f\colon L^n \to L$ be a nondecreasing function. The following conditions are equivalent:*

(i) *f is weakly median decomposable.*

(ii) *f weakly $\overline{\mathcal{R}}_f$-min homogeneous and weakly $\overline{\mathcal{R}}_f$-max homogeneous.*

Remark 5. Using the binary function f given in Remark 4, we can see that Proposition 11 does not hold in the general case of bounded distributive lattices. Indeed, as observed, f is nondecreasing and both weakly $\overline{\mathcal{R}}_f$-min homogeneous and weakly $\overline{\mathcal{R}}_f$-max homogeneous, but $f(b, b) = 1 \neq b = \operatorname{median}(f(0, b), b, f(1, b))$ which shows that f is not weakly median decomposable.

From Proposition 11 and Theorem 8, we obtain the following description of polynomial functions given in terms of weak median decomposability.

Theorem 12. *A nondecreasing function $f\colon L^n \to L$ is a polynomial function if and only if it is weakly median decomposable.*

Remark 6. Note that Theorem 12 does not hold if weak median decomposability would have been defined in terms of vectors in $L_n^{(0,2)}$ only. To see this, let $L = \{0, c, 1\}$ and consider the following nondecreasing function $f : L^3 \to L$, defined by

$$f(x_1, x_2, x_3) = \begin{cases} 1, & \text{if } \operatorname{median}(x_1, x_2, x_3) = 1, \\ c, & \text{if } \operatorname{median}(x_1, x_2, x_3) \\ & = x_1 \wedge x_2 \wedge x_3 = c, \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to this that f is median decomposable for vectors in $L_n^{(0,2)}$, but it is not a polynomial function, e.g., we have $f(0, c, c) = 0$ but $f(0, 1, 1) \wedge c = c$.

4.4 Strong idempotency and componentwise range convexity

By Theorem 1, a nondecreasing function $f\colon L^n \to L$ is a polynomial function if and only if it is strongly idempotent, has a convex range, and a componentwise convex range. In the case of chains, the condition requiring a convex range becomes redundant, since it becomes a consequence of componentwise range convexity. Thus, we obtain the following characterization, which weakens condition (iii) of Theorem 1 when L is a chain.

ISBN: 978-989-95079-6-8

Theorem 13. *A function $f\colon L^n \to L$ is a polynomial function if and only if it is nondecreasing, strongly idempotent, and has a componentwise convex range.*

Remark 7. None of the conditions provided in Theorem 13 can be dropped off. For instance, let L be the real interval $[0,1]$. Clearly, the unary function $f(x) = x^2$ is nondecreasing and has a componentwise convex range, but it is not strongly idempotent. On the other hand, the function $f\colon L^2 \to L$ defined by

$$f(x_1, x_2) = \begin{cases} 1, & \text{if } x_1 = x_2 = 1, \\ 0, & \text{otherwise,} \end{cases}$$

is nondecreasing and strongly idempotent but it does not have a componentwise convex range, e.g., both $f^1_{\mathbf{1}}$ and $f^2_{\mathbf{1}}$ do not have convex ranges.

In the special case of real interval lattices, i.e., where $L = [a,b]$ for reals $a \leqslant b$, the property of having a convex range, as well as the property of having a componentwise convex range, are consequences of continuity. More precisely, for nondecreasing functions $f : [a,b]^n \to \mathbb{R}$, being continuous reduces to being continuous in each variable, and this latter property is equivalent to having a componentwise convex range. In fact, since polynomial functions are continuous, the condition of having a componentwise convex range can be replaced in Theorem 13 by continuity in each variable. Also, we can add continuity and replace $\overline{\mathcal{R}}_f$ by $\mathcal{R}_f$ in Theorems 8 and 9.

Corollary 14. *Assume that L is a bounded real interval $[a,b]$. A function $f\colon L^n \to L$ is a polynomial function if and only if it is nondecreasing, strongly idempotent, and continuous (in each variable).*

4.5 Comonotonic maxitivity and minitivity

Two vectors $\mathbf{x}$ and $\mathbf{x}'$ in L^n are said to be *comonotonic* if $\mathbf{x}, \mathbf{x}' \in L^n_\sigma$ for some permutation σ on $[n]$. A function $f\colon L^n \to L$ is said to be

- *comonotonic minitive* if, for any two comonotonic vectors $\mathbf{x}, \mathbf{x}' \in L^n$, we have
$$f(\mathbf{x} \wedge \mathbf{x}') = f(\mathbf{x}) \wedge f(\mathbf{x}').$$
- *comonotonic maxitive* if, for any two comonotonic vectors $\mathbf{x}, \mathbf{x}' \in L^n$, we have
$$f(\mathbf{x} \vee \mathbf{x}') = f(\mathbf{x}) \vee f(\mathbf{x}').$$

Note that for any $\mathbf{x} \in L^n$ and any $c \in L$, we have that $\mathbf{x}$ and $(c, \ldots, c)$ are comonotonic and that $\mathbf{x} \vee c$ and $[\mathbf{x}]^c$ are comonotonic. These facts lead to the following result.

Lemma 15. *Let S be a nonempty subset of L. If a function $f\colon L^n \to L$ is comonotonic minitive (resp. comonotonic maxitive), then it is horizontally S-minitive (resp. horizontally S-maxitive). Moreover, if f is S-idempotent, then it is S-min homogeneous (resp. S-max homogeneous).*

Let σ be a permutation on $[n]$. Clearly, every comonotonic minitive (or comonotonic maxitive) function $f\colon L^n \to L$ is nondecreasing on the standard simplex L^n_σ. The following lemma shows that this fact can be extended to the whole domain L^n.

Lemma 16. *If $f\colon L^n \to L$ is comonotonic minitive or comonotonic maxitive, then it is nondecreasing. Furthermore, every nondecreasing unary function is comonotonic minitive and comonotonic maxitive.*

We now have the following characterization of polynomial functions.

Theorem 17. *Let $f\colon L^n \to L$ be a function. The following conditions are equivalent:*

(i) *f is a polynomial function.*

(ii) *f is weakly $\overline{\mathcal{R}}_f$-min homogeneous and comonotonic maxitive.*

(iii) *f is comonotonic minitive and weakly $\overline{\mathcal{R}}_f$-max homogeneous.*

(iv) *f is $\overline{\mathcal{R}}_f$-idempotent, weakly horizontally $\overline{\mathcal{R}}_f$-minitive, and comonotonic maxitive.*

(v) *f is $\overline{\mathcal{R}}_f$-idempotent, comonotonic minitive, and weakly horizontally $\overline{\mathcal{R}}_f$-maxitive.*

(vi) *f is $\overline{\mathcal{R}}_f$-idempotent, comonotonic minitive, and comonotonic maxitive.*

Remark 8. (i) As already observed in the remark following Theorem 9, the weak horizontal $\overline{\mathcal{R}}_f$-minitivity (resp. weak horizontal $\overline{\mathcal{R}}_f$-maxitivity) can be replaced with weak horizontal L-minitivity (resp. weak horizontal L-maxitivity) in the assertions (iv)–(v) of Theorem 17.

(ii) The condition requiring $\overline{\mathcal{R}}_f$-idempotency is necessary in conditions (iv)–(vi) of Theorem 17. For instance, let L be the unit interval $[0,1]$. Clearly, the unary function $f(x) = x^2$ is nondecreasing and thus comonotonic minitive and comonotonic maxitive. By Lemma 15, it is also horizontally $\overline{\mathcal{R}}_f$-minitive and horizontally $\overline{\mathcal{R}}_f$-maxitive. However, it is not a polynomial function.

(iii) The concept of comonotonic vectors appeared as early as 1952 in Hardy et al. [12]. Comonotonic minitivity and maxitivity were introduced in the context of Sugeno integrals in de Campos et al. [7]. An interpretation of these properties was given by Ralescu and Ralescu [16] in the framework of aggregation of fuzzy subsets.

5 Discrete Sugeno integrals

In this final section, we consider a noteworthy subclass of polynomial functions, namely, that of discrete Sugeno integrals, and provide its characterizations.

A function $f\colon L^n \to L$ is said to be *idempotent* if it is L-idempotent.

Fact 18. *A polynomial function is $\{0,1\}$-idempotent if and only if it is idempotent.*

In [15, §4], $\{0,1\}$-idempotent polynomial functions are referred to as *discrete Sugeno integrals*. They coincide exactly with those functions $\mathcal{S}_\mu\colon L^n \to L$ for which there is a fuzzy measure μ such that

$$\mathcal{S}_\mu(\mathbf{x}) = \bigvee_{I \subseteq [n]} \big(\mu(I) \wedge \bigwedge_{i \in I} x_i\big).$$

ISBN: 978-989-95079-6-8

(A *fuzzy measure* μ is simply a set function $\mu\colon 2^{[n]} \to L$ satisfying $\mu(I) \leqslant \mu(I')$ whenever $I \subseteq I'$, and $\mu(\varnothing) = 0$ and $\mu([n]) = 1$.)

The following proposition shows how polynomial functions relate to Sugeno integrals; see [15, Proposition 12].

Proposition 19. *For any polynomial function $f\colon L^n \to L$ there is a fuzzy measure $\mu\colon 2^{[n]} \to L$ such that $f(\mathbf{x}) = \langle \mathcal{S}_\mu(\mathbf{x}) \rangle_f$.*

We say that a function $f\colon L^n \to L$ is *Boolean min homogeneous* (resp. *Boolean max homogeneous*) if (1) (resp. (2)) holds for every $\mathbf{x} \in \{0,1\}^n$ and every $c \in L$. Note that every weakly L-min homogeneous (resp. weakly L-max homogeneous) function is Boolean min homogeneous (resp. Boolean max homogeneous). The following result provides a variant of Theorem 8.

Theorem 20. *A function $f\colon L^n \to L$ is a discrete Sugeno integral if and only if it is nondecreasing, Boolean min homogeneous, and Boolean max homogeneous.*

Remark 9. (i) Theorem 20 as well as the characterization of the discrete Sugeno integrals obtained by combining $\{0,1\}$-idempotency with (vi) in Theorem 17 were presented in the case of real variables by [13, §4.3]; see also [14].

(ii) Even though Theorem 20 can be derived from condition (ii) of Theorem 9 by simply modifying the two homogeneity properties, to proceed similarly with conditions (iii) and (iv), it is necessary to add the conditions of $\{1\}$-idempotency and $\{0\}$-idempotency, respectively (and similarly with conditions (ii) and (iii) of Theorem 17). To see this, let L be a bounded chain with at least three elements and consider the unary functions $f(x) = x \wedge d$ and $g(x) = x \vee d$, where $d \in L \setminus \{0,1\}$. Clearly, f is L-min homogeneous and horizontally L-maxitive and g is L-max homogeneous and horizontally L-minitive. However, neither f nor g is a discrete Sugeno integral. To see that these additions are sufficient, just note that L-min homogeneity (resp. L-max homogeneity) implies $\{0\}$-idempotency (resp. $\{1\}$-idempotency).

(iii) Marichal [13, §2.2.3] showed that, when L is a chain, a nondecreasing and idempotent function $f\colon L^n \to L$ is Boolean min homogeneous (resp. Boolean max homogeneous) if and only if we have $f(\mathbf{e} \wedge c) \in \{f(\mathbf{e}), c\}$ (resp. $f(\mathbf{e} \vee c) \in \{f(\mathbf{e}), c\}$) for every $\mathbf{e} \in \{0,1\}^n$ and every $c \in L$.

References

[1] G. Beliakov, A. Pradera, and T. Calvo. *Aggregation Functions: A Guide for Practitioners*. Studies in Fuziness and Soft Computing. Springer, Berlin, 2007.

[2] P. Benvenuti, R. Mesiar, and D. Vivona. Monotone set functions-based integrals. In *Handbook of measure theory, Vol. I, II*, pages 1329–1379. North-Holland, Amsterdam, 2002.

[3] S. Burris and H. P. Sankappanavar. *A course in universal algebra*, volume 78 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1981.

[4] M. Couceiro and J.-L. Marichal. Polynomial functions over bounded distributive lattices. `http://arxiv.org/abs/0901.4888`

[5] M. Couceiro and J.-L. Marichal. Characterizations of discrete Sugeno integrals as polynomial functions over distributive lattices. `http://arxiv.org/abs/0808.2619`

[6] L. M. de Campos and M. J. Bolaños. Characterization and comparison of Sugeno and Choquet integrals. *Fuzzy Sets and Systems*, 52(1):61–67, 1992.

[7] L. M. de Campos, M. T. Lamata, and S. Moral. A unified approach to define fuzzy integrals. *Fuzzy Sets and Systems*, 39(1):75–90, 1991.

[8] J. C. Fodor and M. Roubens. Characterization of weighted maximum and some related operations. *Inform. Sci.*, 84(3-4):173–180, 1995.

[9] R. L. Goodstein. The solution of equations in a lattice. *Proc. Roy. Soc. Edinburgh Sect. A*, 67:231–242, 1965/1967.

[10] M. Grabisch, J.-L. Marichal, R. Mesiar, and E. Pap. *Aggregation functions*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, UK, 2009.

[11] M. Grabisch, T. Murofushi, and M. Sugeno, editors. *Fuzzy measures and integrals*, volume 40 of *Studies in Fuzziness and Soft Computing*. Physica-Verlag, Heidelberg, 2000.

[12] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge, at the University Press, 1952. 2d ed.

[13] J.-L. Marichal. *Aggregation operators for multicriteria decision aid*. PhD thesis, Institute of Mathematics, University of Liège, Liège, Belgium, December 1998.

[14] J.-L. Marichal. On Sugeno integral as an aggregation function. *Fuzzy Sets and Systems*, 114(3):347–365, 2000.

[15] J.-L. Marichal. Weighted lattice polynomials. *Discrete Mathematics*, 309(4):814–820, 2009.

[16] A. L. Ralescu and D. A. Ralescu. Extensions of fuzzy aggregation. *Fuzzy Sets and Systems*, 86(3):321–330, 1997.

[17] S. Rudeanu. *Lattice functions and equations*. Springer Series in Discrete Mathematics and Theoretical Computer Science. Springer-Verlag London Ltd., London, 2001.

[18] M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, Tokyo, 1974.

[19] M. Sugeno. Fuzzy measures and fuzzy integrals—a survey. In M. M. Gupta, G. N. Saridis, and B. R. Gaines, editors, *Fuzzy automata and decision processes*, pages 89–102. North-Holland, New York, 1977.

ISBN: 978-989-95079-6-8

Quasi-polynomial functions on bounded chains

Miguel Couceiro Jean-Luc Marichal

Mathematics Research Unit, University of Luxembourg
162A, avenue de la Faïencerie, L-1511 Luxembourg, Luxembourg
Email: {miguel.couceiro, jean-luc.marichal}@uni.lu

Abstract— *Two emergent properties in aggregation theory are investigated, namely horizontal maxitivity and comonotonic maxitivity (as well as their dual counterparts) which are commonly defined by means of certain functional equations. We present complete descriptions of the function classes axiomatized by each of these properties, up to weak versions of monotonicity, in the cases of horizontal maxitivity and minitivity. While studying the classes axiomatized by combinations of these properties, we introduce the concept of quasi-polynomial function which appears as a natural extension of the well-established notion of polynomial function. We present further axiomatizations for this class both in terms of functional equations and natural relaxations of homogeneity and median decomposability. As noteworthy particular cases, we investigate those subclasses of quasi-term functions and quasi-weighted maximum and minimum functions, and present characterizations accordingly.*

Keywords— Discrete Sugeno integral, quasi-polynomial function, horizontal maxitivity and minitivity, comonotonic maxitivity and minitivity, functional equation.

1 Introduction

Aggregation functions arise wherever aggregating information is important: applied and pure mathematics (probability, statistics, decision theory, functional equations), operations research, computer science, and many applied fields (economics and finance, pattern recognition and image processing, data fusion, etc.). For recent references, see Beliakov et al. [1] and Grabisch et al. [10].

A noteworthy aggregation function is the so-called discrete Sugeno integral, which was introduced by Sugeno [15, 16] and which has been widely investigated in aggregation theory, due to its many applications for instance in fuzzy set theory, decision making, and image analysis. For general background, see also the edited book [11].

A convenient way to introduce the discrete Sugeno integral is via the concept of (lattice) polynomial functions, i.e., functions which can be expressed as combinations of variables and constants using the lattice operations $\wedge$ and $\vee$. As shown by Marichal [13], the discrete Sugeno integrals are exactly those polynomial functions $f : L^n \to L$ which are idempotent, that is, satisfying $f(x, \ldots, x) = x$. Several axiomatizations of the class of discrete Sugeno integrals (as idempotent polynomial functions) have been recently given; see [4].

Of particular interest in aggregation theory, are the so-called horizontal maxitivity and comonotonic maxitivity (as well as their dual counterparts), usually expressed in terms of certain functional equations, and which we now informally describe.

Let L be a bounded chain. For every $\mathbf{x} \in L^n$ and every $c \in L$, consider the horizontal maxitive decomposition of $\mathbf{x}$ obtained by "cutting" it with c, namely $\mathbf{x} = (\mathbf{x} \wedge c) \vee [\mathbf{x}]_c$, where $[\mathbf{x}]_c$ is the n-tuple whose ith component is 0, if $x_i \leqslant c$, and x_i, otherwise. A function $f\colon L^n \to L$ is said to be *horizontally maxitive* if

$$f(\mathbf{x}) = f(\mathbf{x} \wedge c) \vee f([\mathbf{x}]_c)$$

for every $\mathbf{x} \in L^n$ and every $c \in L$.

A function $f\colon L^n \to L$ is said to be *comonotonic maxitive* if, for any two vectors $\mathbf{x}$ and $\mathbf{x}'$ in the same standard simplex of L^n, we have

$$f(\mathbf{x} \vee \mathbf{x}') = f(\mathbf{x}) \vee f(\mathbf{x}').$$

As we are going to see (Lemma 6 below), these (as well as their duals) are closely related and constitute properties shared by discrete Sugeno integrals. Still, and as it will become evident, no combination of these with their dual forms suffices to fully describe the class of Sugeno integrals. Thus, and given their emergence in aggregation theory, it is natural to ask which classes of functions are axiomatized by combinations of these properties or, in fact, by each of these properties.

In this paper, we answer this question for both the maxitive and minitive comonotonic properties, and for horizontal maxitivity and minitivity properties, up to certain weak variants of monotonicity. While looking at combinations of the latter properties, we reach a natural generalization of polynomial functions, which we call *quasi-polynomial functions* and which are best described by the following equation

$$f(x_1, \ldots, x_n) = p(\varphi(x_1), \ldots, \varphi(x_n)),$$

where p is a polynomial function and φ a nondecreasing function (see Theorem 10 below). Searching for alternative descriptions, we introduce weaker versions of well-established properties, such as homogeneity and median decomposability, to provide further axiomatizations of the class of quasi-polynomial functions, accordingly.

This paper is organized as follows. We start by recalling basic notions and terminology in lattice function theory, as well as present some known results, needed throughout this paper (Section 2). In Section 3, we study the properties of horizontal maxitivity and comonotonic maxitivity, as well as their dual forms, and determine those function classes axiomatized by each of this properties. Combinations of the latter are then considered in Section 4.1, where the notion of quasi-polynomial function is introduced. In Section 4.2, we propose weaker versions of homogeneity and median decomposability, and provide further characterizations of quasi-polynomial functions, accordingly. In Section 5, we introduce and axiomatize few noteworthy subclasses of quasi-polynomial functions, namely, those of quasi-term functions and those quasi-weighted maximum and minimum functions.

ISBN: 978-989-95079-6-8

2 Basic notions and preliminary results

In this section we recall basic terminology as well as some results needed in the current paper. For general background we refer the reader to, e.g., Burris and Sankappanavar [3] and Rudeanu [14].

2.1 General background

Throughout this paper, let L be a bounded chain with operations $\wedge$ and $\vee$, and with least and greatest elements 0 and 1, respectively. A subset S of a chain L is said to be *convex* if for every $a, b \in S$ and every $c \in L$ such that $a \leqslant c \leqslant b$, we have $c \in S$. For any subset $S \subseteq L$, we denote by $\overline{S}$ the convex hull of S, that is, the smallest convex subset of L containing S. For every $a, b \in S$ such that $a \leqslant b$, the *interval* $[a, b]$ is the set $[a, b] = \{c \in L : a \leqslant c \leqslant b\}$. For any integer $n \geqslant 1$, let $[n] = \{1, \ldots, n\}$.

For any bounded chain L, we regard the Cartesian product L^n, $n \geqslant 1$, as a distributive lattice endowed with the operations $\wedge$ and $\vee$ given by

$$\begin{aligned}(a_1, \ldots, a_n) \wedge (b_1, \ldots, b_n) &= (a_1 \wedge b_1, \ldots, a_n \wedge b_n),\\ (a_1, \ldots, a_n) \vee (b_1, \ldots, b_n) &= (a_1 \vee b_1, \ldots, a_n \vee b_n).\end{aligned}$$

The elements of L are denoted by lower case letters $a, b, c, \ldots$, and the elements of L^n, $n > 1$, by bold face letters $\mathbf{a}, \mathbf{b}, \mathbf{c}, \ldots$. We also use $\mathbf{0}$ and $\mathbf{1}$ to denote the least element and greatest element, respectively, of L^n. For $c \in L$ and $\mathbf{x} = (x_1, \ldots, x_n) \in L^n$, set

$$\mathbf{x} \wedge c = (x_1 \wedge c, \ldots, x_n \wedge c) \quad \text{and} \quad \mathbf{x} \vee c = (x_1 \vee c, \ldots, x_n \vee c).$$

The *range* of a function $f\colon L^n \to L$ is defined by $\mathcal{R}_f = \{f(\mathbf{x}) : \mathbf{x} \in L^n\}$. A function $f\colon L^n \to L$ is said to be *nondecreasing* (*in each variable*) if, for every $\mathbf{a}, \mathbf{b} \in L^n$ such that $\mathbf{a} \leqslant \mathbf{b}$, we have $f(\mathbf{a}) \leqslant f(\mathbf{b})$. The *diagonal section* of f, denoted δ_f, is defined as the unary function given by $\delta_f(x) = f(x, \ldots, x)$. Note that if f is nondecreasing, then δ_f is nondecreasing and $\overline{\mathcal{R}}_{\delta_f} = \overline{\mathcal{R}}_f = [f(\mathbf{0}), f(\mathbf{1})]$.

2.2 Polynomial functions and their representations

In this paper the so-called polynomial functions will play a fundamental role. Formally, an *n-ary polynomial function* on L is any function $f : L^n \to L$ which can be obtained by finitely many applications of the following rules:

(i) For each $i \in [n]$ and each $c \in L$, the projection $\mathbf{x} \mapsto x_i$ and the constant function $\mathbf{x} \mapsto c$ are polynomial functions from L^n to L.

(ii) If f and g are polynomial functions from L^n to L, then $f \vee g$ and $f \wedge g$ are polynomial functions from L^n to L.

Polynomial functions are also called lattice functions (Goodstein [9]), algebraic functions (Burris and Sankappanavar [3]) or weighted lattice polynomial functions (Marichal [13]). Idempotent polynomial functions (i.e., satisfying $f(c, \ldots, c) = c$ for every $c \in L$) are referred to by aggregation theorists as (*discrete*) *Sugeno integrals*, and those obtained from projections by finitely many applications of (ii) are usually referred to as (*lattice*) *term functions*. A noteworthy term function that we shall make use in the sequel is the median function, defined by

$$\operatorname{median}(x, y, z) = (x \wedge y) \vee (y \wedge z) \vee (z \wedge x).$$

As observed by Goodstein [9] (see also Rudeanu [14]), polynomial functions are exactly those functions which can be represented by formulas in disjunctive and conjunctive normal forms. In fact, each polynomial function $f\colon L^n \to L$ is uniquely determined by its restriction to $\{0, 1\}^n$. Due to their relevance in the sequel, we recall some known results concerning normal form representations of polynomial functions in the special case where L is a chain. The following result is due to Goodstein [9].

Proposition 1. *(a) Every polynomial function is completely determined by its restriction to $\{0, 1\}^n$.*

(b) A function $g\colon \{0, 1\}^n \to L$ can be extended to a polynomial function $f\colon L^n \to L$ if and only if it is nondecreasing. In this case, the extension is unique.

(c) For any $f\colon L^n \to L$, the following are equivalent:

(i) f is a polynomial function.

(ii) There exists $\alpha\colon 2^{[n]} \to L$ such that

$$f(\mathbf{x}) = \bigvee_{I \subseteq [n]} \big(\alpha(I) \wedge \bigwedge_{i \in I} x_i\big). \tag{1}$$

(iii) There exists $\beta\colon 2^{[n]} \to L$ such that

$$f(\mathbf{x}) = \bigwedge_{I \subseteq [n]} \big(\beta(I) \vee \bigvee_{i \in I} x_i\big). \tag{2}$$

The expressions given in (1) and (2) are usually referred to as the *disjunctive normal form* (DNF) representation and the *conjunctive normal form* (CNF) representation, respectively, of the polynomial function f.

Remark 1. By requiring α and β to be nonconstant functions from $2^{[n]}$ to $\{0, 1\}$ and satisfying $\alpha(\varnothing) = 0$ and $\beta(\varnothing) = 1$, respectively, we obtain the analogue of (c) of Proposition 1 for term functions.

As observed by Marichal [13], the DNF and CNF representations of polynomial functions $f\colon L^n \to L$ are not necessarily unique. For instance, we have

$$x_1 \vee (x_1 \wedge x_2) = x_1 = x_1 \wedge (x_1 \vee x_2).$$

However, from among all the possible set functions α (resp. β) defining the DNF (resp. CNF) representation of f, only one is isotone (resp. antitone), namely the function $\alpha_f\colon 2^{[n]} \to L$ (resp. $\beta_f\colon 2^{[n]} \to L$) defined by

$$\alpha_f(I) = f(\mathbf{e}_I) \qquad (\text{resp. } \beta_f(I) = f(\mathbf{e}_{[n] \setminus I})), \tag{3}$$

where $\mathbf{e}_I$ denotes the element of $\{0, 1\}^n$ whose ith component is 1 if and only if $i \in I$.

In the case when L is a chain, it was shown in [4] that the DNF and CNF representations of polynomial functions $f\colon L^n \to L$ can be refined and given in terms of standard simplices of L^n. Let σ be a permutation on $[n]$. The *standard simplex* of L^n associated with σ is the subset $L^n_\sigma \subset L^n$ defined by

$$L^n_\sigma = \{(x_1, \ldots, x_n) \in L^n : x_{\sigma(1)} \leqslant x_{\sigma(2)} \leqslant \cdots \leqslant x_{\sigma(n)}\}.$$

For each $i \in [n]$, define $S^{\uparrow}_\sigma(i) = \{\sigma(i), \ldots, \sigma(n)\}$ and $S^{\downarrow}_\sigma(i) = \{\sigma(1), \ldots, \sigma(i)\}$. As a matter of convenience, set $S^{\uparrow}_\sigma(n+1) = S^{\downarrow}_\sigma(0) = \varnothing$.

ISBN: 978-989-95079-6-8

Proposition 2. *For any function $f\colon L^n \to L$, the following conditions are equivalent:*

(i) *f is a polynomial function.*

(ii) *For any permutation σ on $[n]$ and every $\mathbf{x} \in L^n_\sigma$, we have*

$$f(\mathbf{x}) = \bigvee_{i=1}^{n+1} \big(\alpha_f(S^{\uparrow}_{\sigma}(i)) \wedge x_{\sigma(i)}\big),$$

where $x_{\sigma(n+1)} = 1$.

(iii) *For any permutation σ on $[n]$ and every $\mathbf{x} \in L^n_\sigma$, we have*

$$f(\mathbf{x}) = \bigwedge_{i=0}^{n} \big(\beta_f(S^{\downarrow}_{\sigma}(i)) \vee x_{\sigma(i)}\big),$$

where $x_{\sigma(0)} = 0$.

3 Motivating characterizations

Even though horizontal maxitivity and comonotonic maxitivity, as well as their dual counterparts, play an important role in aggregation theory (as properties shared by noteworthy classes of aggregation functions), they have not yet been described independently. In this section we investigate each of these properties and determine their corresponding function classes (up to weak versions of monotonicity, in the cases of horizontal maxitivity and minitivity).

3.1 Horizontal maxitivity and minitivity

Recall that a function $f\colon L^n \to L$ is said to be

- *horizontally maxitive* if, for every $\mathbf{x} \in L^n$ and every $c \in L$, we have

$$f(\mathbf{x}) = f(\mathbf{x} \wedge c) \vee f([\mathbf{x}]_c),$$

where $[\mathbf{x}]_c$ is the n-tuple whose ith component is 0, if $x_i \leqslant c$, and x_i, otherwise.

- *horizontally minitive* if, for every $\mathbf{x} \in L^n$ and every $c \in L$, we have

$$f(\mathbf{x}) = f(\mathbf{x} \vee c) \wedge f([\mathbf{x}]^c),$$

where $[\mathbf{x}]^c$ is the n-tuple whose ith component is 1, if $x_i \geqslant c$, and x_i, otherwise.

Let us consider the following weak forms of nondecreasing monotonicity:

$(\mathbf{P_1})$ $f(\mathbf{e} \wedge c) \leqslant f(\mathbf{e}' \wedge c)$ for every $\mathbf{e}, \mathbf{e}' \in \{0,1\}^n$ such that $\mathbf{e} \leqslant \mathbf{e}'$ and every $c \in L$.

$(\mathbf{D_1})$ $f(\mathbf{e} \vee c) \leqslant f(\mathbf{e}' \vee c)$ for every $\mathbf{e}, \mathbf{e}' \in \{0,1\}^n$ such that $\mathbf{e} \leqslant \mathbf{e}'$ and every $c \in L$.

$(\mathbf{P_2})$ $f(\mathbf{e} \wedge c) \leqslant f(\mathbf{e} \wedge c')$ for every $\mathbf{e} \in \{0,1\}^n$ and every $c, c' \in L$ such that $c \leqslant c'$.

$(\mathbf{D_2})$ $f(\mathbf{e} \vee c) \leqslant f(\mathbf{e} \vee c')$ for every $\mathbf{e} \in \{0,1\}^n$ and every $c, c' \in L$ such that $c \leqslant c'$.

Theorem 3. *A function $f\colon L^n \to L$ is horizontally maxitive and satisfies $\mathbf{P_1}$ if and only if there exists $g\colon L^n \to L$ satisfying $\mathbf{P_2}$ such that*

$$f(\mathbf{x}) = \bigvee_{I \subseteq [n]} g\Big(\mathbf{e}_I \wedge \bigwedge_{i \in I} x_i\Big).$$

In this case, we can choose $g = f$.

Similarly, we obtain the following dual characterization:

Theorem 4. *A function $f\colon L^n \to L$ is horizontally minitive and satisfies $\mathbf{D_1}$ if and only if there exists $g\colon L^n \to L$ satisfying $\mathbf{D_2}$ such that*

$$f(\mathbf{x}) = \bigwedge_{I \subseteq [n]} g\Big(\mathbf{e}_{[n]\setminus I} \vee \bigvee_{i \in I} x_i\Big).$$

In this case, we can choose $g = f$.

From Theorems 3 and 4 we have the following corollary.

Corollary 5. *A function $f\colon L^n \to L$ is horizontally maxitive (resp. horizontally minitive) and satisfies $\mathbf{P_1}$ (resp. $\mathbf{D_1}$) if and only if there are unary nondecreasing functions $\varphi_I\colon L \to L$, for $I \subseteq [n]$, such that*

$$\begin{aligned} f(\mathbf{x}) &= \bigvee_{I \subseteq [n]} \big(\alpha_f(I) \wedge \bigwedge_{i \in I} \varphi_I(x_i)\big) \\ \text{(resp. } f(\mathbf{x}) &= \bigwedge_{I \subseteq [n]} \big(\beta_f(I) \vee \bigvee_{i \in I} \varphi_I(x_i)\big)\text{)}, \end{aligned}$$

where the set function α_f (resp. β_f) is defined in (3). In this case, we can choose $\varphi_I(x) = f(\mathbf{e}_I \wedge x)$ (resp. $\varphi_I(x) = f(\mathbf{e}_{[n]\setminus I} \vee x)$) for every $I \subseteq [n]$.

Remark 2. (i) Theorem 3 (resp. Theorem 4) provides the description of those horizontally maxitive (resp. horizontally minitive) functions which are nondecreasing.

(ii) Every Boolean function $f\colon \{0,1\}^n \to \{0,1\}$ satisfying $f(\mathbf{0}) \leqslant f(\mathbf{x})$ (resp. $f(\mathbf{x}) \leqslant f(\mathbf{1})$) is horizontally maxitive (resp. horizontally minitive). Moreover, not all such functions are nondecreasing, thus showing that condition $\mathbf{P_1}$ (resp. $\mathbf{D_1}$) is necessary in Theorem 3 (resp. Theorem 4).

(iii) As shown in [4], polynomial functions $f\colon L^n \to L$ are exactly those $\overline{\mathcal{R}}_f$-idempotent (i.e., satisfying $f(c, \ldots, c) = c$ for every $c \in \overline{\mathcal{R}}_f$) which are nondecreasing, horizontally maxitive, and horizontally minitive.

(iv) The concept of horizontal maxitivity was introduced, in the case when L is the real interval $[0,1]$, by Benvenuti et al. [2] as a general property of the Sugeno integral.

3.2 Comonotonic maxitivity and minitivity

Two vectors $\mathbf{x}, \mathbf{x}' \in L^n$ are said to be *comonotonic* if there exists a permutation σ on $[n]$ such that $\mathbf{x}, \mathbf{x}' \in L^n_\sigma$. A function $f\colon L^n \to L$ is said to be

- *comonotonic maxitive* if, for any two comonotonic vectors $\mathbf{x}, \mathbf{x}' \in L^n$, we have

$$f(\mathbf{x} \vee \mathbf{x}') = f(\mathbf{x}) \vee f(\mathbf{x}').$$

ISBN: 978-989-95079-6-8

- *comonotonic minitive* if, for any two comonotonic vectors $\mathbf{x}, \mathbf{x}' \in L^n$, we have

$$f(\mathbf{x} \wedge \mathbf{x}') = f(\mathbf{x}) \wedge f(\mathbf{x}').$$

Note that for any $\mathbf{x} \in L^n$ and any $c \in L$, the vectors $\mathbf{x} \vee c$ and $[\mathbf{x}]^c$ are comonotonic. As a consequence, if a function $f\colon L^n \to L$ is comonotonic maxitive (resp. comonotonic minitive), then it is horizontally maxitive (resp. horizontally minitive). It was also observed in [4] that if f is comonotonic maxitive or comonotonic minitive, then it is nondecreasing. Moreover, we obtain the following result.

Lemma 6. *A function $f\colon L^n \to L$ is comonotonic maxitive (resp. comonotonic minitive) if and only if it is horizontally maxitive (resp. horizontally minitive) and satisfies $\mathbf{P_1}$ (resp. $\mathbf{D_1}$).*

Combining Theorems 3 and 4 with Lemma 6, we immediately obtain the descriptions of the classes of comonotonic maxitive and comonotonic minitive functions.

Theorem 7. *A function $f\colon L^n \to L$ is comonotonic maxitive if and only if there exists $g\colon L^n \to L$ satisfying $\mathbf{P_2}$ such that*

$$f(\mathbf{x}) = \bigvee_{I \subseteq [n]} g\Big(\mathbf{e}_I \wedge \bigwedge_{i \in I} x_i\Big).$$

In this case, we can choose $g = f$.

Theorem 8. *A function $f\colon L^n \to L$ is comonotonic minitive if and only if there exists $g\colon L^n \to L$ satisfying $\mathbf{D_2}$ such that*

$$f(\mathbf{x}) = \bigwedge_{I \subseteq [n]} g\Big(\mathbf{e}_{[n]\setminus I} \vee \bigvee_{i \in I} x_i\Big).$$

In this case, we can choose $g = f$.

As before, we have the following corollary.

Corollary 9. *A function $f\colon L^n \to L$ is comonotonic maxitive (resp. comonotonic minitive) if and only if there are unary nondecreasing functions $\varphi_I\colon L \to L$, for $I \subseteq [n]$, such that*

$$\begin{aligned} f(\mathbf{x}) &= \bigvee_{I \subseteq [n]} \big(\alpha_f(I) \wedge \bigwedge_{i \in I} \varphi_I(x_i)\big) \\ \text{(resp. } f(\mathbf{x}) &= \bigwedge_{I \subseteq [n]} \big(\beta_f(I) \vee \bigvee_{i \in I} \varphi_I(x_i)\big)\text{)}, \end{aligned}$$

where the set function α_f (resp. β_f) is defined in (3). In this case, we can choose $\varphi_I(x) = f(\mathbf{e}_I \wedge x)$ (resp. $\varphi_I(x) = f(\mathbf{e}_{[n]\setminus I} \vee x)$) for every $I \subseteq [n]$.

Remark 3. (i) An alternative description of comonotonic maxitive (resp. comonotonic minitive) functions was obtained in Grabisch et al. [10, §2.5] in the case when L is a real interval.

(ii) It was shown in [4] that polynomial functions $f\colon L^n \to L$ are exactly those $\overline{\mathcal{R}}_f$-idempotent functions which are comonotonic maxitive and comonotonic minitive.

(ii) Comonotonic minitivity and maxitivity were introduced in the context of Sugeno integrals in de Campos et al. [5].

4 Quasi-polynomial functions

Motivated by the results of Section 3 concerning horizontal maxitivity and comonotonic maxitivity, as well as their dual counterparts, we now study combinations of these properties. This will lead to a relaxation of the notion of polynomial function, which we will refer to as *quasi-polynomial function*. Accordingly, we introduce weaker variants of well-established properties, such as homogeneity and median decomposability, which are then used to provide further axiomatizations of the class of quasi-polynomial functions.

4.1 Motivation and definition

We start by looking at combinations of those properties studied in Section 3. These are considered in the following result.

Theorem 10. *Let $f\colon L^n \to L$ be a function. The following assertions are equivalent:*

(i) f is horizontally maxitive, horizontally minitive, and satisfies $\mathbf{P_1}$ or $\mathbf{D_1}$.

(ii) f is comonotonic maxitive and comonotonic minitive.

(iii) f is horizontally maxitive and comonotonic minitive.

(iv) f is comonotonic maxitive and horizontally minitive.

(v) There exist a polynomial function $p\colon L^n \to L$ and a nondecreasing function $\varphi\colon L \to L$ such that

$$f(x_1, \ldots, x_n) = p(\varphi(x_1), \ldots, \varphi(x_n)).$$

If these conditions hold then we can choose for p the unique polynomial function p_f extending $f|_{\{0,1\}^n}$ and for φ the diagonal section δ_f of f.

Theorem 10 motivates the following definition.

Definition 11. We say that a function $f\colon L^n \to L$ is a *quasi-polynomial function* (resp. a *discrete quasi-Sugeno integral*, a *quasi-term function*) if there exist a polynomial function (resp. a discrete Sugeno integral, a term function) $p\colon L^n \to L$ and a nondecreasing function $\varphi\colon L \to L$ such that $f = p \circ \varphi$, that is,

$$f(x_1, \ldots, x_n) = p(\varphi(x_1), \ldots, \varphi(x_n)). \tag{4}$$

Remark 4. (i) Note that each quasi-polynomial function $f\colon L^n \to L$ can be represented as a combination of constants and a nondecreasing unary function φ (applied to the projections $\mathbf{x} \mapsto x_i$) using the lattice operations $\vee$ and $\wedge$.

(ii) In the setting of decision-making under uncertainty, the nondecreasing function φ in (4) can be thought of as a *utility function* and the corresponding quasi-polynomial function as a (qualitative) *global preference functional*; see for instance Dubois et al. [6].

Note that the functions p and φ in (4) are not necessarily unique. For instance, if f is a constant $c \in L$, then we could choose $p \equiv c$ and φ arbitrarily, or p idempotent and $\varphi \equiv c$. We now describe all possible choices for p and φ. For any integers $m, n \geqslant 1$, any vector $\mathbf{x} \in L^m$, and any function $f\colon L^n \to L$, we define $\langle \mathbf{x} \rangle_f \in L^m$ as the m-tuple

$$\langle \mathbf{x} \rangle_f = \operatorname{median}(f(\mathbf{0}), \mathbf{x}, f(\mathbf{1})),$$

where the right-hand side median is taken componentwise.

ISBN: 978-989-95079-6-8

Proposition 12. *Let $f\colon L^n \to L$ be a quasi-polynomial function and let $p_f : L^n \to L$ be the unique polynomial function extending $f|_{\{0,1\}^n}$. We have*

$$\{(p,\varphi)\colon f = p \circ \varphi\} = \{(p,\varphi)\colon p_f = \langle p\rangle_f \text{ and } \delta_f = \langle\varphi\rangle_p\},$$

where p and φ stand for polynomial and unary nondecreasing functions, respectively. In particular, we have $f = p_f \circ \delta_f$.

It was shown in Marichal [13] that every polynomial function $p\colon L^n \to L$ can be represented as $\langle q\rangle_p$ for some discrete Sugeno integral $q\colon L^n \to L$. Combining this with Proposition 12, we obtain the next result.

Corollary 13. *The class of quasi-polynomial functions is exactly the class of discrete quasi-Sugeno integrals.*

4.2 Further axiomatizations

We now recall some properties of polynomial functions, namely homogeneity and median decomposability, and we propose weaker variants of these to provide alternative axiomatizations of the class of quasi-polynomial functions.

4.2.1 Quasi-homogeneity

Let S be a subset of L. A function $f\colon L^n \to L$ is said to be *S-max homogeneous* (resp. *S-min homogeneous*) if for every $\mathbf{x} \in L^n$ and every $c \in S$, we have

$$f(\mathbf{x} \vee c) = f(\mathbf{x}) \vee c \qquad (\text{resp. } f(\mathbf{x} \wedge c) = f(\mathbf{x}) \wedge c).$$

Although polynomial functions $p\colon L^n \to L$ share both of these properties for any $S \subseteq \overline{\mathcal{R}}_p$, this is not the case for quasi-polynomial functions. For instance, let $f_1, f_2\colon [0,1] \to [0,1]$ be respectively given by $f_1(x) = x^2$ and $f_2(x) = \sqrt{x}$. Clearly, f_1 and f_2 are quasi-polynomial functions but, e.g., for $x = c$, we have

$$f_1(x \vee c) < f_1(x) \vee c \quad \text{and} \quad f_2(x \wedge c) > f_2(x) \wedge c.$$

This example motivates the following relaxations. We say that a function $f\colon L^n \to L$ is *quasi-max homogeneous* (resp. *quasi-min homogeneous*) if for every $\mathbf{x} \in L^n$ and $c \in L$, we have

$$f(\mathbf{x} \vee c) = f(\mathbf{x}) \vee \delta_f(c) \qquad (\text{resp. } f(\mathbf{x} \wedge c) = f(\mathbf{x}) \wedge \delta_f(c)).$$

Observe that if f is $\overline{\mathcal{R}}_f$-idempotent (i.e., satisfying $f(c,\ldots,c) = c$ for every $c \in \overline{\mathcal{R}}_f$), then $\overline{\mathcal{R}}_f$-min homogeneity (resp. $\overline{\mathcal{R}}_f$-max homogeneity) is equivalent to quasi-min homogeneity (resp. quasi-max homogeneity).

Lemma 14. *Let $f\colon L^n \to L$ be nondecreasing and quasi-min homogeneous (resp. quasi-max homogeneous). Then f is quasi-max homogeneous (resp. quasi-min homogeneous) if and only if it is horizontally maxitive (resp. horizontally minitive).*

Combining Theorem 10 and Lemma 14, we obtain a characterization of quasi-polynomial functions in terms of quasi-min homogeneity and quasi-max homogeneity.

Theorem 15. *A function $f\colon L^n \to L$ is a quasi-polynomial function if and only if it is nondecreasing, quasi-max homogeneous, and quasi-min homogeneous.*

4.2.2 Quasi-median decomposability

A function $f\colon L^n \to L$ is said to be *median decomposable* [13] if, for every $\mathbf{x} \in L^n$ and every $k \in [n]$, we have

$$f(\mathbf{x}) = \operatorname{median}\big(f(\mathbf{x}_k^0), x_k, f(\mathbf{x}_k^1)\big),$$

where $\mathbf{x}_k^c = (x_1,\ldots,x_{k-1},c,x_{k+1},\ldots,x_n)$ for any $c \in L$. As Marichal [13] showed, the class of polynomial functions are exactly those functions which are median decomposable.

In complete analogy with the previous subsection we propose the following weaker variant of median decomposability. We say that a function $f\colon L^n \to L$ is *quasi-median decomposable* if, for every $\mathbf{x} \in L^n$ and every $k \in [n]$, we have

$$f(\mathbf{x}) = \operatorname{median}\big(f(\mathbf{x}_k^0), \delta_f(x_k), f(\mathbf{x}_k^1)\big).$$

Note that every nondecreasing unary function is quasi-median decomposable.

Observe that $\vee$ and $\wedge$, as well as any nondecreasing function $\varphi\colon L \to L$, are quasi-median decomposable. Also, it is easy to see that any combination of constants and a nondecreasing unary function φ using $\vee$ and $\wedge$ is quasi-median decomposable and hence, by Remark 4 (i), every quasi-polynomial function is quasi-median decomposable. Our following result asserts that quasi-median decomposable functions $f\colon L^n \to L$ with a nondecreasing diagonal section δ_f are exactly the quasi-polynomial functions.

Theorem 16. *A function $f\colon L^n \to L$ is a quasi-polynomial function if and only if δ_f is nondecreasing and f is quasi-median decomposable.*

5 Some special classes of quasi-polynomial functions

In this final section we consider few noteworthy subclasses of quasi-polynomial functions, namely those of quasi-term functions and quasi-weighted maximum and minimum functions, and provide characterizations accordingly.

5.1 Quasi-term functions

We say that a function $f\colon L^n \to L$ is

- *conservative* if, for every $\mathbf{x} \in L^n$, we have $f(\mathbf{x}) \in \{x_1,\ldots,x_n\}$.
- *quasi-conservative* if, for every $\mathbf{x} \in L^n$, we have $f(\mathbf{x}) \in \{\delta_f(x_1),\ldots,\delta_f(x_n)\}$.

Note that, if f is idempotent, then it is quasi-conservative if and only if it is conservative.

Theorem 17. *A quasi-polynomial function $f\colon L^n \to L$ is a quasi-term function if and only if it is quasi-conservative.*

5.2 Quasi-weighted maximum and minimum functions

A function $f\colon L^n \to L$ is said to be a *weighted maximum function* if there are $v_0, v_1, \ldots, v_n \in L$ such that

$$f(\mathbf{x}) = v_0 \vee \bigvee_{i\in[n]} (v_i \wedge x_i). \tag{5}$$

Similarly, $f\colon L^n \to L$ is said to be a *weighted minimum function* if there are $w_0, w_1, \ldots, w_n \in L$ such that

$$f(\mathbf{x}) = w_0 \wedge \bigwedge_{i\in[n]} (w_i \vee x_i). \tag{6}$$

ISBN: 978-989-95079-6-8

We say that a function $f\colon L^n \to L$ is a *quasi-weighted maximum function* (resp. a *quasi-weighted minimum function*) if there exist a weighted maximum function (resp. a weighted minimum function) $p\colon L^n \to L$ and a nondecreasing function $\varphi\colon L \to L$ such that $f = p \circ \varphi$.

To present a axiomatization of each of these classes, we need to recall some terminology. We say that a function $f\colon L^n \to L$ is

- *maxitive* if, for every $\mathbf{x}, \mathbf{x}' \in L^n$, we have $f(\mathbf{x} \vee \mathbf{x}') = f(\mathbf{x}) \vee f(\mathbf{x}')$.
- *minitive* if, for every $\mathbf{x}, \mathbf{x}' \in L^n$, we have $f(\mathbf{x} \wedge \mathbf{x}') = f(\mathbf{x}) \wedge f(\mathbf{x}')$.

We first recall the descriptions of maxitive and minitive functions; see Dubois and Prade [8] and Marichal [12].

Proposition 18. *A function $f\colon L^n \to L$ is maxitive (resp. minitive) if and only if there are nondecreasing unary functions $f_i\colon L \to L$ ($i \in [n]$) such that, for every $\mathbf{x} \in L^n$,*

$$f(\mathbf{x}) = \bigvee_{i\in[n]} f_i(x_i) \qquad (\textit{resp. } f(\mathbf{x}) = \bigwedge_{i\in[n]} f_i(x_i)).$$

Theorem 19. *Let $f\colon L^n \to L$ be a quasi-polynomial function. Then f is a quasi-weighted maximum function (resp. quasi-weighted minimum function) if and only if it is maxitive (resp. minitive).*

Remark 5. (i) Idempotent weighted maximum functions $f\colon L^n \to L$ are those functions (5) for which $v_0 = 0$ and $\vee_{i\in[n]} v_i = 1$. Dually, idempotent weighted minimum functions $f\colon L^n \to L$ are those functions (6) for which $w_0 = 1$ and $\wedge_{i\in[n]} w_i = 0$. These functions were introduced on real intervals by Dubois and Prade [7] in fuzzy set theory.

(ii) As observed in Proposition 12, the underlying weighted maximum function (resp. weighted minimum function) defining a given quasi-weighted maximum function (resp. quasi-weighted minimum function) can be chosen to be idempotent.

References

[1] G. Beliakov, A. Pradera, and T. Calvo. *Aggregation Functions: A Guide for Practitioners*. Studies in Fuziness and Soft Computing. Springer, Berlin, 2007.

[2] P. Benvenuti, R. Mesiar, and D. Vivona. Monotone set functions-based integrals. In *Handbook of measure theory, Vol. I, II*, pages 1329–1379. North-Holland, Amsterdam, 2002.

[3] S. Burris and H. P. Sankappanavar. *A course in universal algebra*, volume 78 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1981.

[4] M. Couceiro and J.-L. Marichal. Representations and characterizations of polynomial functions on chains. `http://arxiv.org/abs/0811.0309`

[5] L. M. de Campos, M. T. Lamata, and S. Moral. A unified approach to define fuzzy integrals. *Fuzzy Sets and Systems*, 39(1):75–90, 1991.

[6] D. Dubois, J.-L. Marichal, H. Prade, M. Roubens, and R. Sabbadin. The use of the discrete Sugeno integral in decision-making: a survey. *Internat. J. Uncertain. Fuzziness Knowledge-Based Systems*, 9(5):539–561, 2001.

[7] D. Dubois and H. Prade. Weighted minimum and maximum operations in fuzzy set theory. *Inform. Sci.*, 39(2):205–210, 1986.

[8] D. Dubois and H. Prade. Aggregation of possibility measures. In *Multiperson decision making models using fuzzy sets and possibility theory*, volume 18 of *Theory Decis. Lib. Ser. B Math. Statist. Methods*, pages 55–63. Kluwer Acad. Publ., Dordrecht, 1990.

[9] R. L. Goodstein. The solution of equations in a lattice. *Proc. Roy. Soc. Edinburgh Sect. A*, 67:231–242, 1965/1967.

[10] M. Grabisch, J.-L. Marichal, R. Mesiar, and E. Pap. *Aggregation functions*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, UK, to appear in 2009.

[11] M. Grabisch, T. Murofushi, and M. Sugeno, editors. *Fuzzy measures and integrals*, volume 40 of *Studies in Fuzziness and Soft Computing*. Physica-Verlag, Heidelberg, 2000. Theory and applications.

[12] J.-L. Marichal. On Sugeno integral as an aggregation function. *Fuzzy Sets and Systems*, 114(3):347–365, 2000.

[13] J.-L. Marichal. Weighted lattice polynomials. *Discrete Mathematics*, 309(4):814–820, 2009.

[14] S. Rudeanu. *Lattice functions and equations*. Springer Series in Discrete Mathematics and Theoretical Computer Science. Springer-Verlag London Ltd., London, 2001.

[15] M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, Tokyo, 1974.

[16] M. Sugeno. Fuzzy measures and fuzzy integrals—a survey. In M. M. Gupta, G. N. Saridis, and B. R. Gaines, editors, *Fuzzy automata and decision processes*, pages 89–102. North-Holland, New York, 1977.

ISBN: 978-989-95079-6-8

Novel Methods for the Design of General Type-2 fuzzy Sets based on Device Characteristics and Linguistic Labels Surveys

Christian Wagner Hani Hagras

The Computational Intelligence Centre, School of Computer Science and Electronic Engineering, University of Essex
Colchester, United Kingdom
Email: chwagn@essex.ac.uk, hani@essex.ac.uk

Abstract—*Fuzzy Logic Systems are widely recognized to be successful at modelling uncertainty in a large variety of applications. While recently interval type-2 fuzzy logic has been credited for the ability to better deal with large amounts of uncertainty, general type-2 fuzzy logic has been a steadily growing research area. All fuzzy logic systems require the accurate specification of the membership functions' (MFs) parameters. While some work for automatic or manual design of these parameters has been proposed for type-1 and interval type-2 fuzzy logic, the problem has not yet been widely addressed for general type-2 fuzzy logic. In this paper we propose two methods which allow the automatic design of general type-2 MFs using either data gathered through a survey on the linguistic variables required or, in the case of physical devices (e.g. sensors, actuators), using data directly gathered from the specific devices. As such, the proposed methods allow for the creation of general type-2 MFs which directly model the uncertainty incorporated in the respective applications. Additionally, we demonstrate how interval type-2, type-1 and the recently introduced zSlices based general type-2 MFs can be extracted from the automatically designed general type-2 MFs. We also present a recursive algorithm that computes the convex approximation of generated fuzzy sets.*

Keywords— Type-2 fuzzy systems, general type-2 fuzzy sets, specification of fuzzy sets.

1 Introduction

Fuzzy Logic is widely applied because of its ability to model the uncertainties which are inherent to most real world applications and environments. The high levels of uncertainty encountered by such fuzzy logic systems can generally be attributed to one or more of the categories below:

- Uncertainties in inputs to FLCs which translate to uncertainties in the antecedent Membership Functions (MFs) as for example sensor measurements are typically noisy and affected by the conditions of observation (i.e. their characteristics are changed by the environmental conditions such as wind, sunshine, humidity, rain, etc.).
- Uncertainties in control outputs which translate to uncertainties in the consequent MFs of FLCs. Such uncertainties can result from the change of the actuators characteristics which can be due to wear, tear, environmental changes, etc.
- Linguistic uncertainties associated with the meaning of words (that are used as antecedents or consequents in FLCs or used as part of Computing With Words (CWW) applications) which can be uncertain as "words mean different things to different people" [1]. For example the perception of the linguistic label "warm" varies for different people while also being heavily subject to context, e.g. "warm" in the Caribbean will usually be associated with a different temperature than "warm" at the North Pole etc.

While the capability of modelling uncertainty provided by type-1 fuzzy sets has been intensively researched and shown to provide good results, in recent years interval type-2 fuzzy sets have been found to provide an even higher potential at modelling large amounts of uncertainty [1],[2],[3],[4]. Additionally, a variety of significant advances has been made in the area of general type-2 fuzzy logic which is hoped to provide the framework for an accurate model of uncertainty.

One of the main challenges during the design of any fuzzy based system (general type-2, interval type-2 or type-1) is the specification of the fuzzy Membership Functions (MFs). The choice of type of membership function (such as gaussian, triangular, etc.) as well as the choice of their specific parameters strongly affects the performance of the fuzzy system. A variety of methods to alleviate this problem have been researched for mainly type-1 and interval type-2 FLCs. Such methods are generally based on the use of expert knowledge, evolutionary techniques (such as genetic algorithms) and neural networks [5], [6]. However, there is still much work needed to standardize and simplify the selection of appropriate MFs.

Recent developments in general type-2 fuzzy theory [7], [8], [9], [10], [11] have made general type-2 applications a realistic option for the future. Hence, we have developed novel techniques to determine the appropriate parameters for general type-2 MFs which due to their complex nature are highly difficult to specify. The proposed techniques are based on directly using existing knowledge about the uncertainty contained within the variables represented by the MFs. Additionally, the proposed techniques have the advantage that they can be used for the generation of all currently common types of fuzzy sets: type-1, interval type-2 and general type-2.

Section 2 describes the design of fuzzy MFs modelling linguistic concepts using surveys while Section 3 details the design of fuzzy MFs based on physical device or variable characteristics. Both sections present real world examples to demonstrate the techniques. Section 4 describes how to extract zSlice based general type-2, interval type-2 and type-1 MFs from the general type-2 MFs generated in Section 2 and 3. In Section 5 we present the recursive convexity algorithm employed throughout the paper and give a detailed example of its application. Finally, in Section 6 we present

ISBN: 978-989-95079-6-8

the conclusions and address the work planned for the near future.

2 Design of general type-2 fuzzy sets based on the perception of linguistic labels determined through surveys.

2.1 The Human Perception of Linguistic Labels.

Fuzzy sets are frequently used to specify linguistic variables such as *temperature* or *height* used by people on an everyday basis. As such, the variable *temperature* can for example be modelled using three linguistic labels, *cold*, *moderate* and *warm* which in turn are represented using fuzzy MFs. While type-1 sets can be used to model such labels, it is clear that as type-1 MFs associate each input with a crisp membership value, hence a given type-1 MF will need to be constantly re-tuned for different perceptions of the same linguistic label. For example, the perception of the linguistic label *warm* can easily be seen to depend on factors such as the country of origin of a person (someone from Alaska will consider warm a different temperature than someone from the Caribbean), the age, sex, current season, etc.

Interval type-2 fuzzy sets can alleviate the above need for the continuous re-tuning of individual MFs as the incorporated Footprint of Uncertainty (FOU) allows for the model to include the full range of values associated with a specific linguistic label by different people in different contexts. Interval type-2 fuzzy sets nevertheless have the significant shortcoming that the uncertainty is spread evenly across the FOU which does not allow for variations of the uncertainty within this FOU to be modelled. For example, considering the example of *warm* temperature, if two people associate warm with 25°C and one person associates it with 28°C, a FOU incorporating 25°C and 28°C will not reflect that the majority of the people (here two) associate 25°C with the linguistic label *warm* - we have effectively lost this information.

This additional information can be modelled using general type-2 fuzzy sets by employing the third dimension associated with the FOU to encode the amount of uncertainty associated with each point within the FOU. In terms of our example, the FOU would be associated in its third dimension with more uncertainty towards 28°C and less uncertainty towards 25°C.

In the next few subsections we will describe how a survey on linguistic labels can be used to gather information on a specific linguistic variable using the example of *Indoor Temperature*. While this example is chosen to illustrate and simplify the process, the mechanism can be applied for any linguistic variable which is modelled by a series of consecutive linguistic labels. We will show how the information can be used to generate fuzzy MFs which model the linguistic labels and their associated uncertainty accurately using general type-2 fuzzy sets. In addition, we will present the mechanisms used to extract appropriate interval type-2 and type-1 MFs as well as zSlices based general type-2 MFs which were introduced in [11].

2.2 The survey on the linguistic variable "Indoor Temperature".

As part of our work on the generation of MFs from surveys we have conducted a survey investigating the perception of the linguistic variable *Indoor Temperature*. We have presented the participants with a questionnaire asking them to identify the position of the three linguistic labels *Cold*, *Moderate* and *Warm* in the context of an indoor environment on a temperature scale ranging from 0°C to 40°C. Further details on the survey and its results are given in Table 1. It should be noted that participants could supply additional information during the survey, specifically their age, sex and country of origin. While maintaining anonymity, we consider this information highly useful as it can be expected that factors such as the country of origin significantly affect people's perception for example of temperature.

Table 1: Details on the linguistic survey on the perception of *Indoor Temperature*. (* based on information supplied)

Number of participants:	62
Age range of participants:	18-58 years*
Mean age of participants:	30.145 years*
Number of different countries of origin:	22 countries*
Location of survey:	Colchester, UK
Date of survey:	January 2009
Minimum values with *cold, moderate, warm*:	0°C, 10°C, 16°C
Maximum values with *cold, moderate, warm*:	24°C, 27°C, 35°C
Mean values for *cold, medium, warm*:	12.062°C, 19.664°C, 25.664°C
Standard deviations for *cold, medium, warm*:	4.637, 3.552, 3.926

2.3 From a Survey to a General Type-2 Fuzzy Set model of the Linguistic Variable

In order to utilise the data provided through the survey to generate a general type-2 fuzzy set model of the linguistic variable which includes all the uncertainty information gathered in the data, we follow the following steps:

(1) Map all the values associated with a specific linguistic label to the frequency with which the values occur in the data. This step can be likened to creating a continuous histogram for the data where the xAxis represents the values (in our case, the temperature) and the yAxis represents the number of occurrences. Fig. 1 shows this mapping for our data using interpolation between the data points.

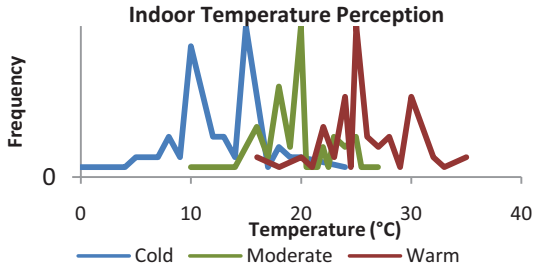


Figure 1: Mapping of survey data for the three linguistic labels *cold, moderate and warm*, showing the large amounts of uncertainty contained within the perceptions of the linguistic labels (e.g. large overlap).

(2) The data mapped in Step (1) and illustrated in Fig. 2 illustrates the significant variation in the perception of the individual linguistic labels as well as their strongly differing interrelation for different people. In order to

ISBN: 978-989-95079-6-8

utilize the information to define the third dimension of general type-2 fuzzy sets we normalise the frequency so it correlates with the uncertainty modelled in the third dimension of a fuzzy set, ie. the more frequent a specific sample was identified in the survey, the less uncertainty is associated with it.

(3) As convexity is a general requirement in several general type-2 fuzzy logic operations such as the join and meet computed on the membership grade presented in [12], we convert the normalised plots created in Step 2 to their convex approximation using a recursive algorithm which we describe in Section 5 in more detail. The original, normalised data as well as its convex approximations using our recursive algorithm and basic triangular approximation are shown for all three linguistic labels in Figs. 2a, 2b and 2c.

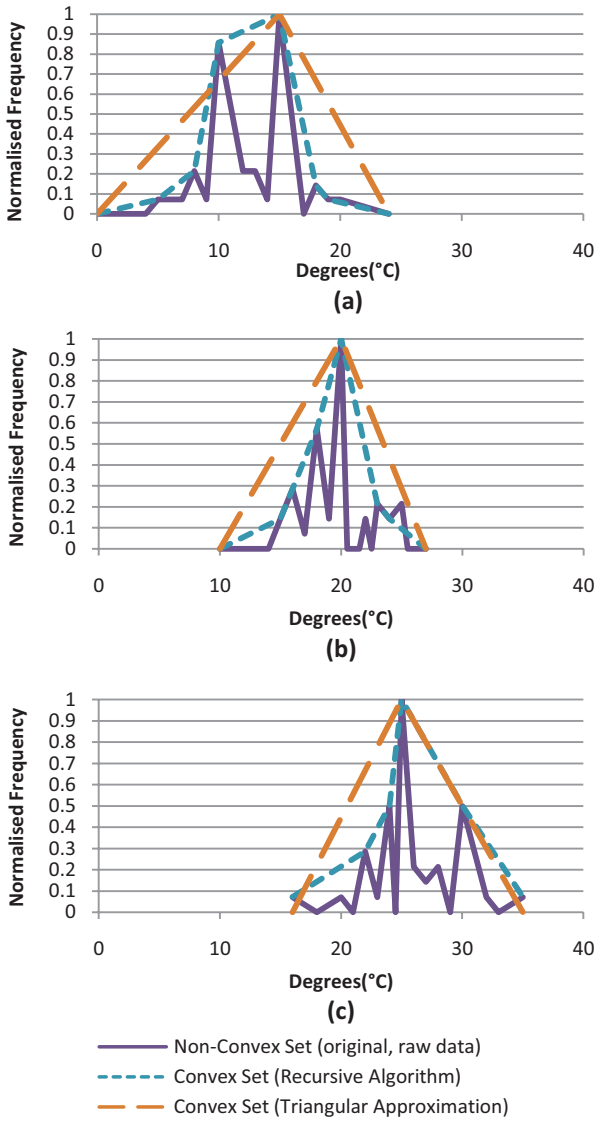


Figure 2: Representations of the linguistic labels, (a) H(cold), (b) H(moderate), (c) H(warm).

Fig. 2a, Fig. 2b and Fig. 2c clearly show that a more accurate convex approximation of the original sets is achieved when employing the proposed recursive algorithm (explained in Section 5) when compared to a simple triangular approximation. We refer to the "histogram based" representations of the linguistic labels cold, moderate and warm as H(cold), H(moderate) and H(warm).

(4) In order to create a fuzzy model of the linguistic variable *Indoor Temperature* we proceed one linguistic label at a time:

a. For *cold*, we associate H(cold) with the level 1 on the yAxis and H(moderate) with level 0 on the yAxis. We e*mploy* interpolation to recreate the general type-2 set for *cold* as shown in Fig. 3a, Fig. 3b and Fig 3c. It s*houl*d be noted that the positioning of H(cold) and H(modera*te)* at *thi*s step is not arbitrary but follows logically from the linguistic labels, i.e. H(cold) represe*nts* the temperature values associated with *cold*, hence they should be related to a yLevel of 1. H(moderate) on the other hand represents the *temp*erature values which are associated with *moderate* and as such not anymore related with *cold*, thus, in terms of the fuzzy set **cold**, it should be ass*ociated* with yLevel 0 at H(moderate).

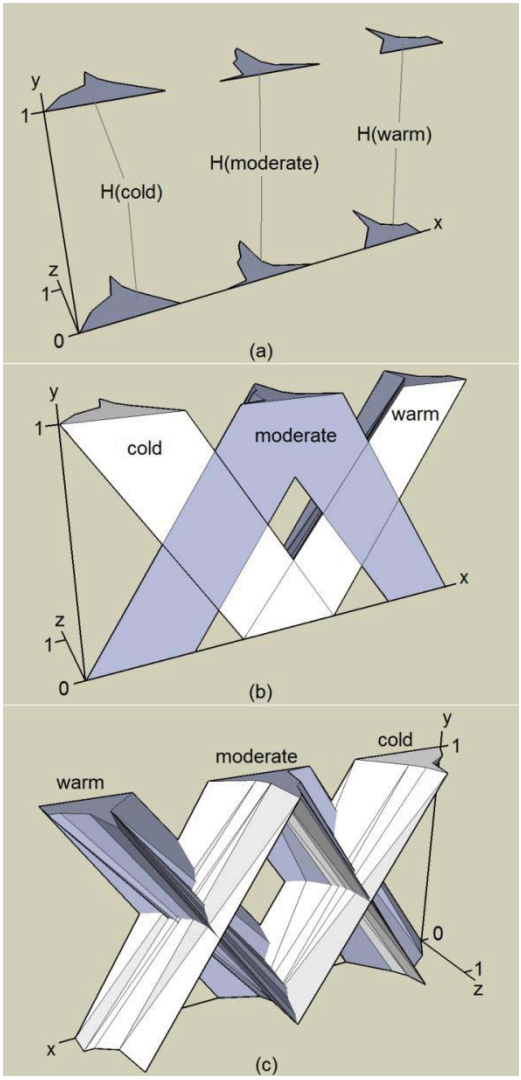


Figure 3: Schematic representation of the steps involved in creating the fuzzy set representation of linguistic labels. (a) "Histogram representations" of the linguistic labels associated with the respective levels on the yAxis before interpolation. (b) Fuzzy sets after interpolation from the front. (c) Fuzzy sets after interpolation from the rear.

ISBN: 978-989-95079-6-8

b. For *moderate*, we associate H(cold) with the level 0 on the yAxis, H(moderate) with level 1 on the yAxis and H(warm) with yLevel 0 on the yAxis. We employ interpolation to recreate the general type-2 set for *moderate* as shown in Fig. 3a, Fig. 3b and Fig. 3c.The reasoning of the positioning on the yAxis of H(cold), H(mod*erate) a*nd H(warm) is analogous to Step a.

c. Fo*r warm,* we associate H(moderate) with level 0 on the yAxis and H(warm) with yLevel 1 on the yAxis. We employ interpolation to recreate the general type-2 set for *warm* as shown in Fig 3(a)-(c).
The reasoning of the positioning on the yAxis of H(moderate) and H(war*m) i*s analogous to Step a.

We have shown in a series of steps how the data on a linguistic variable such as temperature can be used to generate a general type-2 representation of the given variable. It is straightforward to extract zSlices based general type-2 fuzzy sets as introduced in [11], interval type-2 and type-1 sets. The individual steps are shown for the general case in Section 4.

3 Design of fuzzy sets based on device or variable characteristics

In the context of Fuzzy Logic Controllers (FLCs), fuzzy logic sets are generally employed to model the uncertainty associated with the inputs and outputs of the FLC. In real world applications these inputs and outputs are commonly physical devices such as sensors and actuators which are susceptible to a wide range of sources for uncertainty, from wear and tear, to environmental conditions, signal noise etc. General type-2 fuzzy sets are well suited to model this uncertainty and in this section we are describing how information about the device or variable to be modelled can be directly incorporated into the fuzzy sets modelling. We will be using the illustrative example of generating the MFs for a sonar sensor from information which is directly retrieved from the device. While the example is referring to a sonar sensor, the overall method is applicable to every device or artefact that is to be modelled.

Generation of two MFs *Near* and *Far* which model the distance value measured using a Sonar sensor.

1. As a first step, we discretize the maximum sensor range into X steps. A finer level of discretization will result in a more accurate model of the uncertainty contained within the sensor outputs. The process of discretizing the sensor range of a Pioneer robot side sonar is shown in Fig. 4.
2. At every step, the device, in our case the sonar sensor is sampled (for example by presenting the sonar with an obstacle at the specified distance in our case) Y times. A larger number Y of samples will result in a more accurate model of the uncertainty contained within the sensor outputs. The series of Y readings are analysed to create a mapping detailing the normalized frequency of each specific sample. As such, in our case, the reading (or close group of readings) which was measured the largest number of times out of Y samples is mapped to 1, readings that have never been measured are mapped to 0 etc. This approach can be likened to a continuous (high discretization level) histogram. An example of this can be seen in Fig. 4. At each sampling step x, Y samples are taken which are translated to the respective histograms H(x).

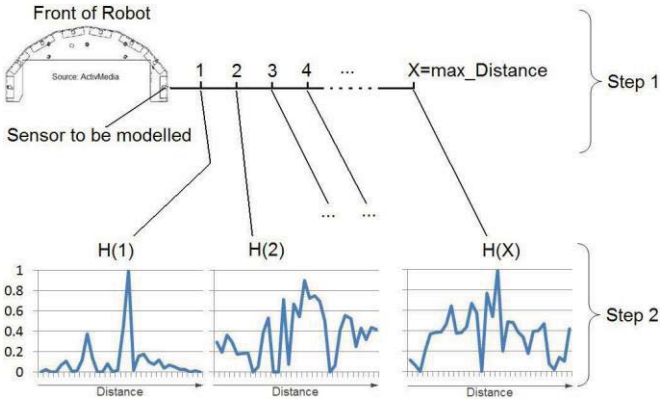


Fig. 4 Step 1 and 2 for the automatic generation of MFs for the sonar sensor mounted on a Pioneer robot.

3. The histograms generated in Step 2 are representative of the uncertainty encountered while sampling the specific distance x, i.e. while usually the majority of samples return the accurate reading, some will return a different reading because of sensor noise, environmental impact, etc. The information contained in the histograms can thus be employed to define the third dimension of general type-2 fuzzy sets which models the uncertainty within the sensor. In order to use the histograms to define the third dimension of general type-2 fuzzy sets, the histograms need to be strictly convex. The convexity of the third dimension of fuzzy sets is a general requirement in several general type-2 fuzzy logic operations such as the join and meet computed on the membership grade presented in [12]. While we are planning to address the direct handling of non-convex sets in the future, at this stage we are converting the existing histogram into a convex function. While this can be done through simple triangular approximation, this results in a very poor representation of the information contained in the initial histogram. We have devised a recursive algorithm which converts the histograms H(x) to their convex counterparts C(x) while preserving as much information as possible. An example comparing an initial histogram to its convex approximations using the triangular approach and our algorithm can be seen in Fig. 5 and an in depth

ISBN: 978-989-95079-6-8

description of the algorithm and further comparisons can be found in Section 5.

It should be noted that the algorithm can be applied to any type-1 fuzzy set and as such to the histograms as presented here as well as any vertical slices of the general type-2 fuzzy set if required.

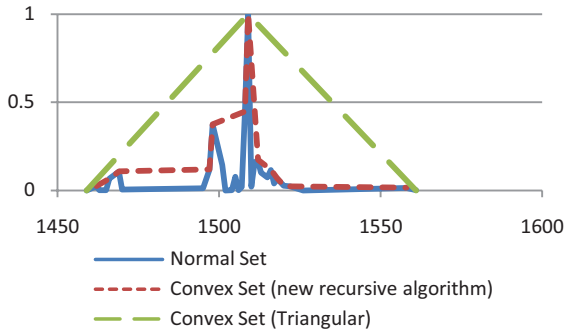


Figure 5: Comparison between an initial non-convex histogram (set) and its convex approximation using the triangular and our recursive algorithm approach.

4. For each linguistic label, we will need to determine the position of the linguistic labels within the device input/output range. In our example we are using triangular MFs and as such would define three points for every linguistic variable: its start, its maximum and its end. Fig.6 shows the Near MF for which in our example we defined the start as Ns, the maximum as Nm and the end as Ne. All previously generated convex H(x) where $x \in [start, end]$ are mapped to the triangle formed by {start, maximum, end}, "raising" them to the respective level on the yAxis. This process is illustrated in Fig. 6 for 5 histograms. In case there is uncertainty (for example due to linguistic uncertainties) about the positions of points Ns, Nm and Ne, we will have various triangles, each one is associated with the relevant histograms as shown in Fig.6. These triangles are then interpolated (as explained in the following step) in the x-y and the y-z domain to generate the FOU and third dimension of the general type-2 fuzzy set.

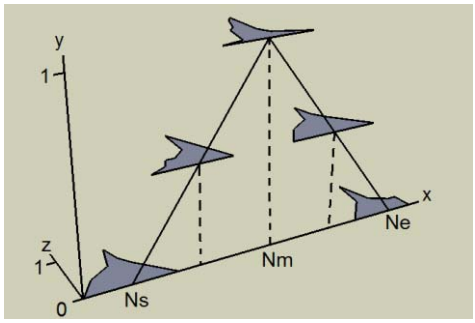


Fig. 6 The fuzzy MF *Near* is composed of the endpoints defined by the linguistic variable combined with the histograms of any sampling po*ints* which lie within the range of the linguistic variable. Here, five histograms are shown which have been "raised" to their respective level on the yAxis which is defined by the intersection of their sampling point x with the triangular {Ns,Nm,Ne}.

5. In the final step towards the generation of a general type-2 MF, the different histograms are linked through interpolation. It should be noted that the choice of interpolation method and the number of histograms directly affect the quality of the uncertainty model of the MF. Fig. 7 illustrates the front and side view of our exemplary MF *Near*. If it is requirement for the vertical slices of the resulting set to be strictly convex, the same convexity algorithm as described above can be applied to the vertical slice.

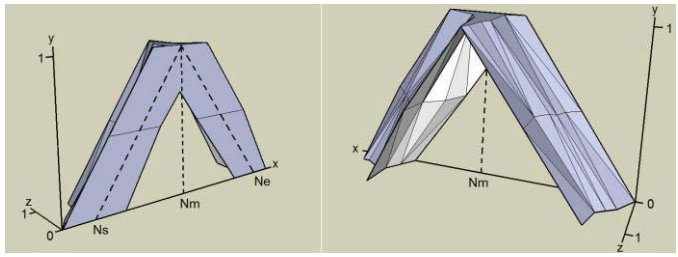


Figure 7: (a) Front view of the generated membership function for *Near*; (b) rear view of the same set.

The procedure *for* the automatic generation of general type-2 MFs presented in this section allows the generation of MFs which reflect and model the uncertainty encountered by or inherent to a specific device. The generated general type-2 MFs can easily be used to extract type-1, interval type-2 and zSlices based general type-2 fuzzy sets as we will be shown in Section 4.

4 Extraction of zSlice based general, interval type-2 and type-1 fuzzy sets.

It is straightforward to extract type-1 MFs and interval type-2 MFs from the general type-2 MFs generated in Sections 2 and 3 by restricting the zLevel (i.e. the value on the zAxis) to 1 or 0 respectively (in the case of interval type-2 MFs, they are subsequently "raised" to a zLevel of 1). This allows the process to generate type-1 and interval type-2 MFs from data without altering the procedure, which should facilitate the performance comparison between different levels of complexity in FLCs. An example of the MFs extracted from the general type-2 fuzzy sets shown in Fig. 7 is shown in Fig. 8.

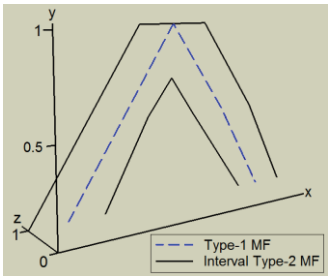


Figure 8: The type-1 and interval type-2 MFs extracted from the general type-2 MF shown in Fig. 7

ISBN: 978-989-95079-6-8

Finally, it is worth noting that the fuzzy sets generated in Sections 2 and 3 can easily be transformed to zSlices based MFs which were introduced in [11]. Fig. 9 shows an example of the generated zSlice based MF from the general type-2 MF shown in Fig. 7.

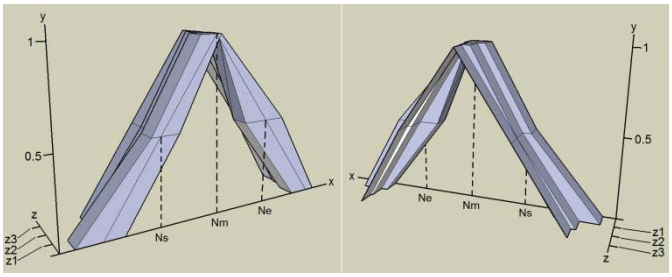


Figure 9: zSlice based general type-2 version (3 zSlices) of the general type-2 set shown in Fig. 15. It should be noted that only the zSlices 1-3 are shown, zSlice 0 has been omitted to improve visibility.

5 Approximation of a convex set from a non-convex set.

5.1 Description of the algorithm.

In order to generate a convex function from a non-convex function (more specifically, the discretized points from this function) while maintaining a high level of similarity to the original function, we have devised a recursive algorithm which is described in detail in this section.

The algorithm takes a discretized non-convex function (such as a type-1 fuzzy set) F as input. The discretization level along the xAxis is referred to as D. The function F matches each point on the xAxis x_d to a point y_d on the yAxis (where $d \in [1, D]$). As such, the function is represented as a series of points P_D in two-dimensional space where the x and y coordinates of a point P_d are referred to as P_{d_x} and P_{d_y}.

In our application of the convex approximation of "continuous histograms" (as discussed in Sections 2 & 3), the x-Axis represents the input axis, for example the distance, if we are considering a distance sensor, while the yAxis represents the number of samples. We denote the convex output set as O.

The recursive algorithm performs the following steps:

0.
   a. Traverse set from left to right (arbitrary- or from right to left) and find the coordinates of the discretized point P_{Max} (P_{Max} is defined in two-dimensional space by its coordinates P_{Max_x} and P_{Max_y}), where
      $P_{Max} \in P_D$ and
      $P_{Max_y} = \max\left(P_{d_y}\right), d \in [1, D]$.
      If more than one point with the same maximum is found, the middle point (according to its x coordinates P_{d_x}) is chosen. If the number of points with the same maximum is even, one of the middle points is chosen arbitrarily.
   b. Add point P_{Max} to the output set O.
   c. Split the series of points into two series: $\{P_1, \dots, P_{Max}\}$ and $\{P_{Max}, \dots, P_D\}$.
   d. Pass the series $\{P_1, \dots, P_{Max}\}$ to Step 1 and the series $P_{Max}, \dots, P_D\}$ to Step 2.

1.
   a. Refer to the given series of points as $[P_{start}, P_{end}]$.
   b. $IF\ (P_{start} == P_{end})$ THEN STOP.
   c. Find the new maximum point $P_{Max'}$ of the given series such that:
      $P_{Max'_y} > P_{start_y}\ AND\ P_{Max'_y} < P_{end_y}$
      $P_{Max'_y} = \max\left(P_{d_y}\right), d \in]start, end[$.
   d. Add point $P_{Max'}$ to the output set O.
   e. Split the given series into two series: $\{P_{start}, \dots, P_{Max'}\}$ and $\{P_{Max'}, \dots, P_{end}\}$ and pass both series separately to Step 1 (recursion).

2.
   a. Refer to the given series of points as $[P_{start}, P_{end}]$.
   b. $IF\ (P_{start} == P_{end})$ THEN STOP.
   c. Find the maximum point $P_{Max'}$ of the given series with endpoints such that
      $P_{Max'_y} < P_{start_y}\ AND\ P_{Max'_y} > P_{end_y}$
      $P_{Max'_y} = \max\left(P_{d_y}\right), d \in]start, end[$.
   d. Add point $P_{Max'}$ to the output set O.
   e. Split the given series into two series: $\{P_{start}, \dots, P_{Max'}\}$ and $\{P_{Max'}, \dots, P_{end}\}$ and pass both series separately to Step 2 (recursion).

Below we give an example of the application of the recursive algorithm.

5.2 Example application of the recursive algorithm.

In Table 2 we indicate the coordinates of a non convex set shown in Fig.10 taken from real world experiments and compare it to the coordinates computed by our algorithm and the triangular approximation.

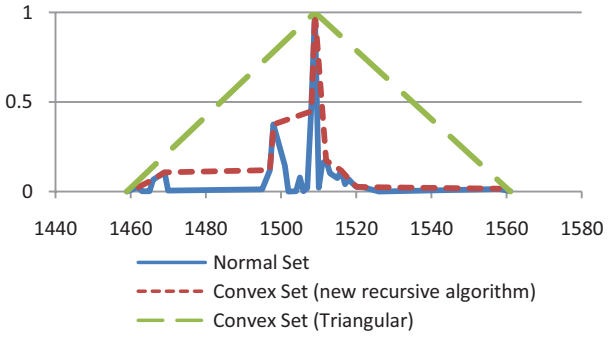


Fig. 10 A plot of the non convex input function as well as both the recursive and triangular approximated

From Table 2 it can be seen that both the recursive and the triangular algorithms omit a significant amount of detail in order to generate a convex approximation of the input set. Nevertheless the number of preserved samples and thus detail and information is much larger in the recursive approximation. This is also visible in Fig. 10 which shows that the recursive algorithm provides a vastly superior fit to

ISBN: 978-989-95079-6-8

the original function in comparison to the triangular approximation.

Table 2 Comparison of z and y tuples preserved by our recursive algorithm and a triangular algorithm while converting the non-convex set shown in Fig. 10.

Non-convex Set				Convex Set (Recursive Algorithm)		Convex Set (Triangular Algorithm)	
x	y	x	y	x	y	x	y
0.00	1459	0.44	1508	0.02	1462	0	1459
0.02	1462	1.00	1509	0.06	1466	1	1509
0.00	1463	0.02	1510	0.10	1469	0	1561
0.00	1465	0.15	1511	0.11	1497		
0.06	1466	0.17	1512	0.37	1498		
0.10	1469	0.10	1513	0.44	1508		
0.00	1470	0.07	1515	1.00	1509		
0.01	1495	0.11	1516	0.17	1512		
0.11	1497	0.04	1517	0.11	1516		
0.37	1498	0.06	1518	0.06	1518		
0.14	1501	0.04	1519	0.04	1519		
0.00	1502	0.02	1520	0.02	1520		
0.00	1504	0.02	1522	0.02	1522		
0.07	1505	0.00	1526	0.01	1558		
0.00	1506	0.01	1558				
0.02	1507	0.00	1561				

6 Conclusions

In this paper we have addressed the reoccurring problem in fuzzy systems of specifying the parameters for fuzzy membership functions (MFs). The correct selection of parameter of fuzzy MFs is crucial in order for the fuzzy systems to model the uncertainty contained within the system correctly and thus to allow the generated fuzzy system to provide good performance. While the problem has been previously addressed to some extent for interval type-2 and type-1 fuzzy systems, the use of general type-2 fuzzy sets requires the specification of an even larger number of parameters for each MF and as such is set to exasperate the problem of choosing those parameters.

We have presented two novel methods which allow the design of general type-2 fuzzy MFs directly from data and as such eliminate the need for a manual design of the MFs or the use of artificial intelligence techniques in order to tune the MFs.

The first method allows the design of general type-2 fuzzy MFs using data on linguistic labels which is collected through a survey. The technique allows for an accurate representation of the perception of linguistic variables gathered through the survey without any loss of information.

The second method specifies the general type-2 MFs based on data which is directly sampled from the device (such as a sensor) which is to be modelled and as such creates an accurate representation of the uncertainty associated with the specific device.

Both methods further allow the extraction of type-1, interval type-2 and the recently introduced zSlices based general type-2 fuzzy sets.

Additionally, we have presented a recursive algorithm that computes a convex approximation of fuzzy sets while preserving significantly more detail than simple triangular approximation which in turn enables a more accurate modelling of the uncertainty within the fuzzy MFs.

In the future, we are planning to further refine the algorithms for the automatic definition of fuzzy sets while investigating their performance in real world applications. Particularly, we are looking to extend the theoretical frameworks to non-convex fuzzy sets and to investigate non-singleton fuzzification in conjunction with the automatically created sets.

Acknowledgment

This work was supported in part by the European Commission under Grant FP7-ICT-1-8.2/216837 entitled “Adaptive and Trusted Ambient Ecologies” (ATRACO).

References

[1] Mendel and R. John; “Type-2 Fuzzy Sets Made Simple” ; IEEE Transactions on Fuzzy Systems, Vol. 10, pp.117-127, April 2002.

[2] H. Hagras, "A Hierarchical type-2 fuzzy logic control architecture for autonomous mobile robots,” *IEEE Transactions On Fuzzy Systems*, vol. 12 No. 4, pp. 524-539, August 2004.

[3] J. Figueroa, J. Posada, J. Soriano, S. Rojas and M. Melgarejo, “A Type-2 Fuzzy Controller for Tracking Mobile Objects in the Context of Robotic Soccer Games”, Proceedings of the International Conference of Fuzzy System, May 2005, pp.359 – 364.

[4] J. Mendel, Uncertain Rule-Based Fuzzy Logic Systems, Prentice Hall 2001.

[5] A. Homaifar and E. McCormick, “Simultaneous design of membership functions and rule sets for fuzzy controllers using genetic algorithms,” IEEE Transactions Fuzzy Systems, vol. 3, No. 2, pp. 129-139 May 1995.

[6] C. Lynch. H. Hagras and V. Callaghan, “Embedded Interval type-2 Neuro-Fuzzy Speed Controller for Marine Diesel Engines,” Proceedings of the 2006 Information Processing and Management of Uncertainty in Knowledge-based Systems conference, Paris, France, July 2006, pp. 1340-1347.

[7] S. Coupland and R. John, “Geometric type-1 and type-2 fuzzy logic systems,” IEEE Transactions on Fuzzy Systems, vol. 15, pp. 3-15, Feb. 2007.

[8] S. Greenfield and R. John, “Optimised Generalised Type-2 Join and Meet Operations,” Proceedings of the IEEE International Conference of Fuzzy Systems, London, UK, July 2007, pp. 141-146.

[9] L. Lucas, T. Centeno and M. Delgado, “General type-2 fuzzy inference systems: analysis, design and computational aspects,” Proceedings of the IEEE International Conference of Fuzzy Systems, London, UK, July 2007, pp. 1107-1112.

[10] J. Mendel, Feilong Liu, “On new quasi-type-2 fuzzy logic systems,” Proceedings of the IEEE International Conference of Fuzzy Systems, Hong Kong, China, June 2008, pp. 354-360.

[11] C. Wagner and H. Hagras, “zSlices – Towards bridging the gap between Interval and General Type-2 Fuzzy Logic,“ Proceedings of the IEEE International Conference of Fuzzy Systems, Hong Kong, China, June 2008, pp. 489-497.

[12] N. Karnik and J. Mendel, "Operations on type-2 fuzzy sets," Fuzzy Sets and Systems, vol. 122, pp. 327-348, 2001.

ISBN: 978-989-95079-6-8

Flow Line Systems with Possibilistic Data: a System with Waiting Time in Line Uncertain

David de la Fuente [1] María José Pardo [2]

1 Polytechnic School of Engineering-Gijon, University of Oviedo
Gijón, Spain
2 Faculty of Economics, University of Basque Country
Bilbao, Spain
Email: david@uniovi.es, mjose.pardo@ehu.es

***Abstract*—** *This paper proposes to analyze two flow line systems in which we include possibilistic data -the priority-discipline is possibilistic instead of probabilistic- and measure the performances of the systems with the effectiveness measure "waiting time in queue". In a previous work we have analyzed and developed a queuing model with uncertain priority-discipline, using Zadeh's extension principle. Because of it in this paper the analysis of the model takes the work realized by Prade as a starting point to incorporate an uncertain priority-discipline in a flow line system by means of a possibilistic distribution. To verify the validity of the proposed method, we calculate the possibility distribution of the priority-discipline and the possibility distribution of the performance measure "waiting time in queue", in two flow line systems with different waiting lines: the determinist flow line system $D/D/1$ and the classic flow line system $M/M/1/N$ with finite capacity. Since the performance measure is expressed by membership function rather than by crisp value, the fuzziness of input information is conserved completely, and more information is provided for capacity planning in flow line systems.*

Keywords— Flow Line Systems, Fuzzy Numbers, Possibility Theory, Priority-Discipline

1 Introduction

In this work we present two flow line systems to which we add a possibilistic priority-discipline instead of a probabilistic priority-discipline. For the development of the models we will follow the methodology exposed by *Prade* [1] about possibilistic models of queuing systems. In the literature regarding the queuing models with uncertain data we have not found any article that uses and develops his method in order to incorporate uncertainty in the queuing models with priority-discipline (except for *Prade*'s work). For example, the works of Li and Lee [2], Negi and Lee [3], or more recently Chen [4] or Pardo and de la Fuente [5,6] among others, study the flow line systems with uncertain data according to the Zadeh's extension principle [7], and in the work of Pardo and de la Fuente [8] the optimization of the fuzzy queuing model with priority-discipline is also studied according to the Zadeh's extension principle.

2 Calculation of the membership function of possibilistic priority-discipline

The possibilistic priority-discipline indicates the possibility that has the unit that arrives to the system in the position *n* to be overtaken, displaced or served in a different position that *n* in the waiting line. To develop the proposed models in this work, we need to define a possibility distribution of the priority-discipline but the possibility distribution can be calculated from the different existing methods to construct possibility distributions from real data or from the opinion of the experts (for example in Dubois et al. [9]).

In this way, we define the following possibility distribution (we denote by μ_i^n) and it provides us with the possibility that has the unit that has arrived in the position *n* to be served in the position $n+i$. Logically, this is valid when there is a queue length of one or more units at the arrival of one unit to the system, depending on the case. Besides and because of simplicity, the possibility distribution of the priority-discipline will be considered independent of the unit *n*. For this reason, the possibility of the unit *n* to advance one or more positions in the queue is:

1. The possibility of being served in the position *n* is: $\mu_0^n = 1$.
2. The possibility of being served in the position $n-1$ is: $\mu_{-1}^n = \sigma_1 \left(\sigma_1 \le 1\right)$.
3. The possibility of being served in the position $n-2$ is: $\mu_{-2}^n = \sigma_2 \left(\sigma_2 \le 1\right)$. In this case we consider that the unit $n-1$ and the unit $n-2$ move back one position in the waiting line.
4. The possibility of being served in the position $n+i$ with $i = -3,-4,...$ is: $\mu_i^n = 0$. In this way one unit can not advance more than two positions in the waiting line.

From the following hypotheses, we calculate the possibility that has the waiting unit to be moved back one position in the waiting line because of the arrival of one unit with higher priority. Then:

1. There are two reasons for the unit *n* to move back one position in the waiting line:
   i) Because the unit $n+1$ overtakes the unit *n*. That happens because it has priority to advance one position in the queue or because it has priority to advance two positions in the queue. The

ISBN: 978-989-95079-6-8

possibility of the event "the unit $n+1$ overtakes the unit n" is the association of the two previous events and it is know that it has a certain rate according to the fuzzy logic (Zimmermann [10]):

$$\max\{\sigma_1, \sigma_2\} \tag{1}$$

ii) If the unit $n+1$ does not have higher priority to the unit n, but the unit $n+2$ has higher priority to the units n and $n+1$ and advances two positions, the units n and $n+1$ move back one position and that happens with a possibility grade of σ_2.

Then, the possibility grade of the unit n to move one position back once it has arrived to the system is the association of the previous events, and according to the fuzzy logic (Zimmermann [10]):

$$\mu_1^n = \max\{\max\{\sigma_1, \sigma_2\}, \sigma_2\} = \max\{\sigma_1, \sigma_2\} \tag{2}$$

2. The unit n can move back two positions in the queue if it moves back twice consecutively, that means if the event "the unit n moves back one position" happens twice, what in fuzzy logic has a certain rate (Zimmermann [10]):

$$\mu_2^n = \min\{\mu_1^n, \mu_1^n\} = \mu_1^n = \max\{\sigma_1, \sigma_2\} \tag{3}$$

3. Likewise, the possibility grade of the unit to be moved back three or more positions in the waiting line can be calculated and in all of them we have:

$$\mu_i^n = \max\{\sigma_1, \sigma_2\} \quad \text{with} \quad i = 3, 4, \ldots \tag{4}$$

Then, the possibility grade of the unit n to move back in the waiting line, regardless of the number of positions is:

$$\mu_i^n = \max\{\sigma_1, \sigma_2\} \quad \text{with} \quad i = 1, 2, \ldots \tag{5}$$

The possibility distribution of the priority-discipline in the waiting line is:

$$\mu_i^n = \begin{cases} 0 & \text{if} \quad i = -3, -4, \ldots \\ \sigma_2 & \text{if} \quad i = -2 \\ \sigma_1 & \text{if} \quad i = -1 \\ 1 & \text{if} \quad i = 0 \\ \max\{\sigma_1, \sigma_2\} & \text{if} \quad i = 1, 2, \ldots \end{cases} \tag{6}$$

Once we know μ_i^n we obtain the "possibility distribution of the permanence time in line of the unit n" (denoted $\tilde{W}_q^n$), taking into account that if the unit n advances two positions, it does not have to wait in the waiting line to the service of the units $n-2$ and $n-1$. I it advances one position, we have to take away the service of the unit $n-1$. If it maintains its arrival position, its permanence time in queue does not change. And if it moves back in the queue it has to wait its permanence time and also so many service times as units that had overtaken it. Using Zadeh's notation to denote a fuzzy number (Zadeh [7]) (the possibility grade is previous to the symbol "/", the possible value is posterior, and the symbol "+" represents the association), if it is denoted by b the service time of one unit, the possibility distribution of the permanence time in queue of the unit n is:

$$\begin{aligned} \tilde{W}_q^n = {} & \mu_{-2}^n / \left(W_q^n - 2b\right) + \mu_{-1}^n / \left(W_q^n - b\right) + \\ & + \mu_0^n / W_q^n + \mu_1^n / \left(W_q^n + b\right) + \mu_2^n / \left(W_q^n + 2b\right) + \ldots \end{aligned} \tag{7}$$

Depending on the particular features of the flow line system to which is added this possibilistic priority-discipline, we obtain some results that differ among the different models. In the following sections we show the obtained theoretical results with two flow line systems with possibilistic priority-discipline.

3 Determinist flow line system $D/D/1$ with times among arrivals longer than the service time and possibilistic priority-discipline

In the determinist flow line system, the units arrive to the system with time among arrivals constant and are also served in constant times. It is supposed that the system is organised in a way that the arrivals happen in regular intervals and they are also regularly served in an only channel. The arrival source is considered unlimited and there is no limitation in the capacity of the system.

Saaty [11] analyses the determinist queuing system D/D/1 in transitory state. That means that the initial conditions (the number of units that there are in the system at the initial moment) affect the obtained results in the model. In addition to this, Saaty studies the system with times among arrivals longer than the service times. This condition supposes that the units take in arriving to the system more time that the system needs to serve them, that is why there will never be a waiting line when starting the system there are no units in it (initial condition $I = 0$ when $t = 0$, with I the number of initial units in the system). Each unit that arrives to the system is served before the arrival of the next unit. Due to this, Saaty studies the determinist model supposing that it is able to empty (the units in the system are not accumulated) under the initial existence conditions of I units in the system at the initial moment.

To know the evolution of this system and to calculate its performance measure, we have to analyse the behaviour of this flow line. The I units are found in the facilities when the system start to work and since the system is able to absorb the next arrivals (they are served very fast) then there will be an moment (denoted T) in which there will be no units in the system and each one arriving to the system will be served before the arrival of the next one. In T, the system reaches the steady state: the server stays idle until a new arrival, there are no units in the waiting line and the permanence time in queue is zero. In order to define the system completely, we need to calculate the number of units that arrive to the system before T (denoted A). They are the units that have to stay in the waiting line together with the initials.

ISBN: 978-989-95079-6-8

Regarding the performance measure "waiting time in queue of the unit *n*", W_q^n Saaty does not calculate it for the different values of *n*. That is why before incorporating a possibilistic priority-discipline, we will first calculate the variable *A* and the performance measure W_q^n. It will be so denoted:

a: Time among consecutive arrivals.

b: Service time.

I: Number of units in the system at the initial moment (including the unit that arrives to the system at the initial moment and the units that were already in it). The case $I=1$ means that at the arrival of the first unit, the system was empty.

The system reaches the steady state in *T* once the server has attended the initial units *I* and the new arrivals *A*. The arrival $A+1$ is the first one that finds the server free and goes directly to be attended. This arrival happens in $(A+1)a$. From here, the moment in that the first unit $A+1$ arrives to the system has to be longer than the time of service of the $I+A$ units, otherwise it is no possible that such arrival finds the system free. Then:

$$(A+1)a > (I+A)b \Leftrightarrow A > \frac{Ib-a}{a-b} \tag{8}$$

The variable *A* has to be the first entire number that fulfils the previous expression, then:

$$A = \left[\frac{Ib-a}{a-b}\right] + 1 = \left[\frac{b(I-1)}{a-b}\right] \tag{9}$$

where $[x]$ represents the whole part of the number *x*, that is to say, the largest integer less than or equal to a number *x* (floor function).

The performance measure of the permanence time in queue of the unit *n*, W_q^n is:

$$W_q^n = \begin{cases} (n-1)b & \text{if} \quad n \le I \\ (n-1)b-(n-I)a & \text{if} \quad I < n \le I+A \\ 0 & \text{if} \quad n > I+A \end{cases} \tag{10}$$

We obtain the result (10) from the following reasoning: If the arrival time to the system of the unit *n* is:

$$T_{ll}^n = \begin{cases} 0 & \text{if} \quad n \le I \\ (n-I)a & \text{if} \quad n > I \end{cases} \tag{11}$$

and the time when the unit *n* is served is:

$$T_s^n = \begin{cases} (n-1)b & \text{if} \quad n \le I+A \\ (n-I)a & \text{if} \quad n > I+A \end{cases} \tag{12}$$

then $W_q^n = T_s^n - T_{ll}^n$.

3.1 Example

Once we know the behaviour of the determinist flow line systems, following we show the incorporation of a possibilistic priority-discipline in the system $D/D/1$ with practice application, calculating the possibility distribution of the permanence time in queue for each unit that when arrive to the system have to stay in the waiting line for a flow line system with time among arrivals of 30 min., service time 20 min., 8 initial units and with the possibility distribution of the priority-discipline: $\mu_{-2}^n = 0.2$; $\mu_{-1}^n = 0.4$ and $\mu_0^n = 1$, from where we have that:

$$\mu_i^n = \max\{\sigma_1, \sigma_2\} = \max\{0.4, 0.2\} = 0.4\,,\ i = 1,2,...$$

First of all with the data of the example it will be $I=8$, $A=14$, and:

$$T_{ll}^n = \begin{cases} 0 & \text{if} \quad n \le 8 \\ (n-8)30 & \text{if} \quad n > 8 \end{cases} \qquad T_s^n = \begin{cases} (n-1)20 & \text{if} \quad n \le 22 \\ (n-8)30 & \text{if} \quad n > 22 \end{cases}$$

$$W_q^n = \begin{cases} (n-1)20 & \text{if} \quad n \le 8 \\ (n-1)20-(n-8)30 & \text{if} \quad 8 < n \le 22 \\ 0 & \text{if} \quad n > 22 \end{cases}$$

Before calculating the possibility distribution of the permanence time in queue of the client *n* and given the features of the system, we have to make the following remarks:

- Because of the evolution of this system there is a unit *N*, such that $I < N \le I+A$ $(8 < N \le 22)$, which fulfils that every unit $n > N$ is served before the arrival of the next unit and it is not possible for the unit $n+1$ to overtake the unit *n*. That means, if it is denoted by T_s^N at the same moment that the unit *N* is served and by T_{ll}^{N+1} the arrival time of the unit $N+1$, then *N* is the first unit that fulfils: $T_s^N < T_{ll}^{N+1}$, so there is no possibility for the position *n* to be moved back in the waiting line, then $\mu_j^n = 0$, $j = 1,2,...$ and $n > N$, with:

  $$T_s^N = (N-1)20 < (N+1-8)30 = T_{ll}^{N+1}$$

  From here, we obtain $N > 19$, then in the example it is $N = 20$. In this way we obtain:

  1. If $n < 20$ is $\mu_j^n = 0.4$ with $j = 1,2,...,20-n$ and $\mu_j^n = 0$ with $j = 20-n+1,....$
  2. If $n \ge 20$ is $\mu_j^n = 0$ with $j = 1,2,...$
  3. And if from the unit 20 it is not possible to move back in the waiting line, then it is not possible to overtake in the line from the unit 21 and the following ones. Then, if $n \ge 21$ is $\mu_{-2}^n = \mu_{-1}^n = 0$.

- The possibilities μ_{-2}^n and μ_{-1}^n when $n \le 20$ depend on the unit *n*:

ISBN: 978-989-95079-6-8

1. If $n=1$, that is, the first unit in the system, then: $\mu_{-2}^{1}=\mu_{-1}^{1}=0$.
2. If $n=2$, then: $\mu_{-2}^{2}=0$ and $\mu_{-1}^{2}=0.4$, given that the second unit arrives at the same time that the first one and there is a possibility that it advances in the waiting line.
3. If $3\le n\le I=8$, they are units that arrive to the system in the moment zero and there is a possibility that all of them change their position advancing one or two positions in the waiting line, then: $\mu_{-2}^{n}=0.2$ and $\mu_{-1}^{n}=0.4$
4. If $9\le n\le 20$, because of the features of that system we know that the waiting line diminishes for any unit that enters, that means that it is possible for the last arrivals that the level of the queue is lower than 2 or 1 unit and they can not advance one or two positions respectively. To determine what units can make its priority effective, we follow the next reasoning: We denote N_1 the unit with priority to advance two positions in the waiting line, in order to make this priority effective, the arrival time of the unit N_1 must be lower or the same that the service time of the unit N_1-2 (the unit previous to N_1 in two positions) so it can occupy its position, then $T_{ll}^{N_1}\le T_s^{N_1-2}$, and from here we get $N_1\le 18$. And we denote N_2 the unit with priority to advance one position in the waiting line, this priority can be likewise effective if the arrival time of the unit N_2 is lower or the same that the service time of the unit N_2-1 (the unit previous to N_2 one position), then $T_{ll}^{N_2}\le T_s^{N_2-1}$, from where it is $N_2\le 20$. Then:

   i) If $9\le n\le 18$, then $\mu_{-2}^{n}=0.2$ and $\mu_{-1}^{n}=0.4$

   ii) If $n=19$ or $n=20$, then $\mu_{-2}^{n}=0$ and $\mu_{-1}^{n}=0.4$

With all this we calculate the possibility distribution of the permanence time in queue of the unit *n*:

$$\begin{aligned}\tilde{W}_q^n=\ &\mu_{-2}^{n}/\left(W_q^n-40\right)+\mu_{-1}^{n}/\left(W_q^n-20\right)+\\ &+1/W_q^n+\mu_1^n/\left(W_q^n+20\right)+\mu_2^n/\left(W_q^n+40\right)+\\ &+\ldots+\mu_{20-n}^{n}/\left(W_q^n+(20-n)20\right)\end{aligned}$$

Following is detailed this possibility distribution for some values of *n*:

$$\begin{aligned}\tilde{W}_q^1&=1/0+0.4/20+\ldots+0.4/380\\ \tilde{W}_q^2&=0.4/0+1/20+0.4/40+\ldots+0.4/380\\ \tilde{W}_q^8&=0.2/100+0.4/120+1/140+0.4/160+\ldots+0.4/380\\ \tilde{W}_q^{15}&=0.2/30+0.4/50+1/70+0.4/90+\ldots+0.4/170\\ \tilde{W}_q^{20}&=0.4/0+1/20\\ \tilde{W}_q^{21}&=1/10\\ \tilde{W}_q^n&=1/0\quad\text{if}\quad n\ge 22\end{aligned}$$

Just as the possibilistic priority-discipline has been adapted to the determinist flow line system $D/D/1$, with time among arrivals longer than the service time, it can be incorporated to other determinist queuing systems only taking into account the features of the system and the evolution of the waiting line.

4 Flow line systems $M/M/1/N$ with possibilistic priority-discipline

The classic flow line system $M/M/1/N$ with finite capacity is considered, with times among arrivals and service time distributed according to an exponential of parameters λ and μ respectively, equals to all the units and regardless the kind of priority. The system has an unlimited enter source and the capacity is *N* units in the system.

The system is analysed in steady state so we know the probabilities of the system to be empty [12]:

$$P_0=\left(\sum_{n=0}^{N}\left(\frac{\lambda}{\mu}\right)^n\right)^{-1} \tag{13}$$

and that the units *n* are found in the system:

$$P_n=P_0\left(\frac{\lambda}{\mu}\right)^n \text{ with } n=1,2,\ldots,N \tag{14}$$

and we also know that the average permanence time in queue of each unit is:

$$W_q=\frac{L_q}{\lambda\left(1-P_N\right)} \tag{15}$$

where:

$$L_q=P_0\frac{\lambda}{\mu}\left(\sum_{n=2}^{N}(n-1)\left(\frac{\lambda}{\mu}\right)^{n-1}\right) \tag{16}$$

is average length of the units in queue.

To incorporate the possibilistic priority-discipline before indicated we have to take into account that in the stochastic flow line systems we do not know for sure the moment when the unit arrives to the system, which the length of the queue is when the unit arrives and if, after that, one ore more units can arrive and make that the previous unit move back in the queue from its arrival position. Then, we should calculate the average possibility to advance one or two

ISBN: 978-989-95079-6-8

positions in line, the average possibility to remain in the arrival position or the average possibility of moving back in the queue:

1. When the unit *n* arrives to the system, if it has priority to overtake 2 units and besides there is a waiting line with 2 or more units in the queue (3 units or more in the system), the priority-discipline will be made effective. The possibility of overtaken two units is μ_{-2}^{n} and the probability that 3 or more units are found in the system is $P_{n\geq 3} = 1 - P_0 - P_1 - P_2$. For this reason the average possibility of one unit advancing two positions in the queue (denoted p_{-2}) happens when both happen at the same time, and using the definition of Zadeh [13] of the probability of a fuzzy event is:

$$p_{-2} = \mu_{-2}^{n} \cdot P_{n\geq 3} \tag{17}$$

2. With a similar reasoning to the previous one, we conclude that the average possibility of one unit advancing one position is:

$$p_{-1} = \mu_{-1}^{n} \cdot P_{n\geq 2} \tag{18}$$

with $P_{n\geq 2} = 1 - P_0 - P_1$

3. The possibility of remaining in its position in the waiting line always happens, regardless the number of units that there are in the system at the moment that the unit *n* arrives, then:

$$p_0 = \mu_0^{n} = 1 \tag{19}$$

4. The average possibility of moving back in the queue, p_i with $i = 1, 2, ..., N-1$ happens when the unit has to wait to be attended since another unit can arrive meanwhile with higher priority. One unit has to wait if the system is busy when it arrives to it and this happens with probability $P_{n>0} = 1 - P_0$, then:

$$p_i = \mu_i^{n} \cdot P_{n>0} \tag{20}$$

Once we know the average possibility of advancing or moving back in the waiting line, the possibility distribution of the average permanence time in queue is:

$$\begin{aligned} \tilde{W}_q = {} & p_{-2} / \left(W_q - 2/\mu\right) + p_{-1} / \left(W_q - 1/\mu\right) + \\ & + p_0 / W_q + p_1 / \left(W_q + 1/\mu\right) + p_2 / \left(W_q + 2/\mu\right) + ... \end{aligned} \tag{21}$$

4.1 Example

To conclude, we illustrate with an example the incorporation of a possibilistic priority-discipline calculating the possibility distribution of the average permanence time in queue for a flow line system $M/M/1/10$ of parameters $\lambda = 6$ and $\mu = 8$ units per hour, and $N = 10$ units. The possibility distribution of the waiting discipline is: $\mu_{-2}^{n} = 0.4$, $\mu_{-1}^{n} = 0.6$ and $\mu_0^{n} = 1$. From here: $\mu_i^{n} = \max\{0.4, 0.6\} = 0.6$, with $i = 1, 2, ..., 9$.

With this data we have: $P_0 = 0.261$, $P_1 = 0.1958$ and $P_2 = 0.1468$, then $P_{n\geq 3} = 0.3964$, $P_{n\geq 2} = 0.5432$ and $P_{n>0} = 0.739$. The average possibility of advancing or moving back in the waiting line is: $p_{-2} = 0.1586$, $p_{-1} = 0.3259$, $p_0 = 1$ and $p_i = 0.4434$ with $i = 1, 2, ..., 9$ The average permanence time in the waiting line of one unit is: $W_q = 0.3004$. And the possibility distribution of the average permanence time in the queue is:

$$\begin{aligned} \tilde{W}_q = {} & 0.1586/0.0504 + 0.3259/0.1754 + \\ & + 1/0.3004 + 0.4434/0.4254 + \\ & + 0.4434/0.5504 + ... \end{aligned}$$

The possibility distribution $\tilde{W}_q$ is represented in the Fig. 1.

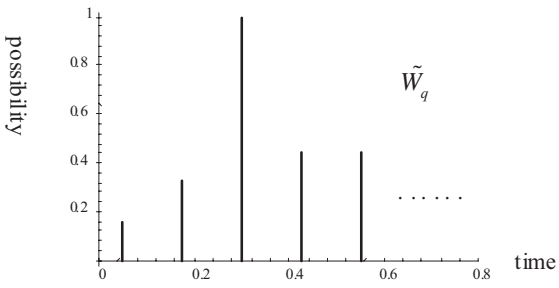


Figure 1: Possibility distribution $\tilde{W}_q$

The possibilistic priority-discipline can be adapted to other stochastic flow line system only by following a similar procedure to the one carried out in the last example.

5 Conclusions

In this work, taking the work started by *Prade* [1] as a starting point and incorporating it a possibilistic priority-discipline, we have given a more realistic representation of the flow line systems. For this we have defined a possibility distribution for the waiting-discipline but it can be calculated from real data or from the opinion of the experts. To illustrate the proposed method and the obtained theoretical results we have carried out two applications: one, to incorporate the possibilistic priority-discipline in a pure determinist flow line system and the other one to incorporate the uncertain priority-discipline in a classic flow line model. In both applications we have developed the technique to incorporate the priority-discipline given the particular features of the system and on the other hand, we have exposed an example that explains the proposed methodology.

Even so, the flow line systems with uncertain waiting line presented and developed in this work provide the decision maker with more complete and informative results and also with a better knowledge about the behaviour of the system due to the results that are possibility distributions that

ISBN: 978-989-95079-6-8

include all the uncertainty contained in the system. It is also a simple methodology that can be adapted to the different flow line systems with uncertain priority-discipline. For this reason we think that the proposed flow line system with uncertain data can have more applications than the classic one.

References

[1] H.M. Prade, *An outline of fuzzy or possibilistic models for queuing systems*, in: Fuzzy Sets: Theory and Applications to Policy Analysis and Information Systems. Wang, P.P. and Chang, S.K. (eds.) New York: Plenum Press, 1980.

[2] R.J. Li and E.S. Lee, Analysis of fuzzy queues. *Computers and Mathematics with Applications*, 17 (7):1143-1147, 1989.

[3] D.S. Negi and E.S. Lee, Analysis and simulation of fuzzy queues. *Fuzzy Sets and Systems*, 46:321-330, 1992.

[4] S.P. Chen, Solving fuzzy queuing decision problems via a parametric mixed integer nonlinear programming method. *European Journal of Operational Research*, 177 (1):445-457, 2007.

[5] M.J. Pardo and D. de la Fuente, Optimizing a priority-discipline queuing model using fuzzy set theory. *Computers and Mathematics with Applications*, 54 (2):267-281, 2007.

[6] M.J. Pardo and D. de la Fuente, Optimal selection of the service rate for a finite input source fuzzy queuing system. *Fuzzy Sets and Systems*,. 159 (3):325-342, 2008.

[7] L.A. Zadeh, Fuzzy sets. *Information and control*, 8:338-353, 1965.

[8] M.J. Pardo and D. de la Fuente, Design of a fuzzy finite capacity queuing model based on the degree of customer satisfaction: analysis and fuzzy optimization. *Fuzzy Sets and Systems*, 159 (24):3313-3332, 2008.

[9] D. Dubois, H.M. Prade and R.R. Yager, *Merging fuzzy information*, in: Fuzzy Sets in Approximate Reasoning and Information Systems. Boston: Kluwer, 1999.

[10] H.J. Zimmermann, *Fuzzy Set Theory and its applications*, 4th ed. Boston: Kluwer, 2001.

[11] T.L. Saaty, *Elements of queuing theory with applications*. New York: Dover, 1983.

[12] W.L. Winston, *Operations Research: Applications and Algorithms*, 4th ed. Boston: Duxbury Press, 2003.

[13] L.A. Zadeh, Probability measures of fuzzy events. *Journal of Mathematical Analysis and Applications*, 23:421-427, 1968.

ISBN: 978-989-95079-6-8

Valuations on the Algebra of Intervals

Mirko Navara

Center for Machine Perception, Department of Cybernetics
Faculty of Electrical Engineering, Czech Technical University in Prague
Technická 2, CZ-166 27, Praha, Czech Republic
Email: navara@cmp.felk.cvut.cz

Abstract— *Interval-valued fuzzy sets are based on the algebra of subintervals of the unit interval* $[0,1]$. *We study valuations as a special type of measures on this algebra. We present a description of all valuations which preserve the standard fuzzy negation and extend the identity on the elements of the form* (x,x). *Consequences for sublattices are formulated.*

Keywords— Girard algebra, interval-valued fuzzy set, intuitionistic fuzzy set, measure, valuation

1 Introduction

In order to work with vague quantities, fuzzy sets were suggested [13]; their membership degrees are in the unit interval $I = [0,1]$. Sometimes this approach appeared too restrictive; instead of one value, the membership degree of a point was restricted to an interval in I. This led to the notion of *interval-valued fuzzy sets*. The original term was *intuitionistic fuzzy sets* [1, 2], but it causes misunderstandings because there is no relevance to the intuitionistic logic. Other synonyms found in the literature are *IFS events* [8], etc. The idea of using more general membership degrees is rather old, cf. [4].

Another motivation of interval-valued fuzzy sets is the following: If we determine the membership degree of a point p by an interval $(x,y) \in I^2$, $x \le y$, we say that p *belongs* to the fuzzy set with the degree at least x and *does not belong* to it with the degree at least $z = 1-y$; the remainder $1-(x+z) = y-x \ge 0$ represents missing knowledge. This approach works with pairs $(x,z) \in I^2$, $x+z \le 1$, representing *bipolar* information. It is distinguished for philosophical, psychological, and methodological reasons; from the mathematical point of view, the descriptions by the pair (x,y) or (x,z) lead to isomorphic models and results can be easily translated from one context to the other.

We refer to [3, 5, 6, 7, 8] for recent investigation of properties of interval-valued fuzzy sets, operations and measures on them.

2 Standard interval algebra

Here we define the standard algebra of all intervals in I. In the sequel, we shall also consider its subalgebras.

We shall work with the set

$$I_2 = \left\{(x,y) \in I^2 \mid x \le y\right\} . \tag{1}$$

Equipped with the usual pointwise ordering,

$$(x_1,y_1) \le (x_2,y_2) \iff (x_1 \le x_2 \text{ and } y_1 \le y_2) , \tag{2}$$

I_2 becomes a lattice with the least element $\mathbf{0} = (0,0)$ and the greatest element $\mathbf{1} = (1,1)$ and the lattice operations

$$(x_1,y_1) \wedge (x_2,y_2) = (\min(x_1,y_1), \min(x_2,y_2)) , \tag{3}$$
$$(x_1,y_1) \vee (x_2,y_2) = (\max(x_1,y_1), \max(x_2,y_2)) . \tag{4}$$

The sublattice

$$D = \{(x,x) \mid x \in I\} \subset I_2 \tag{5}$$

consists of the elements $(x,x) \in I_2$ which represent real membership degrees corresponding to $x \in I$ in standard fuzzy sets. Thus I_2 represents an extension of $D \cong I$ by elements $(x,y) \in I^2$, $x < y$.

The standard fuzzy negation $'\colon I \to I$, $x' = 1-x$, has a natural extension to the negation $'\colon I_2 \to I_2$,

$$(x,y)' = (y',x') . \tag{6}$$

Thus we have a bounded lattice with an antitone involutive operation $'$. We call this structure the *standard interval algebra* and—with some abuse of notation—we denote it again by I_2.

The standard interval algebra is often considered with additional operations. These are often derived from the *standard MV-algebra* $([0,1],0,\oplus,')$, where $'$ is the standard fuzzy negation and $\oplus$ is the Łukasiewicz t-conorm $a \oplus b = \min(1, a+b)$. Its dual operation is the Łukasiewicz t-norm $a \odot b = (a' \oplus b')' = \max(0, a+b-1)$. These operations are applied to elements of I_2 coordinatewise,

$$(x_1,y_1) \odot (x_2,y_2) = (x_1 \odot x_2, y_1 \odot y_2) , \tag{7}$$
$$(x_1,y_1) \oplus (x_2,y_2) = (x_1 \oplus x_2, y_1 \oplus y_2) . \tag{8}$$

Then $D \subset I_2$ becomes isomorphic to the standard MV-algebra.

Remark 1. *Other t-norms and t-conorms on* I_2, *not originated this way, are studied, e.g., in [3]. We do not deal with them here.*

Another line of research is due to S. Weber [10, 11]. He extends the MV-algebra D to a Girard algebra with operations

$$(x_1,y_1) \sqcap (x_2,y_2) = (x_1 \odot x_2, \max((x_1 \odot y_2), (y_1 \odot x_2))) , \tag{9}$$
$$(x_1,y_1) \sqcup (x_2,y_2) = (\min((x_1 \oplus y_2), (y_1 \oplus x_2)), y_1 \oplus y_2) . \tag{10}$$

Although rather strange, these operations arise as a unique solution of well-motivated algebraic conditions [11]. On D, they

ISBN: 978-989-95079-6-8

coincide with the Łukasiewicz operations and with (7), (8). The lattice reduct $(I_2, \vee, \wedge)$ of the Girard algebra is the same as in the former approach.

As we shall deal only with the lattice operations, the (standard) negation, and with the Łukasiewicz operations on D, we need not deal with the differences between the above two approaches; we study features which are common to both of them.

3 Measures on the interval algebra

In order to study probability of fuzzy events, we need to introduce (probability) measures on fuzzy sets. First of all, we need measures on the algebras of membership degrees; in our case, on the standard interval algebra I_2. Different conditions are imposed in order to extend the notion of probability from classical sets to fuzzy sets and interval-valued fuzzy sets. Here we follow one of the approaches suggested by S. Weber in [11], but analogous conditions are studied in many other papers. We refer to [12] for a comparison of many approaches to measures on collections of fuzzy sets.

Without referring to a specific definition, let us accept that the standard MV-algebra admits one very natural candidate for a measure—the identity. It commutes with the standard negation and satisfies all forms of additivity conditions (see [9] for details). Thus here we assume that the only measure on the standard MV-algebra is the identity. Also S. Weber in [11] studies measures m which are extensions of this measure on D, i.e.,

$$\forall x \in I : m(x,x) = x\,. \tag{M1+}$$

(For simplicity, we write $m(x,x)$ instead of $m((x,x))$.) For all elements $(x,y) \in I_2$, he imposes the following conditions as minimal requirements on so-called *uncertainty measures*:

$$m(\mathbf{0}) = 0\,, \qquad m(\mathbf{1}) = 1\,, \tag{M1}$$

$$A \le B \implies m(A) \le m(B)\,, \tag{M2}$$

$$m\,(A') = 1 - m(A)\,. \tag{M3}$$

(Condition (M1) follows from (M1+).) One of the additional conditions studied in [11] is the *valuation property*:

$$m(A \wedge B) + m(A \vee B) = m(A) + m(B)\,. \tag{V}$$

In this paper, we study mappings $m\colon I_2 \to [0,1]$ satisfying (M1+), (M2), (M3), and (V); we call them *valuation measures*.

3.1 Characterization of valuation measures on the standard interval algebra

Our principal contribution is the following (μ is a $\frac{1}{2}$-*Lipschitz* function if $|\mu(u) - \mu(v)| \le \frac{1}{2}\ |u - v|$ for all u, v):

Theorem 2. *A mapping $m\colon I_2 \to [0,1]$ is a valuation measure if and only if there is a $\frac{1}{2}$-Lipschitz function μ on $\left[0, \frac{1}{2}\right]$ such that m is of the form*

$$m(x,y) = \frac{x+y}{2} + \mu\left(\left|x - \tfrac{1}{2}\right|\right) - \mu\left(\left|y - \tfrac{1}{2}\right|\right)\,. \tag{12}$$

Proof: First, let us suppose that μ is a $\frac{1}{2}$-Lipschitz function and m is given by (12). If $x = y$, we obtain (M1+). Condition (M3) is verified as follows:

$$\begin{aligned} &m(1-y, 1-x)\\ &= \frac{1-y+1-x}{2} + \mu\left(\left|(1-y) - \tfrac{1}{2}\right|\right) - \mu\left(\left|(1-x) - \tfrac{1}{2}\right|\right)\\ &= 1 - \frac{x+y}{2} - \mu\left(\left|x - \tfrac{1}{2}\right|\right) + \mu\left(\left|y - \tfrac{1}{2}\right|\right)\\ &= 1 - m(x,y)\,. \end{aligned} \tag{13}$$

Monotonicity (M2) can be verified in each coordinate separately and follows from the $\frac{1}{2}$-Lipschitz condition. For $x_1 < x_2$, we obtain

$$\begin{aligned} &m(x_2,y) - m(x_1,y)\\ &= \frac{x_2 - x_1}{2} + \mu\left(\left|x_2 - \tfrac{1}{2}\right|\right) - \mu\left(\left|x_1 - \tfrac{1}{2}\right|\right)\\ &\ge \frac{x_2 - x_1}{2} - \frac{1}{2}\left|\left|x_2 - \tfrac{1}{2}\right| - \left|x_1 - \tfrac{1}{2}\right|\right|\\ &\ge \frac{x_2 - x_1}{2} - \frac{|x_2 - x_1|}{2} = 0\,. \end{aligned} \tag{14}$$

For $y_1 < y_2$, we obtain

$$\begin{aligned} &m(x,y_2) - m(x,y_1)\\ &= \frac{y_2 - y_1}{2} + \mu\left(\left|y_2 - \tfrac{1}{2}\right|\right) - \mu\left(\left|y_1 - \tfrac{1}{2}\right|\right)\\ &\ge \frac{y_2 - y_1}{2} - \frac{1}{2}\left|\left|y_2 - \tfrac{1}{2}\right| - \left|y_1 - \tfrac{1}{2}\right|\right|\\ &\ge \frac{y_2 - y_1}{2} - \frac{|y_2 - y_1|}{2} = 0\,. \end{aligned} \tag{15}$$

To check the valuation property (V), take $A = (x_1, y_1)$, $B = (x_2, y_2)$. Notice that, independently of all possible orderings of x_1, x_2 (resp. y_1, y_2), $m(A \wedge B) + m(A \vee B)$ expressed by (12) contains the same summands as $m(A) + m(B)$:

$$\begin{aligned} &m(A \wedge B) + m(A \vee B)\\ &= \frac{x_1 + x_2 + y_1 + y_2}{2}\\ &+ \mu\left(\left|x_1 - \tfrac{1}{2}\right|\right) + \mu\left(\left|x_2 - \tfrac{1}{2}\right|\right)\\ &- \mu\left(\left|y_1 - \tfrac{1}{2}\right|\right) - \mu\left(\left|y_2 - \tfrac{1}{2}\right|\right)\\ &= m(A) + m(B)\,. \end{aligned} \tag{16}$$

Thus m is a valuation measure.

To prove the reverse implication, let us assume that m is a valuation measure. Whenever $x + y = 1$, the element (x,y) is invariant to the negation and, due to (M3), $m(x,y) = \frac{1}{2}$.

We may define

$$\mu(z) = \tfrac{z}{2} + m\left(\tfrac{1}{2} - z, \tfrac{1}{2}\right)\,, \tag{17}$$

in particular,

$$\mu(0) = m\left(\tfrac{1}{2}, \tfrac{1}{2}\right) = \tfrac{1}{2}\,. \tag{18}$$

For $x \le y = \frac{1}{2}$, the right-hand side of (12) is

$$\begin{aligned} \tfrac{x}{2} + \tfrac{1}{4} + \mu\left(\tfrac{1}{2} - x\right) - \mu\left(0\right) &= \tfrac{x}{2} - \tfrac{1}{4} + \mu\left(\tfrac{1}{2} - x\right)\\ &= m\left(x, \tfrac{1}{2}\right) \end{aligned} \tag{19}$$

as desired.

ISBN: 978-989-95079-6-8

For $x \le y < \frac{1}{2}$, we take $A = \left(x, \frac{1}{2}\right)$, $B = (y, y)$. Then $A \wedge B = (x, y)$, $A \vee B = \left(y, \frac{1}{2}\right)$. Using the valuation property, (19), and (M1+), we obtain

$$\begin{aligned} m(x,y) &= m(A \wedge B) = m(A) + m(B) - m(A \vee B) \\ &= m\left(x, \tfrac{1}{2}\right) + m(y,y) - m\left(y, \tfrac{1}{2}\right) \\ &= \tfrac{x}{2} + \mu\left(\tfrac{1}{2} - x\right) + y - \tfrac{y}{2} - \mu\left(\tfrac{1}{2} - y\right) \\ &= \tfrac{x}{2} + \mu\left(\tfrac{1}{2} - x\right) + \tfrac{y}{2} - \mu\left(\tfrac{1}{2} - y\right) \end{aligned} \tag{20}$$

in accordance with (12).

The case $\frac{1}{2} < x \le y$ is obtained by duality.

In the remaining case $x \le \frac{1}{2} < y$, without loss of generality we may suppose that $x + y \le 1$ (the other case is dual). We take $A = (x, y)$, $B = \left(1 - y, \frac{1}{2}\right)$. Then $A \wedge B = \left(x, \frac{1}{2}\right)$, $A \vee B = (1 - y, y)$. Using the valuation property and (19), we obtain

$$\begin{aligned} m(x,y) = m(A) &= m(A \wedge B) + m(A \vee B) - m(B) \\ &= m\left(x, \tfrac{1}{2}\right) + m(1-y, y) - m\left(1-y, \tfrac{1}{2}\right) \\ &= \tfrac{x}{2} + \mu\left(\tfrac{1}{2} - x\right) + \tfrac{1}{2} - \tfrac{1-y}{2} + \mu\left(\tfrac{1}{2} - (1-y)\right) \\ &= \tfrac{x}{2} + \mu\left(\tfrac{1}{2} - x\right) + \tfrac{y}{2} - \mu\left(y - \tfrac{1}{2}\right) . \end{aligned} \tag{21}$$

It remains to prove that μ is $\frac{1}{2}$-Lipschitz. Suppose that $0 \le x_1 \le x_2 \le \frac{1}{2}$. One inequality follows from the monotonicity of m:

$$\begin{aligned} \mu(x_2) - \mu(x_1) &= \tfrac{x_2}{2} + m\left(\tfrac{1}{2} - x_2, \tfrac{1}{2}\right) - \tfrac{x_1}{2} - m\left(\tfrac{1}{2} - x_1, \tfrac{1}{2}\right) \\ &\le \tfrac{x_2 - x_1}{2} . \end{aligned} \tag{22}$$

To prove the second inequality, we take $A = \left(\frac{1}{2} - x_2, \frac{1}{2}\right)$, $B = \left(\frac{1}{2} - x_1, \frac{1}{2} - x_1\right)$. Then $A \wedge B = \left(\frac{1}{2} - x_2, \frac{1}{2} - x_1\right)$, $A \vee B = \left(\frac{1}{2} - x_1, \frac{1}{2}\right)$. We obtain

$$m(A \vee B) - m(A) = m(B) - m(A \wedge B) , \tag{23}$$

$$\begin{aligned} & m\left(\tfrac{1}{2} - x_1, \tfrac{1}{2}\right) - m\left(\tfrac{1}{2} - x_2, \tfrac{1}{2}\right) \\ &= m\left(\tfrac{1}{2} - x_1, \tfrac{1}{2} - x_1\right) - m\left(\tfrac{1}{2} - x_2, \tfrac{1}{2} - x_1\right) \\ &\le m\left(\tfrac{1}{2} - x_1, \tfrac{1}{2} - x_1\right) - m\left(\tfrac{1}{2} - x_2, \tfrac{1}{2} - x_2\right) = x_2 - x_1 , \end{aligned} \tag{24}$$

$$\begin{aligned} & \mu(x_1) - \mu(x_2) \\ &= \tfrac{x_1}{2} + m\left(\tfrac{1}{2} - x_1, \tfrac{1}{2}\right) - \tfrac{x_2}{2} - m\left(\tfrac{1}{2} - x_2, \tfrac{1}{2}\right) \\ &\le \tfrac{x_1 - x_2}{2} + x_2 - x_1 = \tfrac{x_2 - x_1}{2} . \end{aligned} \tag{25}$$

Corollary 3. *Every valuation measure on I_2 is continuous, even 1-Lipschitz.*

Function μ in (12) is determined by m uniquely up to an additive constant. Thus we may choose, e.g., $\mu(0) = 0$, then μ attains values from $\left[-\frac{1}{4}, \frac{1}{4}\right]$.

Example 4. *For $\mu_0 = 0$, we obtain the valuation measure*

$$m_0(x,y) = \frac{x+y}{2} \tag{26}$$

which is additive with respect to $\oplus, \odot$.

The extreme cases are $\mu_1(z) = \frac{z}{2}$, $\mu_{-1}(z) = -\frac{z}{2}$, which lead to

$$m_1(x,y) = \begin{cases} y & \text{if } x \le y \le \frac{1}{2} , \\ \frac{1}{2} & \text{if } x \le \frac{1}{2} < y , \\ x & \text{if } \frac{1}{2} < x \le y , \end{cases} \tag{27}$$

$$m_{-1}(x,y) = \begin{cases} x & \text{if } x \le y \le \frac{1}{2} , \\ x + y - \frac{1}{2} & \text{if } x \le \frac{1}{2} < y , \\ y & \text{if } \frac{1}{2} < x \le y . \end{cases} \tag{28}$$

The choice of function μ is not limited to linear functions in (12).

3.2 Convex structure of the space of valuation measures

In this section, we consider the set $\mathcal{M}$ of all valuation measures on I_2. We consider $\mathcal{M}$ equipped with the product topology inherited from $[0,1]^{I_2} \supseteq \mathcal{M}$. As the defining conditions of valuation measures can be written in the form of equations using only continuous operations, the set $\mathcal{M}$ is closed. It can be easily checked that a convex combination of valuation measures is a valuation measure; thus $\mathcal{M}$ is also convex. Valuation measures μ_1, μ_{-1} of Example 4 are (some of) its extreme points.

Moreover, convex combinations are preserved by (12) in the following sense:

Proposition 5. *Let m_a, resp. m_b, be a valuation measure on I_2 and μ_a, resp. μ_b, the corresponding $\frac{1}{2}$-Lipschitz function satisfying (12). Let $\lambda \in [0,1]$. Then the convex combination $\lambda\, m_a + (1-\lambda)\, m_b$ is a valuation measure which satisfies (12) for $\mu = \lambda\, \mu_a + (1-\lambda)\, \mu_b$.*

4 Generalizations

Theorem 2 described valuation measures on the standard interval algebra. This case is important, but very special. It describes all interval-valued fuzzy subsets of a singleton. Here we extend Theorem 2 to more general cases.

4.1 Valuation measures on the set of all interval-valued fuzzy sets

Now we shall extend Theorem 2 to finite products of algebras isomorphic to the standard interval algebra. These products are isomorphic to the collections of all interval-valued fuzzy subset of a finite set.

Let $n \in \mathbb{N}$. We denote by I_2^n the product of n standard interval algebras, i.e., the cartesian product $\prod_{i=1}^n I_2$ with the lattice operations and negation applied to each coordinate separately.

Theorem 6. *Let $n \in \mathbb{N}$. A mapping $m\colon I_2^n \to [0,1]$ is a valuation measure if and only if there are valuation measures $m_1, \ldots, m_n$ on I_2 and real coefficients $c_1, \ldots, c_n \in [0,1]$ such that $\sum_{i=1}^n c_i = 1$ and m is of the form*

$$m(A_1, \ldots, A_n) = \sum_{i=1}^n c_i\, m_i(A_i) . \tag{29}$$

Proof: Suppose first that m is of the form (29), i.e., a convex combination of valuation measures $m_1, \ldots, m_n$ on I_2. For each $i = 1, \ldots, n$, we denote by π_i the projection on the ith coordinate,

$$\pi_i(A_1, \ldots, A_n) = A_i . \tag{30}$$

ISBN: 978-989-95079-6-8

It is a lattice homomorphism of I_2^n onto I_2. Thus the composition $m_i \circ \pi_i \colon I_2^n \to [0,1]$,

$$(m_i \circ \pi_i)(A_1, \ldots, A_n) = m_i(\pi_i(A_1, \ldots, A_n)) = m_i(A_i)\,, \tag{31}$$

is a valuation measure on I_2^n. As a convex combination of valuation measures, m is also a valuation measure.

Second, suppose that m is a valuation measure on I_2^n. Let $\mathcal{B}$ be the collection of all sharp (=crisp) elements of I_2^n, i.e., all n-tuplets $(A_1, \ldots, A_n)$ such that $A_i \in \{\mathbf{0}, \mathbf{1}\}$ for all $i = 1, \ldots, n$. Then $\mathcal{B}$ is a finite Boolean algebra with n atoms $(\delta_{i1}, \ldots, \delta_{in})$, $i = 1, \ldots, n$, where

$$\delta_{ik} = \begin{cases} \mathbf{1} & \text{if } i = k\,, \\ \mathbf{0} & \text{if } i \neq k\,. \end{cases} \tag{32}$$

This means that $(\delta_{i1}, \ldots, \delta_{in}) = (\mathbf{0}, \ldots, \mathbf{0}, \mathbf{1}, \mathbf{0}, \ldots, \mathbf{0}) \in I_2^n$, where the unit is in the ith coordinate. A valuation measure on a Boolean algebra is a classical probability measure. Thus m it is of the form

$$m(A_1, \ldots, A_n) = \sum_{i=1}^{n} c_i\, s_i(A_1, \ldots, A_n)\,, \tag{33}$$

where $c_i = m(\delta_{i1}, \ldots, \delta_{in})$, $i = 1, \ldots, n$, and $s_1, \ldots, s_n$ are valuation measures on I_2^n such that

$$s_i(\delta_{i1}, \ldots, \delta_{in}) = 1\,. \tag{34}$$

As

$$s_i((\delta_{i1}, \ldots, \delta_{in})') = s_i(\delta'_{i1}, \ldots, \delta'_{in}) = 0\,, \tag{35}$$

we obtain

$$s_i(B_1, \ldots, B_n) = 0 \tag{36}$$

whenever

$$(B_1, \ldots, B_n) \leq (\delta_{i1}, \ldots, \delta_{in})'\,. \tag{37}$$

Thus $s_i(A_1, \ldots, A_n)$ depends only on the ith coordinate, $A_i = \pi_i(A_1, \ldots, A_n)$. We may express s_i in the form $s_i = m_i \circ \pi_i$ for some mapping $m_i \colon I_2 \to [0,1]$. The image of I_2^n under π_i is isomorphic to I_2. As π_i is a homomorphism, m_i must be a valuation measure on I_2. The proof is finished.

The characterization can be extended also to infinite products of the standard interval algebra, thus to collections of interval-valued fuzzy subsets of an infinite universe. This requires valuation measures concentrated not only in coordinates, but also in the points of the underlying Stone space of the Boolean algebra in question. This way, we may also admit subalgebras of products, so that the underlying Boolean algebra is not the whole power set. We leave these extensions for future research.

Theorems 2 and 6 can be combined. We obtain a characterization of valuation measures on the product I_2^n using $\frac{1}{2}$-Lipschitz functions μ_i, $i = 1, \ldots, n$. These functions can be different in different coordinates.

4.2 Extension to subalgebras

S. Weber in [11] investigates valuation measures on finite Girard algebras, mainly on finite subalgebras of I_2 of the form

$$\tilde{\mathbb{C}}_n = \left\{ \left(\tfrac{j}{n}, \tfrac{k}{n}\right) \middle| j, k \in \mathbb{Z},\ 0 \leq j \leq k \leq n \right\}\,, \tag{38}$$

where $n \geq 1$ is a fixed integer. Valuation measures on I_2, when restricted to $\tilde{\mathbb{C}}_n$, become valuation measures on $\tilde{\mathbb{C}}_n$. Conversely, each valuation measure on $\tilde{\mathbb{C}}_n$ admits an extension to a valuation measure on I_2. This holds also for other sublattices of I_2:

Theorem 7. *Let L be a sublattice of I_2. Suppose that $(0,0), (1,1) \in L$ and $(x,y) \in L$ implies (x,x), (y,y), $(y',x') \in L$. A mapping $m \colon L \to [0,1]$ is a valuation measure if and only if it satisfies the condition of Theorem 2.*

The proof follows the pattern of Theorem 2. Analogously, we can also generalize Theorem 6 and obtain a characterization of valuation measures on subalgebras of I_2^n. These subalgebras are isomorphic to general collections of (not all possible) interval-valued fuzzy sets. For this, it is necessary to add assumptions on closedness of these collections under the lattice operations. We leave further details to future research.

5 Conclusions

We studied measures on finite products of the algebra of subintervals of $[0,1]$. We gave a characterization of measures which (besides the obvious properties of monotonicity and boundary conditions)

1. are compatible with the negation,
2. extend the identity on elements of the form (x,x), $x \in [0,1]$,
3. are valuations.

This is aimed as the first step towards the characterization of such measures on collections of interval-valued fuzzy sets.

Our conditions do not represent the only choice, rather a necessary minimum. Other conditions could be added, depending on the application domain. Our general characterization forms a basis for description of measures restricted by additional conditions, too.

Acknowledgment

The author acknowledges the support by the Czech Ministry of Education under project MSM 6840770038.

References

[1] K. T. Atanassov: Intuitionistic fuzzy sets. *Fuzzy Sets Syst.* 20(1) (1986) 87–96.

[2] K. T. Atanassov: *Intuitionistic Fuzzy Sets*. Physica-Verlag, Heidelberg/New York, 1999.

[3] G. Deschrijver, C. Cornelis, and E. E. Kerre: On the representation of intuitionistic fuzzy t-norms and t-conorms. *IEEE Trans Fuzzy Syst.* 12(1) (2004) 45–61.

[4] J. Goguen: L-fuzzy sets. *J. Math. Anal. Appl.* 18(1) (1967) 145–174.

[5] P. Grzegorzewski and E. Mrówka: Probability of intuitionistic fuzzy events. *Soft Methods in Probability, Statistics and Data Analysis*, Physica Verlag, New York, 2002, 105–115.

ISBN: 978-989-95079-6-8

[6] M.G. Karunambigai, P. Rangasamy, K. Atanassov, N. Palaniappan: An intuitionistic fuzzy graph method for finding the shortest paths in networks. In: O. Castillo, P. Melin, O.M. Ross, R.S. Cruz, W. Pedrycz, J. Kacprzyk: *Theoretical Advances and Applications of Fuzzy Logic and Soft Computing*, Proc. IFSA Congress 2007, Springer, Berlin/Heidelberg, 2007, 3–10.

[7] B. Riečan B.: A descriptive definition of the probability on intuitionistic fuzzy sets. *Proc. EUSFLAT'2003*, Zittau, Goerlitz Univ. Appl. Sci., 2003, 263–266.

[8] B. Riečan B.: Representation of probabilities on IFS events. *Advances in Soft Computing, Soft Methodology and Random Information Systems*, Springer, Berlin, 2004, 243–246.

[9] B. Riečan and D. Mundici: Probability on MV-algebras. In: E. Pap (ed.), *Handbook of Measure Theory*, Elsevier Science, Amsterdam, 2002, Chapter 21, 869–910.

[10] S. Weber: Uncertainty measures, decomposability and admissibility. *Fuzzy Sets Syst.* 40 (1991) 395–405.

[11] S. Weber: Uncertainty measures—Problems concerning additivity. *Fuzzy Sets Syst.* 160(3) (2009) 371–383.

[12] S. Weber and E. P. Klement: Fundamentals of a generalized measure theory. In: U. Höhle et al. (eds.), *Mathematics of Fuzzy Sets. Logic, Topology, and Measure Theory*, Kluwer Academic Publishers, Dordrecht, 1999, 633–651.

[13] L. A. Zadeh: Fuzzy Sets: *Inform. Control* 8 (1965) 338–353.

ISBN: 978-989-95079-6-8

Impute Missing Assessments by Opinion Clustering in Multi-Criteria Group Decision Making Problems

Jun Ma[1], Guangquan Zhang[1], Jie Lu[1] Da Ruan[2,3]

1. School of Software Engineering, Faculty of Engineering and Information Technology
University of Technology, Sydney (UTS), PO Box 123, Broadway, NSW 2007, Australia
2. Beglian Nuclear Research Centre (SCK•CEN), 2400, Boeretong, Mol, Belgium
3. Department of Applied Math. & Computer Science, Ghent University, 9000 Gent, Belgium
Email: {junm, zhangg, jielu}@it.uts.edu.au, druan@sckcen.be, da.ruan@ugent.be

Abstract— Multi-criteria group decision-making and evaluation (MCGDME) method typically aggregates information in evaluation tables. For various reasons, evaluation tables (decision matrix) often include missing data that highly affect correct decision-making and evaluation. Most existing imputation methods of missing data are based on statistical features which do not exist in an MCGDME setting. This paper proposes an imputation method of missing data (IMD) in evaluation tables. The IMD method measures the similarity betweent two evaluators' mental models. Evaluators are then classed into several groups based on their similarities by using fuzzy clustering methods. Finally, missing data are imputated under the assumption that the imputated value of missing data does not change the previous clustering results. The proposed IMD method is implemented and tested in two numerical experiments.

Keywords— decision making, missing data, multi-criteria evaluation, opinion clustering, aggregation

1 Introduction

Multi-criteria group decision-making (MCGDM) methods have been widely used in various fields such as new product development in industries [9], clinical diagnose, service strategies selection in business, as well as public economic policy evaluation [11]. In a typical MCGDM process, evaluating alternatives and aggregating assessments based on decision matrices are two crucial processes. Evaluators (or experts) assess alternatives in terms of a set of identified criteria. The assessments are recorded into several decision matrices, also known as decision tables, which are further analyzed and integrated. The aggregation process implements a specific aggregation model to achieve an overall assessment for each alternative and based on which the best alternative(s) is/are selected. Majority of existing research focuses particularly on the problem of how to efficiently integrate heterogeneous quantitative or qualitative information from multiple information sources. Most processes are conducted mainly on the assumption that those obtained decision matrices are intact, i.e., those tables do not include missing values. However, this assumption may not be completely satisfied in real situations.

Various reasons result in missing values in decision matrices. Values may be missed in the course of data collection by device faults or human errors. Evaluators who are lack of relevant background for some criteria or are not willing to provide opinions on some criteria may leave blanks in their decision tables. Security or privacy policies may also prevent evaluators from providing answers to some criteria. An evidently result of missing values reduce the reliability of decision making or evaluation and to produce extraordinary bias towards appropriate decisions. Moreover, missing values may result in invalid evaluation and decision making. Hence, necessary and efficient missing data imputation is required in real applications such as in nuclear material safeguards information management.

Literature indicates missing data processing has a long research history [4–6, 8, 12]. However, most research is mainly conducted on the theory and application of statistics sciences [5]. Existing techniques, such as the LOCF (last observation carried forward), best or worst case imputation, mean imputation, and case deletion [1], generally require the collected data has specific and identifiable features. Such features include, for examples, the data is time-related or follows a particular statistics distribution, as well as has sufficient volume. These features do not hold in a typical MCGDM settings in general. Assessments in a typical MCGDM problem have uniqueness in the sense that each assessment is uniquely linked to a unique alternative, a unique evaluator, and a unique criterion. Moreover, because evaluators are generally not asked or are not willing to repeat same evaluation procedures, the obtained assessment values are always onetime data which is irrelevant to observation time. Therefore, not enough volume of data exists with the same features of a missing value. Besides the uniqueness and small volume of assessments, assessment values are generated through information processing in evaluators' mental model of the given decision problem which cannot be strictly expressed through a formalized and quantified model. Moreover, the links among the selected criteria are generally weak and unclear, because evaluation criteria are always deliberately selected to represent different considerations for alternatives. These features make it very hard to use statistics-based missing data imputation methods to MCGDM problems.

This paper presents an imputation method for missing assessments (IMD) in decision tables for MCGDM. The presented method mainly includes two steps. A fuzzy clustering method is first used to classify evaluators into several groups based on their opinions reflected in their assessments. Then missing assessments are imputed according to those assessments of evaluators belonging to the same group. The rest of the paper is organized as follows. Section 2 reviews the main steps of the presented IMD method. Details of each step is introduced in Section 3. To further illustrate the IMD method,

ISBN: 978-989-95079-6-8

two case studies are conducted. Finally, main results and future works are discussed in Section 5.

2 Overview

2.1 Formalization of typical MCGDM problems

A typical MCGDM problem is composed of a set of alternatives (e.g., action choices, policies), a hierarchy of attributes (i.e., decision criteria), weights associated with criteria, a set of decision matrices (i.e., decision tables), as well as a set of evaluators who present those decision matrices. Three main steps are involved in selections of alternatives [13]: 1) determine the relevant criteria and alternatives; 2) evaluate the relative impacts of alternatives on those criteria; and 3) determine a ranking of each alternatives. Roughly, an MCGDM process can be expressed by a 4-tuple model:

$$\mathcal{M} = (\mathcal{C}, \mathcal{E}, \mathcal{A}, \mathcal{T}), \tag{1}$$

where $\mathcal{C} = \{(c_j, wc_j)|j = 1, 2, \ldots, n\}$ is a set of criteria and their corresponding weights (importance); $\mathcal{E} = \{(e_k, we_k)|k = 1, 2, \ldots, m\}$ is a set of evaluators (experts) and their corresponding weights (reliabilities); $\mathcal{A} = \{a_i|i = 1, 2, \ldots, p\}$ is a set of alternatives (action choices); and $\mathcal{T} = \{T_i = (v^i_{jk})_{n\times m}|i = 1, 2, \ldots, p\}$ is a set of decision matrices. For each a_i, the overall assessment y_i based on the model $\mathcal{M}$ is explicitly expressed as

$$\begin{aligned} y_i &= W_C \circ T_i \diamond W_E^{\mathrm{T}} \\ &= (wc_1, \cdots, wc_n) \circ \begin{pmatrix} v^i_{11} & v^i_{12} & \cdots & v^i_{1m} \\ v^i_{21} & v^i_{22} & \cdots & v^i_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ v^i_{n1} & v^i_{n2} & \cdots & v^i_{nm} \end{pmatrix} \diamond \begin{pmatrix} we_1 \\ we_2 \\ \vdots \\ we_m \end{pmatrix} \end{aligned}$$

where $\circ$ and $\diamond$ are operations in $\mathcal{M}$, W_C is the weight vector for criteria, and W_E the weight vector for evaluators. Once $\circ$ and $\diamond$ are given in explicit forms, the model $\mathcal{M}$ is then determined. For example, let $\circ$ and $\diamond$ be a weighted sum. For any vectors $\boldsymbol{w}$ and $\boldsymbol{v}$

$$\boldsymbol{w} \circ \boldsymbol{v}^{\mathrm{T}} = (w_1, \cdots, w_n) \circ (v_1, \cdots, v_n)^{\mathrm{T}} = \sum_{i=1}^{n} w_i v_i, \tag{2}$$

then y_i can be written as

$$y_i = \sum_{j=1}^{n} \left(wc_i \cdot \sum_{k=1}^{m} v_{jk} we_k \right) = \sum_{k=1}^{m} \left(\sum_{j=1}^{n} wc_j v_{jk} \right) \cdot we_k. \tag{3}$$

Note that most research works on MCGDM focus on how to define these two operations which are mostly studied in the framework of aggregation operators.

In the following, let $T_i = (\boldsymbol{v}^i_1, \boldsymbol{v}^i_2, \cdots, \boldsymbol{v}^i_k)$, where $\boldsymbol{v}^i_k = (v^i_{1k}, v^i_{2k}, \cdots, v^i_{nk})^{\mathrm{T}}$ is called the evaluation vector of evaluator e_k on alternative a_i. T_i is called complete if no missing values in T_i. T_i is called incomplete if at least one evaluation vector includes missing values. For convenience, the following analysis will omit the alternative index i.

2.2 Distance and similarity of evaluator opinions

For each $\boldsymbol{v}_k$ in T, $\boldsymbol{v}_k$ can be seen as an output of evaluator e_k's mental model with the physical nature of alternative as input and denoted by x, i.e., this can be formally described as

$$\boldsymbol{v}_k = f_k(x), \quad k = 1, 2, \ldots, m. \tag{4}$$

Figure 1 illustrates a generalized procedure of obtaining an evaluation table. Notice that the input of each evaluator's mental model is the same, the difference among the obtained evaluation vectors can only be introduced by the difference among evaluators' mental models. Assume that the relationship between evaluators' mental models exists and can be found through their outputs, i.e., the evaluation vectors, then missing values can be approximately imputed with the aid of those relationships. In the light of this idea, the following strategy is adopted to impute missing values. First, a distance measure d between two evaluation vectors is defined, which is then taken as the distance between two evaluators' opinions. Next, the similarity of two evaluators' mental models is induced from the distance. Based on the similarities between each pair of evaluators, a fuzzy clustering method is used to segment evaluators to several clusters. For each of missing values, a predicted value is obtained according to the evaluation vectors provided by evaluators belonging to the same group. This value is the imputed value.

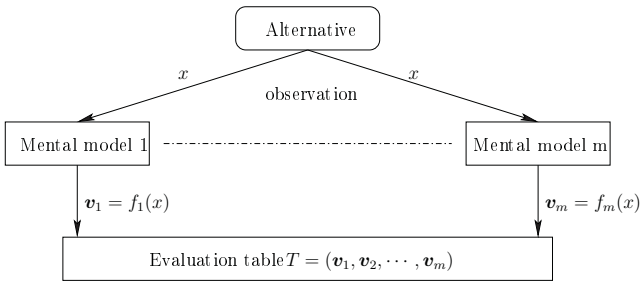


Figure 1: Producing of an evaluation table.

To implement the above-mentioned strategy, we consider some relevant issues. First, the difference of criteria weights is considered when defining a distance measurement. The weight of a criterion indicates the importance degree or relevance degree of that criterion to the decision goal. Some criteria are more important than others. Hence, criteria with bigger weights should contribute more distance to the overall distance between two evaluation vectors. For this reason, we use an aggregation operator to define the overall distance between two evaluation vectors. Another issue is concerned about criteria to that not all missing values must be imputed. For criteria with very small weights, the imputation of missing data is not essential because the bias from appropriate decision induced by discrepancy in those criteria is insignificant. Thus, a parameter α is defined to filter criteria with small weights. Missing data with respect to those filtered criteria are not imputed. Moreover, a second parameter β is used in the segmentation of evaluators, which indicates the at least similarity required when two evaluators are treated with the similar opinion. It can be proved that the bigger the β is, the more obtained clusters and the reduced number of evaluators belong to a certain cluster.

Because the presented strategy adopts two extra parameters α and β, the presented method is thus called an α-β imputation. Next section will give details of the α-β imputation method.

ISBN: 978-989-95079-6-8

3 An α-β imputation method

This section illustrates the details of the α-β imputation method for missing data (assessments, evaluations). For convenience, let all weights for criteria and evaluators be real numbers in the real unit interval [0, 1] and all assessment values be real numbers in [0, 1], too. Without loss of generality, the following discussion will consider the decision matrix for only one alternative.

3.1 Distance between evaluation vectors

An evaluation vector refers to a column in a decision matrix T^i which records the assessments presented by a certain evaluator. As each evaluation vector reflects the opinion of an evaluator, distance of evaluation vectors could be used to reflect the difference between two evaluators' opinions. Based on this idea, we first measure the distance between evaluation vectors. Suppose $\boldsymbol{u}$ and $\boldsymbol{v}$ are two evaluation vectors.

For each criterion c_j ($j = 1, 2, \ldots, n$), let u and v be two assessments from different evaluators about c_j. Then the distance between u and v can be calculated. Suppose the distance is calculated through a distance function d_j which is with respect to c_j and $d_j(u, v)$ is a real number in $[0, 1]$, then d_j can be treated as a function from $[0, 1] \times [0, 1]$ to $[0, 1]$. In the following, let $d_1, d_2, \ldots, d_n$ be the selected distance functions with respect to criteria $c_1, c_2, \ldots, c_n$, respectively without other specifications. Here, we do not require that all d_j be of the same form.

Based on the selected d_j, $j = 1, 2, \ldots, n$, the overall distance between two evaluation vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ can be calculated as follows. Let Agg be an aggregation operator [10] defined from $[0, 1]^n$ to $[0, 1]$ and $\beta_1, \beta_2, \ldots, \beta_n$ be the normalized weights of those criteria, i.e.,

$$\beta_j = \frac{wc_j}{\sum_{j=1}^{n} wc_j}, \quad j = 1, 2, \ldots, n. \tag{5}$$

The distance function $\hat{d}$ between the two evaluation vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ is a function from $[0, 1]^n$ to $[0, 1]$ such that

$$\hat{d}(\boldsymbol{u}, \boldsymbol{v}) = \mathrm{Agg}\left\{(\beta_j, d_j(u_j, v_j))|j = 1, 2, \ldots, n\right\}. \tag{6}$$

Example 3.1 *Suppose Table 1 gives an example setting. Let* Agg *be the weighted-mean. Then the distance between evaluation vectors* $\boldsymbol{u}$ *and* $\boldsymbol{v}$ *is:*

$$\frac{0.12}{1.52} \times 0.75 + \frac{0.47}{1.52} \times 0.11 + \frac{0.93}{1.52} \times 0.57 = 0.44.$$

If Agg *is the OWA operator, then the distance becomes*

$$\frac{0.12}{1.52} \times 0.75 + \frac{0.47}{1.52} \times 0.57 + \frac{0.93}{1.52} \times 0.11 = 0.30.$$

Table 1: An example settings

	d_j	wc_j	$\boldsymbol{u}$	$\boldsymbol{v}$
c_1	$\lvert u - v\rvert$	0.12	0.07	0.82
c_2	$\lvert u - v\rvert^2$	0.47	0.64	0.97
c_3	$\lvert u - v\rvert^3$	0.93	0.93	0.1

Equation (6) indicates the distance between two evaluation vectors that are determined by the selected aggregation operator and the distance functions for each of criteria. It is possible to choose different operators and measures according to the real problem.

3.2 Similarity between two evaluation vectors

Similar to the definition of distance, similarity for individual criterion is first given and then an overall similarity is calculated by a selected aggregation operator.

Let g_j ($j = 1, 2, \ldots, n$) be the selected similarity function with respect to criterion c_j. g_j should satisfy the following two conditions: for any u and v,

$$g_{j_1}(u, v) \geqslant g_{j_2}(u, v) \text{ if } wc_{j_1} \leqslant wc_{j_2}, \quad j_1, j_2 \in \{1, \cdots, n\}; \tag{7}$$

and g_j is a non-increasing function with respect to d_j. Then function g_j is also called a similarity measure with respect to criterion c_j, $j = 1, 2, \ldots, n$. (7) indicates that a more important criterion has less similarity with respect to the same distance between two assessments.

For example, the following functions can be treated as a set of similarity measures.

$$g_j(u, v) = 1 - d_j(u, v)^{\sigma/wc_j}, \quad j = 1, 2, \ldots, n, \tag{8}$$

where $\sigma \in \{wc_j | j = 1, 2, \ldots, n\}$.

In the light of individual similarity measures $\{g_j\}$, an overall similarity $\hat{g}$ between two evaluation vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ is defined, which is a non-increasing function from $[0, 1]^n$ to $[0, 1]$ such that

$$\hat{g}(\boldsymbol{u}, \boldsymbol{v}) = \mathrm{Agg}\{(\beta_j, g_j(u_j, v_j)), j = 1, 2, \ldots, n\}, \tag{9}$$

where Agg is a selected aggregation operators, β_j is the normalized weights of those criteria, $j = 1, 2, \ldots, n$.

In Example 3.1, suppose $g_j(u, v) = 1 - d_j(u, v)$, then the similarities for the weighted-sum aggregation and OWA aggregation are 0.56 and 0.70, respectively. Obviously $\hat{g}(\boldsymbol{u}, \boldsymbol{v}) = 1 - \hat{d}(\boldsymbol{u}, \boldsymbol{v})$. However, this does not hold if we change the relationship between g_j and d_j for any j. This indicates that in real problems we can select the appropriate operators to define the similarity. Moreover, we should notice that $\hat{g}(\boldsymbol{u}, \boldsymbol{v})$ is given through individual distance d_j and individual similarity g_j, $j = 1, 2, \ldots, n$, rather than the overall distance $\hat{d}(\boldsymbol{u}, \boldsymbol{v})$. The reason for separately defining them is to provide more flexibilities to consider the individual influence of criteria.

3.3 Evaluator segmentation

The similarity between two evaluation vectors reflects to what extend two evaluators' opinions are similar, which is used to define the similarity between two evaluators' mental models. Evaluators, thus, can be clustered into several groups based on the similarity of their mental models.

Suppose the similarities between any two pair of evaluators are calculated and recorded in a similarity matrix $M = (s_{jk})_{m \times m}$ where s_{jk} is the similarity between evaluators e_j and e_k. To segment evaluators based on M, we use the fuzzy clustering technique here. Let $\bar{M}$ be the transitive closure of M obtained by

$$\bar{M} = \bigcup_{i=1} M^i, \tag{10}$$

where M^i is the i-th power of M by the max-min composition [11]. Therefore, a clustering result can be obtained from the

ISBN: 978-989-95079-6-8

matrix $\bar{M}$. For example, suppose the matrix $\bar{M}$ is

$$\bar{M} = \begin{pmatrix} 1.0 & 0.6 & 0.8 & 0.6 \\ 0.6 & 1.0 & 0.6 & 0.7 \\ 0.8 & 0.6 & 1.0 & 0.6 \\ 0.6 & 0.7 & 0.6 & 1.0 \end{pmatrix}. \tag{11}$$

Then the clustering result is shown in the following figure.

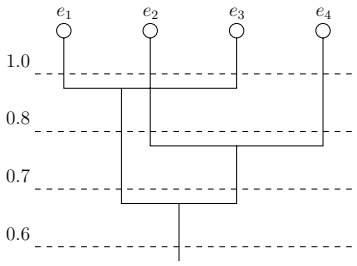


Figure 2: An evaluator clustering results.

From Figure 2, the segmentation results may change with different levels of similarity. For instance, the segmentation result at $\hat{g} = 0.8$ is $\{e_1, e_3\}$, $\{e_2\}$, and $\{e_4\}$; while all evaluators belong to the unique group at $\hat{g} = 0.6$. Hence, a parameter β is used here to indicate at which level the segmentation result is concerned about. In the following let the segmentation result with respect to the given β be $\Gamma(\beta)$.

3.4 Imputation of missing data

To impute missing data, we consider relevant issues and illustrate corresponding solutions here.

First, some missing data need not to be imputed when the associated criteria have infinitesimal effects to the decision G. Therefore, a parameter α is introduced to filter criteria with smaller weights than α. For those criteria, missing data are not imputed.

Next, the evaluator segmentation in Section 3.3 is conducted under the assumption that the decision matrix is a complete one. When the decision matrix is incomplete, we make the following adjustment. Without loss of generality, set

$$T = \begin{pmatrix} A \\ B \end{pmatrix}, \tag{12}$$

where A is a sub-matrix of T which is complete and B is the remainder of T in which each row contains at least one missing data. The evaluator segmentation is obtained based on A, i.e., taking A as the basis to calculate distance and similarity between evaluation vectors. Values in B are used to generate imputations of missing values. Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be two evaluation vectors and divide them into two parts according to A, B,

$$\boldsymbol{u} = \begin{pmatrix} u_A \\ u_B \end{pmatrix}, \qquad \boldsymbol{v} = \begin{pmatrix} v_A \\ v_B \end{pmatrix}.$$

Suppose $u_j \in u_B$ is a missing value and $v_j \in v_B$ exists. To estimate u_j from v_j, we follow the principle that adding missing value does not change the established similarity between the two evaluation vectors. That means if the established similarity between u_A and v_A is g, then the similarity between $u_{A\cup\{u_j\}}$ and $v_{A\cup\{v_j\}}$ is still g after imputation the missing value u_j. Formally, this can be expressed as the following equation

$$\begin{aligned} &\mathrm{Agg}\Big\{(\beta'_l, g_l(u_l, v_l)), l \in \delta(A)\cup\{j\}\Big\} \\ &\quad = \mathrm{Agg}\Big\{(\beta_l, g_l(u_l, v_l)), l \in \delta(A)\Big\}, \end{aligned} \tag{13}$$

where $\delta(A) = \{l | u_l \in A\}$, β' is the normalized weight after adding the criterion c_j and is calculated by

$$\beta'_l = \frac{wc_l}{\sum_{r\in\delta(A)\cup\{j\}} wc_r} = \frac{wc_l}{wc_j + \sum_{r\in\delta(A)} wc_r}. \tag{14}$$

Notice in (13) only u_j is unknown, therefore, the solution(s) of the equation is/are possible estimation(s) of the missing value u_j in $\boldsymbol{u}$ in terms of $\boldsymbol{v}$.

For example, suppose the Agg is the weighted mean and $g_j(\boldsymbol{u}, \boldsymbol{v}) = 1 - d_j(\boldsymbol{u}, \boldsymbol{v})$, $j = 1, 2, \ldots, n$. Then by (13), we have

$$\begin{aligned} &\mathrm{Agg}\Big\{(\beta'_l, g_l(u_l, v_l)), l \in \delta(A)\cup\{j\}\Big\} \\ &\quad = \mathrm{Agg}\Big\{(\beta_l, g_l(u_l, v_l)), l \in \delta(A)\Big\}, \end{aligned}$$

Hence,

$$\sum_{l\in\delta A\cup\{j\}} \beta'_l g_l(u_l, v_l) = \sum_{l\in\delta(A)} \beta_l g_l(u_l, v_l),$$

$$\sum_{l\in\delta(A)\cup\{j\}} \frac{wc_l \cdot g_l(u_l, v_l)}{wc_j + \sum_{r\in\delta(A)} wc_r} = \sum_{l\in\delta(A)} \frac{wc_l \cdot g_l(u_l, v_l)}{\sum_{r\in\delta(A)} wc_r},$$

$$1 - \sum_{l\in\delta(A)\cup\{j\}} \frac{wc_l \cdot d_l(u_l, v_l)}{wc_j + \sum_{r\in\delta(A)} wc_r} = 1 - \sum_{l\in\delta(A)} \frac{wc_l \cdot d_l(u_l, v_l)}{\sum_{r\in\delta(A)} wc_r}$$

So,

$$d_j(u_j, v_j) = \frac{\sum_{l\in\delta(A)} wc_l \cdot d_l(u_l, v_l)}{\sum_{r\in\delta(A)} wc_r} = \hat{d}(u_A, v_A). \tag{15}$$

Equation (15) indicates that u_j can be obtained simply by the d_j and $\hat{d}$ under the above assumptions. Notice that (15) is obtained from g_j and $\hat{g}$, this equation also indicates that the estimation of u_j is determined by the selection of g_j and $\hat{g}$. Hence given different definitions of them, the relationship in (15) may vary.

3.5 Discussions

To find solution of (15), we notice that the equation may produce more than one estimation of the missing value. The question, then, raises naturally that how to determine a unique one. This paper partly resolves this problem through calculating the change $\tilde{d}$ of overall distance induced by an obtained estimation. The estimation that produces the least $\tilde{d}$ will be taken as the imputation of missing data. The overall distance change $\tilde{d}$ is defined as

$$\tilde{d}(u_j) = \sum_{v\in\Gamma(u)\setminus\{u\}} d_j(u_j, v_j), \tag{16}$$

where $\Gamma(u)$ is the cluster of evaluation vectors in which u belongs to. By (16), the best estimation is the one which produces the least distance sum between itself to other evaluation

ISBN: 978-989-95079-6-8

vectors belonging to the same cluster. Examples in the following section have shown that this method can reduce some estimations although they cannot grantee the uniqueness of estimation. Hence, more work is being done to provide other strategy to resolve this problem.

Another question raising from the above imputation procedure is the obtained evaluator segmentation may not be correct because the segmentation is conducted based on A only. Accounting for this problem, we extend the above procedure as follows. For any $\boldsymbol{u}$ and $\boldsymbol{v}$, the distance $\hat{d}(\boldsymbol{u}, \boldsymbol{v})$ and the similarity $\hat{g}(\boldsymbol{u}, \boldsymbol{v})$ are calculated in terms of criteria on which both $\boldsymbol{u}$ and $\boldsymbol{v}$ are not missed rather than limited to criteria corresponding to A. That means the obtained distance and similarity between $\boldsymbol{u}$ and $\boldsymbol{v}$ are the most possible values from the current incomplete decision table. An experiment of the extended process is given in the next section.

4 Experiments and observations

The presented α-β imputation method of missing data is programmed in the Java^{TM} programming language running on a Linux system.

4.1 Experiment 1

Experiment 1 is conducted on a complete decision table with 16 criteria and 5 evaluators. Missing data are randomly selected from the decision table.

To impute the missing data, we set the distance measure for each criteria as $d_j(u_j, v_j) = |u_j - v_j|$ $(j = 1, 2, \ldots, 16)$ and the similarity measure for each criterion as $g_j(u_j, v_j) = 1 - d_j(u_j, v_j)$, the aggregation operators used in calculating overall distance and similarity are the same, say the weighted mean; and set the two parameters $\alpha = 0.2$ and $\beta = 0.6$.

Table 2: The original evaluation table in which missing data are marked by bold font.

wc_j	e_1	e_2	e_3	e_4	e_5
0.262	0.877	0.035	0.88	0.772	0.941
0.334	0.977	0.458	0.903	0.898	0.227
0.823	0.845	0.065	0.894	0.122	0.555
0.216	0.532	0.630	**0.676**	0.353	0.858
0.803	0.862	0.073	0.183	0.618	0.784
0.227	0.797	0.046	**0.721**	0.551	0.950
0.794	0.159	0.995	0.621	0.808	0.307
0.811	0.227	0.936	0.994	0.973	0.134
0.945	0.304	0.901	0.636	0.225	**0.927**
0.892	0.967	0.717	0.851	0.835	**0.035**
0.344	0.641	**0.336**	0.666	0.281	0.598
0.016	**0.475**	0.915	0.743	0.876	0.651
0.235	0.026	0.999	0.501	0.431	0.586
0.905	0.307	0.371	0.757	0.832	0.66
0.316	0.541	0.003	0.438	0.113	0.275
0.068	0.042	0.27	0.929	0.675	0.206

By the given distance and similarity measure, we get the distance matrix $\mathcal{D}$ and the similarity matrix $\mathcal{M}$ as follows:

$$\mathcal{D} = \begin{pmatrix} 0 & 0.571 & 0.35 & 0.392 & 0.233 \\ 0.571 & 0 & 0.348 & 0.332 & 0.55 \\ 0.35 & 0.348 & 0 & 0.255 & 0.382 \\ 0.392 & 0.332 & 0.255 & 0 & 0.394 \\ 0.233 & 0.55 & 0.382 & 0.394 & 0 \end{pmatrix} \tag{17}$$

and

$$\mathcal{M} = \begin{pmatrix} 1 & 0.429 & 0.65 & 0.608 & 0.767 \\ 0.429 & 1 & 0.652 & 0.668 & 0.45 \\ 0.65 & 0.652 & 1 & 0.745 & 0.618 \\ 0.608 & 0.668 & 0.745 & 1 & 0.606 \\ 0.767 & 0.45 & 0.618 & 0.606 & 1 \end{pmatrix}. \tag{18}$$

By the fuzzy clustering method, we get the following clustering result.

$$\bar{\mathcal{M}} = \begin{pmatrix} 1 & 0.65 & 0.65 & 0.65 & 0.767 \\ 0.65 & 1 & 0.668 & 0.668 & 0.65 \\ 0.65 & 0.668 & 1 & 0.745 & 0.65 \\ 0.65 & 0.668 & 0.745 & 1 & 0.65 \\ 0.767 & 0.65 & 0.65 & 0.65 & 1 \end{pmatrix} \tag{19}$$

i.e., we have only one class $\{e_1, e_2, e_3, e_4, e_5\}$. By the imputation method in Section 3.4, the following result is obtained (shown in Table 3).

Table 3: Decision table after imputation of missing data (bold number for estimations and # for still missing data).

wc_j	e_1	e_2	e_3	e_4	e_5
0.262	0.877	0.035	0.88	0.772	0.941
0.334	0.977	0.458	0.903	0.898	0.227
0.823	0.845	0.065	0.894	0.122	0.555
0.216	0.532	0.63	**0.608**	0.353	0.858
0.803	0.862	0.073	0.183	0.618	0.784
0.227	0.797	0.046	**0.296**	0.551	0.95
0.794	0.159	0.995	0.621	0.808	0.307
0.811	0.227	0.936	0.994	0.973	0.134
0.945	0.304	0.901	0.636	0.225	**0.071**
0.892	0.967	0.717	0.851	0.835	**0.734**
0.344	0.641	#	0.666	0.281	0.598
0.016	**0.418**	0.915	0.743	0.876	0.651
0.235	0.026	0.999	0.501	0.431	0.586
0.905	0.307	0.371	0.757	0.832	0.66
0.316	0.541	0.003	0.438	0.113	0.275
0.068	0.042	0.27	0.929	0.675	0.206

4.2 Experiment 2

In the presented imputation method of missing data, aggregation operators are used for calculation of overall distance and overall similarity of two evaluation vectors. Different aggregation operators may lead to different clustering results and, in turn, may result in different imputations. Experiment 2 compares three widely-used aggregation operators, i.e., the weighted mean (Agg_w); the arithmetic mean (Agg_a)

$$\mathrm{Agg}_a(x_1, x_2, \cdots, x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i; \tag{20}$$

and the OWA operator (Agg_o)

$$\mathrm{Agg}_o(x_1, x_2, \cdots, x_n) = \sum_{i=1}^{n} w_i x_{\sigma(j)}, \tag{21}$$

where σ is a permutation such that $x_{\sigma(1)} \geqslant x_{\sigma(2)} \geqslant \cdots \geqslant x_{\sigma n}$. To conduct this experiment, we change the distance measure and the similarity measure for each individual criterion,

ISBN: 978-989-95079-6-8

i.e.,

$$d_j(u_j, v_j) = |u_j - v_j|, \quad g_j(u_j, v_j) = 1 - d_j(u_j, v_j).$$

The original evaluation table is still the one shown in Table 2. The comparison results are shown in Table 4.

Table 4: Comparison results.

missing data		original value	imputations		
row	col		Agg_w	Agg_o	Agg_a
3	2	0.676	0.608	0.527	0.581
5	2	0.721	0.296	0.756	0.323
8	4	0.927	0.071	/	0.557
9	4	0.035	0.734	/	0.714
10	1	0.336	/	/	/
11	0	0.475	0.418	/	0.398

From the comparison, we can observe that different aggregation operators may produce different imputations. Even more, some aggregation operators may not produce estimation of missing data. Moreover, the bias between an imputation and an original value is bigger for some criteria. The independence nature of the data and the unsupervised nature of the presented method are two possible reasons for these results.

5 Conclusions and Future Works

Imputation of missing data is a vital step in decision making and evaluation problems. Due to various reasons, decision tables are often including missing data. Therefore, how to efficient impute the missed data will affect correct decision making and evaluation. However, few work is reported on imputation of missing data for decision tables. In this paper, an α-β imputation method of missing data (IMD) in a decision table is proposed. The IMD method estimates missing data based on the similarity between evaluators' mental models. First, the IMD method uses an overall distance to measure the similarity of two evaluators mental models through their evaluation vectors. Then a fuzzy clustering method is used to segment evaluators based on the similarity of their mental models. Finally, missing data are imputed by the clustering results under the assumption that an imputed data should not change the clustering results. Two numerical experiments are conducted to test the proposed IMD method. Experiments indicate the IMD method can efficiently estimate some missing data. Moreover, those experiments also show that selecting different aggregation operators, distance measure, and similarity measure may change the imputed values. However, from these experiments, we notice that some estimations of missing data have bigger bias from the originals due to the independence nature of the data and lack of backgrounds as well as other possible reasons. Hence, more work still need to be done such as how to improve the accuracy of an estimation of missing data and how to combine relevant background to the imputation. Now, some related work is undertaking.

Currently, we notice that another missing value processing problem in fuzzy preference relations for group decision-making problem has been discussed in [2, 3]. Method based on additive consistency has been discussed [7]. Although the missing value in that method is different from the one discussed in the presented work, ideas in those work can be introduced into the presented work to improve the efficiency and accuracy. Some work is conducting.

Acknowledgment

The work presented in this paper was supported by Australian Research Council (ARC) under Discovery Projects DP0559213 and DP0880739.

The authors sincerely appreciate the reviewers comments and suggestions to improve the presented work.

References

[1] Points to consider on missing data, 2001. Committee for Proprietary Medicinal Products (CPMP),The European Agency for the Evaluation of Medicinal Products.

[2] S. Alonso, F. J. Cabrerizo, F. Chiclana, F. Herrera, and E. Herrera-Viedma. Group decision making with incomplete fuzzy linguistic preference relations. *International Journal of Intelligent Systems*, 24(2):201–222, 2009.

[3] S. Alonso, F. Herrera-Viedma, F. Chiclana, and F. Herrera. Individual and social strategies to deal with ignorance situations in multi-person decision making. *International Journal of Information Technology and Decision Making*, 2009. in press.

[4] Q. Chen, J. G. Ibrahim, M.-H. Chen, and P. Senchaudhuri. Theory and inference for regression models with missing responses and covariates. *Journal of Multivariate Analysis*, 99:1302–1331, 2008.

[5] A. Farhangfar, L. Kurgan, and J. Dy. Impact of imputation of missing values on classification error for discrete data. *Pattern Recognition*, 41:3692–3705, 2008.

[6] O. Harel. Inferences on missing information under multiple imputation and two-stage multiple imputation. *Statistical Methodology*, 4(1):75–89, January 2007.

[7] E. Herrera-Viedma, F. Chiclana, F. Herrera, and S. Alonso. Group decision-making model with incomplete fuzzy preference relations based on additive consistency. *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, 37(1):176–189, 2007.

[8] J. Liesiö, P. Mild, and A. Salo. Robust portfolio modeling with incomplete cost information and project interdependencies. *European Journal of Operational Research*, 190:679–695, 2008.

[9] J. Lu, G. Zhang, D. Ruan, and F. Wu. *Multi-Objective Group Decision Making – Methods, Software and Applications with Fuzzy Set Technology*. Imperial College Press, London, 2007.

[10] R. Mesiar, A. Kolesárová, T. Calvo, and M. Komorníková. A review of aggregation functions. In H. Bustince, F. Herrera, and J. Montero, editors, *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models*, volume 220 of *Studies in Fuzziness and Soft Computing*, pages 121–144. Springer Berlin/Heidelberg, 2008.

[11] G. Munda. A conflict analysis approach for illuminating distributional issues in sustainability policy. *European Journal of Operational Research*, 194(1):307–322, 2009.

[12] J. Siddique and T. R. Belin. Using an approximate bayesian bootstrap to multiply impute nonignorable missing data. *Computational Statistics & Data Analysis*, 53:405–415, 2008.

[13] E. Triantaphyllou. *Multi-Criteria Decision Making Methods: A Comparative Study*. Kluwer Academic Publishers, Dordrecht/Boston/London, 2000.

ISBN: 978-989-95079-6-8

Application of Zadeh's Impossibility Principle to Approximate Explanation

Vesa A. Niskanen

Dept. of Economics & Management, University of Helsinki, PO Box 28, 00014 Helsinki, Finland
Email : vesa.a.niskanen@helsinki.fi

Abstract: *We consider application of Zadeh's impossibility principle and extended logic FLe to approximate scientific explanation from the standpoint of the philosophy of science. First, traditional explanation models are sketched. Second, Zadeh's principle is considered in the broader context which includes reasoning and theory formation. Third, application ideas for using the impossibility principle to approximate explanation are provided. We focus on the role of approximate reasoning, approximate hypothesis assessment and approximate probability in explanation. It seems that approximate explanation can be usable when imprecise entities are involved in the conduct of inquiry.*

Keywords: *approximate explanation, approximate reasoning, impossibility principle.*

1 Introduction

Scientific explanations make our objects of research intelligible for us [28]. In practice this means that we aim to find the causes, functions or purposes of the phenomena under study. An explanation constitutes of two parts, the phenomenon or problem to be explained (explanandum) and our explanation for it (explanans). Explanations play an important role in theory formation and hypothesis assessment [12,17,18,28,29].

The study on the causes of the phenomena is a central issue in explanation. If we study the causes, Aristotle already suggested four types of causes, viz. material, formal, effective and final causes. Today, only two of these are usually applied in the conduct of inquiry. The natural sciences focus on the effective causes, whereas in the human sciences final causes are also considered [18].

Today, the mainstream seems to aim at precise explanation models, but thanks for the fuzzy systems and Zadeh's impossibility principle, it seems that we can also construct corresponding approximate models in the future.

Below we consider the prospects for applying the impossibility principle and other Zadeh's recent ideas to scientific explanation from the standpoint of the philosophy of science. Due to our philosophical approach, we mainly provide some general ideas for applying these methods. Lack of space puts also restrictions on our examination. Section 2 briefly presents the prevailing explanation models. In section 3 we integrate the impossibility principle with reasoning, theory formation, hypothesis testing and explanation. Section 4 considers aspects of approximate explanation, and section 5 concludes our examination.

2 Traditional explanation models

The pragmatic approach to explanation, which seems widely-used today, considers such phenomena as facts, regularities and occurrences. Our task is to provide the cause, origin, history, function or goal of a given phenomenon. In practice we provide such questions as "why?", "how?", "what is the purpose for?", "what is the reason for?" and "what is the function of?" [18,27,29].

The positivistic tradition, in turn, has emphasized syntactical and semantical aspects of explanation in which case concept analysis and logical structures are essential [18,27].

Another categorization of explanation stems from the causes. If we attempt to find effective causes of a phenomenon, we consider which agents, factors or events have initiated the phenomena under study. The well-known example of this idea is the subsumption model in the natural sciences [12,13,17].

The final causes, in turn, are those goals, purposes, intentions or motives of agents, in particular those of the human beings, which guide their behavior. Aristotle's and von Wright's practical syllogism models [29] apply this idea. The subsumption approach mainly provides why?- and how?-questions, and the latter tradition usually asks what is the purpose for or what is the reason for something.

We usually search for effective causes in the inanimated world, but in the animated world the problem arises whether we should also search for final causes. From the philosophical standpoint, the former approach mainly stems from the positivistic tradition and Marxism and the latter, which studies both of these causes, is known as the Geisteswissenschaft approach (hermeneutics, phenomenology etc.). In [29] these competitive approaches are referred to as the Galilean and Aristotelian traditions, respectively, and we use these terms below.

The explanation of the behavior of the human being has been a long-lasting controversy between these two traditions. The Galilean tradition assumes that our behavior is based on our stimulus-response functions and such feedback systems which are analogous to homeostats. In this framework we can even deal with *goal-directed* behavior of beings.

The Aristotelian tradition presupposes that the human beings are *goal-oriented* agents and thus we must also study their aims, intentions, motives and other underlying causes of behavior. Our goals can be both conscious and subconscious, and intentionality is usually related to the former category. A classical example within this tradition is that we should explain (inanimate) nature but understand history, i.e., historical research also takes into account the intentional factors of beings [2,7,15,26].

If we consider the nature of causes on the time axis, the distinction between the Galilean and Aristotelian explanations essentially means that in the former case the cause precedes and in the latter case it follows the effect. Another way to state this is that we use causal and teleological explanations, respectively.

ISBN: 978-989-95079-6-8

For example, if we state in the first day of the IFSA '09 Congress that Lotfi Zadeh is present because he was requested to deliver the keynote speech, we use the Galilean explanation. If, in turn, we state that Lotfi Zadeh is present in order to deliver the keynote speech, the Aristotelian approach is adopted. Naturally, we can often use these explanations simultaneously.

Hence, in this sense we can establish that

1. Requested for speech (causal explanans) → **Zadeh is present in the IFSA '09 (explanandum).**
2. **Zadeh is present in the IFSA '09 (explanandum)** ← In order to deliver speech (teleological explanans):

Other models for explanation are also available. For example, the genetic explanations aim to discover causes by examining the history or chronological order of events of a phenomenon. In the historical research and forensic sciences, inter alia, abduction is usually applied in this context.

The statistical and probabilistic explanations, in turn, are based on statistical results and probabilistic models, respectively. An example of the former type would be the reasoning that most of the participants in the IFSA '07 Congress were males because about 90 % of the researchers in the fuzzy community are males. In the latter case we could provide an explanation that if we are tossing a coin, about 50 % of the outcomes are heads because the frequency probability for this event is .5. In particular in the human sciences these explanation models are relevant because the human behavior is usually indeterministic by nature.

The traditional explanation models aim to operate with precise entities and traditional deduction, and thus certain problems can arise. First, they usually apply bivalent logic and this approach can yield overly rough models. Second, if mathematical expressions are used, they can be oversimplified or too idealized for modeling the corresponding real-world phenomena. Finally, this approach lacks an appropriate uniform language which can be used in the context of imprecise entities. Fuzzy systems, in turn, can provide some resolutions to these problems.

3 Impossibility principle and FLe

Recently, Zadeh has established the principles of the extended fuzzy logic, *FLe* (a.k.a. FL+), which is a combination of "traditional" provable and precisiated fuzzy logic, *FLp*, as well as a novel meta-level unprecisiated fuzzy logic, *Flu* [31,38]. He maintains that in the *FLp* the objects of discourse and analysis can be imprecise, uncertain, unreliable, incomplete or partially true, whereas the results of reasoning, deduction and computation are expected to be provably valid. In the *Flu*, in turn, membership functions and generalized constraints are not specified, and they are a matter of perception rather than measurement.

In our framework below this means that we can apply both traditional bivalent-based and novel approximate validity, definitions, axioms, theories and explanations, inter alia. In the former case we thus operate with precise theorems, classical deducibility, syllogisms and formal logic, whereas the latter is related to "f-entities" within informal and approximate reasoning and incomplete information (Fig. 1).

For example [31, p. 2],

"A simple example of a f-theorem in f-geometry is: f-medians of f-triangle are f-concurrent. This f-theorem can be f-proved by fuzzification of the familiar proof of the crisp version of the theorem."

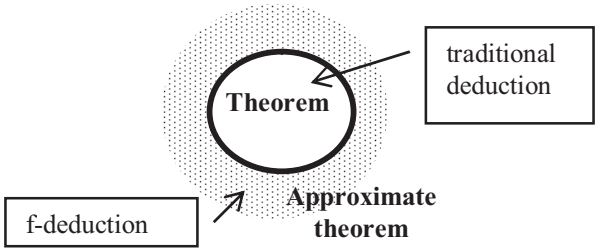


Figure 1. Precise and approximate theorem by applying the *FLe*.

F-validity and f-theoremhood in the *Flu* are examples of Zadeh's Impossibility Principle [31]. This principle informally states that in an environment of imprecision, uncertainty, incompleteness of information, conflicting goals and partiality of truth, p-validity is not, in general, an achievable objective.

The fuzzy modus ponens is the well-known application of this principle in semantic validity, i.e., we can replace the traditional syllogism [8,11,14,19]

A
If A, then B
Thus, B

with its approximate counterpart

A'
If A, then B
Thus B'

in which A≈A' and B≈B'. In the latter case we can thus draw approximate conclusions, i.e., if A' resembles the antecedent A, the conclusion B' resembles the consequent B. In other words, the approximate conclusion is in the proximity to its true counterpart. F-validity is the corner stone in the *FLe* and it is also essential in approximate theory formation and explanation.

In [21,22,23], Zadeh's *FLe* is considered from the standpoint of reasoning and theory formation and it was assumed that the approximation stems essentially from the former. We assume below that if our statements are true, it means that they are identical with their true counterparts. The non-true statements, in turn, are more or less distinct from their true counterparts, or in other words, their distances between their true counterparts are non-zero.

For example, given that John is young (true counterpart), we can reason that the statement

"John is young"

is true. On the other hand, "John is fairly young" is non-true because the predicates "is young" and "is fairly young" are dissimilar.

In fact, we can thus operate with a f-similarity relation of the *Flu,* and then the truth values are specified according

ISBN: 978-989-95079-6-8

to this operation. The higher this similarity, the more true the statement in consideration [23].

Then it seems plausible that, for example, the truth value of the statement

"John is fairly young", provided that he is young

is fairly true [23].

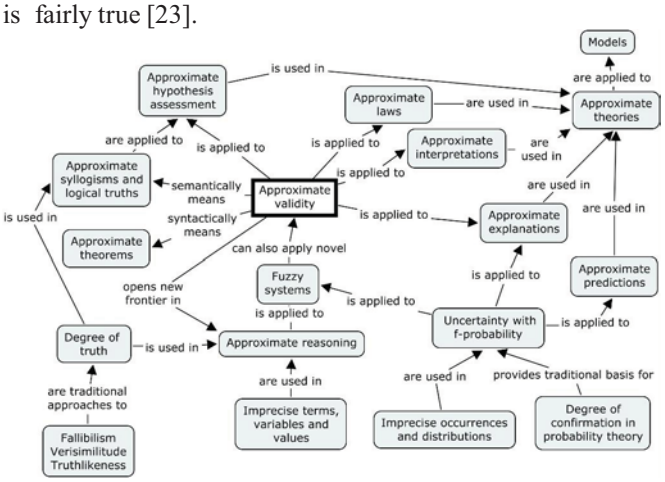


Figure 2. The role of Zadeh's *FLe* in the conduct of inquiry.

We thus assume that approximate theorems are in the neighborhood of their true counterparts. We can also assume that in theory formation approximate laws and theories are in the proximity to their true counterparts and approximate explanations are close to the corresponding true explanations. Figure 2 depicts these ideas at a general level when we establish that our examination on the *FLe* is mainly based approximate validity.

In the case of theory formation, for example, the constituent or "law" of one physical theory states that the boiling point of water is 100 ºC at sea level, but we can also apply it as an approximate law in the low altitudes in general, i.e., we can state that the boiling point of water is *approximately 100 ºC close to* sea level [19,22].

In addition, within the traditional approach to the theory formation and model construction we only use their approximations and idealizations on some occasions. In the former case the degree of truth of a theory or model often increases when we replace the approximate theory, law or model with its less approximate counterpart [18,19]. Hence, in this respect the law

the boiling point of water is 100 ºC at sea level

is a more true than

the boiling point of water is 100 ºC.

Idealizations, in turn, are typical when mathematical or statistical models are used. For example, in statistics such ideal curves as the Gaussian curve are used when the empirical data sets, the distributions of which are only more or less similar to their mathematical counterparts, are examined. Thus, we actually draw statistical conclusions on approximate grounds (c.f. also section 4).

Popper has already considered the foregoing aspects in his theory of verisimilitude from the bivalent standpoint [24,25]. He presupposes that we aim to formulate true theories (cognitivism) but in practice our theories can be non-true. If we perform successful research, our theories will approach their true counterparts and thus they are always corrigible by nature. The similar idea is applied in Peirce's fallibilism [17,18], and in general, in scientific realism (Peirce, Lenin, Popper, Hempel). Niiniluoto [20] has considered approximate theories and explanations within his theory of truthlikeness, but he has also applied bivalent logic in this context. In fact, if we replace bivalent logic with *FLe*, our approach is fairly close to his ideas [23].

4 Approximate explanation

As was stated above, our examination is based on the idea of approximate reasoning in the *FLe*. Thus, in an approximate explanation model the explanans is in the neighborhood of its true counterpart. In practice we now assess the degree of similarity between our explanans and its true counterpart. The higher the degree of similarity, the higher the degree of truth for our explanans (Fig. 3).

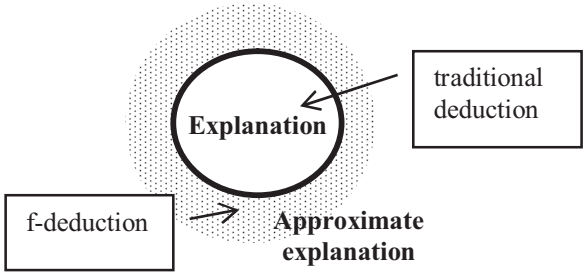


Figure 3. Precise and approximate explanation by applying the *FLe*.

Hence, we have to consider again which logical tools are appropriate to us. As above, within simple statements we assess the similarity between the predicates, i.e.,

truth(John is fairly young), provided that he is young,

is based on the similarity between "young" and "fairly young".

The more complex statements can also be examined within the *FLe* because we can apply the fuzzy similarity measure to compound and quantified statements [31-38]. Then, for example, we should be able to assess such truth values as

truth(John is fairly young and very decent),

provided that he is young and decent.

Another example is the assessment of such quantified expressions as

truth(all Swedes are fairly tall),

provided that most Swedes are tall.

In the former case some concrete resolutions are already available in the *FLe*, whereas fuzzy quantification still awaits usable truth evaluation results. Approximate quantification is problematic because then we operate at two levels, with approximate quantifiers and (compound) statements. Application of the extension principle and Zadeh's idea on

ISBN: 978-989-95079-6-8

f-transformation [38] can be a possible resolution for these, and one method is suggested in [23]. Naturally, these approaches are context-dependent by nature. Hence, more concrete results are expected for assessing the truth values and validity within the *FLe*.

In the probability and statistical explanation models mentioned above we can apply corresponding approximate theories. For example, if in the former case we are tossing the coin in the real world, only approximately 50 % will be heads. Hence, if we ask why about 50 % of these outcomes are heads, we can provide an approximate explanation that it is due to this approximate frequency probability. In other words, our approximate probability statements are close to their true counterparts.

The same goes with the statistical explanations. If we explain why most of the attendants in the IFSA '07 Congress were males, we rely on the approximate statistical result that about 90 % of persons in the fuzzy community are males.

We have to bear in mind that both probabilistic and statistical explanation models above are distinct from the forecasting models. In the former case we know the conclusion and we derive it from the relevant premises, whereas in forecasting the conclusion is unknown for us. Naturally, the systematic power of a theory presupposes both of these features.

In the foregoing models fuzzy probability approach usually means approximate probability variables and values of these variables, i.e., statements of the type

the probability that John is young is very high,

and this area has already been studied quite much in the literature [4,5,6,9,31,37]. In Zadeh's *FLe* we can also apply such approximate probability distributions as approximate normal distributions by using the f-transformation. In this case even such precise variable values as

probability(John's age is 20 years)

yield approximate probabilities. With imprecise values (e.g., John is fairly old) we obtain even more imprecise outputs (Fig. 4).

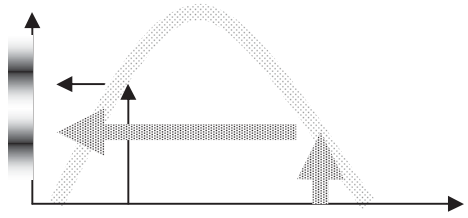


Figure 4. Approximate probability distributions yield approximate outputs even from precise inputs.

The foregoing examples of probability are mainly based on the physicalistic cases, i.e., on the facts and data of the real world. However, Zadeh's theories seem also applicable to epistemic probability in which case these outcomes base on our inference and beliefs [1,30]. In practice we consider the relationship between the hypotheses and their evidence, and we thus assess the degrees of truth, confirmation or belief of hypotheses according to the available relevant evidence.

The gradation of truth was already studied above. As regards the hypothesis testing in general, due to the lack of space we only focus on the degree of confirmation in the context of the explanatory hypotheses and the hypothetico-deductive method [12,13,17,18].

In this hypothesis assessment we usually apply three methods. First, Mill's method of difference [17,18], i.e.,

> *"If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance in common save one, that one occurring only in the former; the circumstance in which alone the two instances differ, is the effect, or the cause, or an indispensable part of the cause, of the phenomenon."*

Second, the disjunctive syllogism

A or B
not A
thus, B.

Third, the modus tollens syllogism

if A, then B
not B
thus, not A.

Mill's principle and the disjunctive syllogism provide us with the general roadmap for excluding the false hypotheses and within the hypothetico-deductive method this task is carried out in practice by applying the modus tollens syllogism.

Hence, given the hypothesis A and its observable or testable logical consequence B, the justifiability of B is assessed according to our knowledge, experiments and observations. If these facts are inconsistent with B, the modus tollens leads us to the conclusion that our hypothesis is false. According to the disjunctive syllogism, in turn, this hypothesis can be rejected or excluded.

If, on the other hand, our experiments and observations correspond with B, the traditional modus tollens will not provide us with any resolution. Hence, in the latter case we have to replace deduction with induction, and we can use such expressions as "our hypothesis is confirmed" in this context [17,18,24,25]. Sufficient degree of confirmation, in turn, will lead to the truth or acceptance of the hypothesis (c.f. below).

In the traditional conduct of inquiry this method thus uses the hypotheses, which stem from the researcher's context of discovery and inventions, deduces tests and experiments from these hypotheses and finally either rejects or confirms the hypotheses according to the known facts and empirical evidence. However, this approach does not provide any resolution to the problem how we can invent or discover novel hypotheses.

Statistical tests also apply the disjunctive method and physicalistic probability when we consider the acceptance of the null and alternative hypotheses according to the tests of significance. We accept the null hypothesis if the value of our test variable does not deviate too much from the "usual" case. Hence, in fact we consider the so-called type 1 error, i.e., we attempt to find justified reasons to reject the null hypotheses. In practice this error is estimated with the p-

ISBN: 978-989-95079-6-8

value in which case we consider the rejection of the null hypotheses if the p-value is sufficiently small (e.g., $p < .05$), i.e., the p-value is our risk to make an erroneous decision if we reject the null hypothesis.

In practice, we have three alternatives in this case [10,16], viz.

1. the null hypothesis is true but our data set is exceptional
2. the null hypothesis is false and our data set is typical
3. the null hypothesis is false and our data set is exceptional

In the conventional case the statistical hypotheses are mutually exclusive, and thus the rejection of one means the acceptance of the other. Hence, according to the disjunctive method, our reasoning boils down to the metarules

1. if the p-value is not small, accept the null hypothesis
2. if the p-value is small, reject the null hypothesis (and accept the alternative hypothesis)

The statistical decision making is thus in this respect based on approximate reasoning and probability. Zadeh's *FLe* can make this reasoning more formal and informative if we operated with the degrees of acceptance and rejection in this context. Then, we could apply such metarule as

> the smaller the p-value, the lower the risk to reject the null hypothesis (the higher the degree of rejection for the null hypothesis)

With the *FLe* we could also take into account better the borderline of these two, viz. the area in which we hesitate over these alternatives. This would mean the fuzzy rules which we have already stated above, i.e.,

1. If the p-value is not small, then accept the null hypothesis.
2. If the p-value is small, then reject the null hypothesis.

With these rules the corresponding fuzzy system could also operate fluently with the borderline cases of p. However, more concrete tools for considering this problem are still expected within the *FLe*.

As regards the modus tollens syllogism in hypothesis assessment, we already considered its bivalent version above. If we apply its approximate version, we can also use linguistic and approximate constituents. Thus, given that $A \approx A'$ and $B \approx B'$, these syllogisms are such as [23]

> if A, then B
> B'
> thus, A'

In this case the truth values of the premises may also be between true and false. This syllogism is analogous to the bivalent case when A' and B' are the antonyms of A and B, respectively, because given that the implication and B' are true, B is false, and thus A must be false (and A' is thus true).

Within the *FLe* the other extreme could be that B' is false in which case B is true and A is thus anything from false to true (and A' is also anything from false to true). In practice this could mean that our hypothesis, A, may be true at least to some extent.

For example,

> If John is young, then he is still a schoolboy.
> John is retired.
> Thus, John is old and reject the hypothesis "John is young".

> If John is young, then he is still a schoolboy.
> John is still a schoolboy.
> Thus, the degree of truth for John being young can be non-zero.

It seems that the approximate version of the modus tollens is more versatile because it is also usable when the second premise is non-false. Then we can at least approximately apply the metarule that the more convincing the evidence for the hypothesis, the higher the degree of truth of our hypothesis. In addition, the more various experiments and observations support our hypotheses, the more true hypothesis is obtained. We can also use the concepts "degree of confirmation" or "degree of acceptance / rejection" in this context if necessary.

Within the *FLe* we can also assume that the implication in the approximate modus tollens syllogism is non-true. For example, if this implication is only fairly true, we can establish that even the false B does not necessarily lead to mere false hypotheses. Equally the truth value of B close to true may already lead to conclusion that the truth value of our hypothesis is anything between false and true. In general, we may thus assume that the non-true implications cause more "dispersion", granulation or imprecision to our conclusions, and loose reasoning links of this type are typical in the human sciences in which we usually operate with noisy data and the complex interrelationships between the variables.

We may also apply the foregoing hypothesis testing to interpretation, which is a widely-used technique in the qualitative research. In this context we attempt to provide a "true" interpretation according to our preliminary interpretation hypothesis (*foreknowledge* in hermeneutics). This procedure of assessing hypotheses is often more subjective by nature than in the quantitative research. However, the *FLe* also seems appropriate to qualitative hypotheses assessment because this inquiry is usually linguistic and approximate by nature [2,3,23].

Zadeh has also considered the idea of the second-order probability under the *FLe* [31,37]. In this context we examine such statements as

> The probability that "the probability of John being very young is fairly high" is high.

This subject matter would also extend a new frontier within both the fuzzy systems and the probability theory, but it also awaits much further studies.

Approximate scientific explanation is related to reasoning, theory formation and hypothesis testing. These, in turn, base much on approximate and probabilistic reasoning. Hence, this object of research is very wide. Zadeh's *FLe* seems to provide promising tools for resolving many of the foregoing problems, but much further studies are still expected in this area.

ISBN: 978-989-95079-6-8

5 Conclusions

We have provided some guidelines how Zadeh's impossibility principle and the extended logic *FLe* could be applied to scientific explanation. We have also examined reasoning, hypothesis assessment and theory formation because these are closely related to explanation.

First traditional reasoning, theory formation and explanation, which use precise entities and bivalent logic in particular in the quantitative research, were considered.

Second, we dealt with Zadeh's *FLe* and his other recent theories, and we examined how we could design usable approximate theories and explanations which essentially base on approximate reasoning and probability.

In particular, we provided examples on our approximation approach to scientific laws, theory formation, statistical and probabilistic reasoning, hypothesis assessment and two well-known syllogisms. These ideas, in turn, were relevant when approximate explanation was considered.

It seems that the foregoing methods are usable in approximate explanation and in this context we could apply the general principle that approximate explanations are in the neighborhood of their true counterparts. This aim is essentially achieved by using approximate reasoning.

Zadeh's novel ideas seem promising in particular in the qualitative research because then we operate with imperfect information, and in this context we still lack a uniform and usable logico-linguistic system. It also seems possible to apply his ideas to the quantitative research, in particular if we attempt to mimic human reasoning in a computer environment.

We must bear in mind that several of the foregoing issues are still controversial even in the philosophy of science, and the same problem will likely concern Zadeh's and Author's ideas above.

However, due to the lack of space, this paper only provides at metalevel some general guidelines for approximate scientific explanation and related topics from the standpoint of philosophy of science, and hence this relevant subject matter still awaits much further studies in the future.

Acknowledgements

The author expresses his deep gratitude to Prof. Lotfi Zadeh at the UC Berkeley for the original ideas, articles, several discussions and encouragement which made it possible to write this article. I also express my thanks to the anonymous referees for their valuable comments.

References

[1] Carnap, R. (1962): The Logical Foundations of Probability. The University of Chicago Press, Chicago.

[2] Denzin, N. & Lincoln, Y. (1994): Handbook of Qualitative Research. Sage, Thousand Oaks.

[3] Dimitrov, V. & Hodge, B. (2002): Social Fuzziology – Study of Fuzziness of Social Complexity. Physica Verlag, Heidelberg.

[4] Dubois, D. & al. (1991): Fuzzy Sets in Approximate Reasoning, Part 2: Logical Approaches. Fuzzy Sets and Systems 40, 203 - 244.

[5] Dubois, D. & Prade, H. (1980): Fuzzy Sets and Systems, Theory and Applications. Academic Press, New York.

[6] Dubois, D. & Prade, H. (1991): Fuzzy Sets in Approximate Reasoning, Part 1: Inference with Possibility Distributions. Fuzzy Sets and Systems 40, 143-202.

[7] Fetzer, J. (1993): Philosophy of Science. Paragon House, New York, N. Y.

[8] Gadamer, H.-G. (1975): Truth and Method. Sheed & Ward, London.

[9] Grzegorzewski, P. & al. (Eds.) (2002): Soft Methods in Probability, Statistics and Data Analysis. Physica Verlag, Heidelberg.

[10] Guilford, J. & Fruchter, B. (1978): Fundamental Statistics in Psychology and Education. McGraw-Hill, London.

[11] Haack, S. (1978): Philosophy of Logics. Cambridge University Press, Cambridge.

[12] Hempel, C. (1965): Aspects of Scientific Explanation and Other Essays in the Philosophy of Science. The Free Press, New York.

[13] Hempel, C. (1966): Philosophy of Natural Science. Prentice Hall, Englewood Cliffs.

[14] Hintikka, J. & Suppes, P. (Eds.) (1970): Information and Inference. Reidel, Dordrecht.

[15] Kaplan, A. (1973): The Conduct of Inquiry: Methodology for Behavioral Sciences. Chandler, New York, N. Y.

[16] Kakkuri-Knuuttila, M.-L. (1998): Argumentti ja kritiikki (Argument and Criticism). Tammer-Paino, Tampere.

[17] Nagel, E. (1961): The Structure of Science. Routledge & Kegan Paul, London.

[18] Niiniluoto, I. (1983): Tieteellinen päättely ja selittäminen (Scientific Reasoning and Explanation). Otava, Keuruu.

[19] Niiniluoto, I. (1984): Johdatus tieteenfilosofiaan (Introduction to the Philosophy of Science). Otava, Keuruu.

[20] Niiniluoto, I. (1987): Truthlikeness. Reidel, Dordrecht.

[21] Niskanen, V. A. (2009): Fuzzy Systems and Scientific Method – Meta-Level Reflections and Prospects. In Seising, R. (Ed.): Fuzzy Set Theory - Philosophy, Logics, and Criticism. Springer.

[22] Niskanen, V. A. Meta-Level Prospects for Scientific Theory Formation by Adopting the Information Granulation Approach. Proceedings of the IMS '08 Conference, Sakarya University, pp. 33-41.

[23] Niskanen, V. A. (2004): Soft Computing Methods in Human Sciences. Studies in Fuzziness and Soft Computing, Vol. 134. Springer Verlag, Berlin.

[24] Popper, K. (1963): Conjectures and Refutations. Routledge & Kegan Paul, London.

[25] Popper, K. (1959): The Logic of Scientific Discovery. Hutchinson, London.

[26] Rescher, N. (1980): Empirical Inquiry. The Athlone Press, London.

[27] Rescher, N. (1970): Scientific Explanation. The Free Press, New York.

[28] Runes, D. (Ed.) (1976): Dictionary of Philosophy. Littlefield, Adams & Co., Totowa, New Jersey.

[29] Von Wright, G. H. (1971): Explanation and Understanding. Cornell University Press, Cornell.

[30] Von Wright, G. H. (1957): The Logical Foundations of Probability. Blackwell, Oxford.

[31] Zadeh, L. Computation with Imprecise Probabilities. Plenary lecture, International Conference on Information Processing and Management of Uncertainty for Knowledge-Based Systems (IPMU), Malaga, Spain, June 22-27, 2008.

[32] Zadeh, L. (2006): From Search Engines to Question Answering Systems? The Problems of World Knowledge, Relevance, Deduction and Precisiation. In E. Sanchez, (Ed.), Fuzzy Logic and the Semantic Web, Elsevier, Amsterdam.

[33] Zadeh, L. (1975): Fuzzy Logic and Approximate Reasoning. Synthese 30, 407-428.

[34] Zadeh, L. (1996): Fuzzy Logic = Computing with Words. IEEE Transactions on Fuzzy Systems, 2, 103-111.

[35] Zadeh, L. (1997): Toward a Theory of Fuzzy Information Granulation and Its Centrality in Human Reasoning and Fuzzy Logic. Fuzzy Sets and Systems 90/2, 111-127.

[36] Zadeh, L. (1999): From Computing with Numbers to Computing with Words - From Manipulation of Measurements to Manipulation of Perceptions. IEEE Transactions on Circuits and Systems 45, 105-119.

[37] Zadeh, L. (2002): Toward a Perception-Based Theory of Probabilistic Reasoning with Imprecise Probabilities. Journal of Statistical Planning and Inference 105/2, 233-264.

[38] Zadeh, L., (2009): Toward Extended Fuzzy Logic – A First Step. Fuzzy Sets and Systems (to appear).

ISBN: 978-989-95079-6-8

On the completion mechanism produced by the Choquet integral on some decision strategies

Christophe Labreuche[1]

1. Thales Research & Technology
Campus de Polytechnique, 1 avenue Augustin Fresnel
91767 Palaiseau Cedex- France
christophe.labreuche@thalesgroup.com

Abstract— Preference modeling consists in constructing a preference relation from initial preferences given by a decision maker. We are interested in the preference relation obtained from the use of the Choquet integral. The necessity preference is constructed as the intersection of all preference relations corresponding to a Choquet integral which are compatible with the initial preferences of the decision maker. We study some properties of this necessity relation on some classes of initial preferences. This gives some properties on the completion or extension ability of the Choquet integral.

Keywords— Choquet integral, capacity, initial preferences.

1 Introduction

Multi-Criteria Decision Aid aims at representing the preferences of a Decision Maker (DM) over some options described by several points of view or attributes. We are interested in this paper in the case where the preference relation can be described by an overall utility in which the aggregation of the attributes is obtained with the Choquet integral. The preference model is then thoroughly constructed once the parameters of the Choquet integral - namely the fuzzy measure - are determined.

To this end, the facilitator asks the DM to provide some preferential information. The most broadly used type of preferential information is when the DM says that an option is preferred to another one. The major difficulty the facilitator faces is that there usually does not exist one single fuzzy measure that fulfils this preferential information. Most of the elicitation methods based on the Choquet integral consist in selecting a fuzzy measure (fulfilling the preferential information) that maximizes some functional, which may be an entropy for instance [1, 2, 3]. This is not quite satisfactory for the DM since he does not usually understand what maximizing the functional really means. The use of a maximization problem introduces some additional information that does not come from the DM.

The facilitator shall rather stick strictly to what the DM says and add no further information. One then looks for a robust way to recommend some comparisons among the options from the preferential information [4]. The concept of *necessity preference relation* has been recently introduced for robust quantitative multi-criteria decision models [5, 6]. It has been applied to the Choquet integral in [7]. An option is *necessarily* preferred to another option according to the necessity preference relation, if the first option is preferred to the second one according to all models that fulfill the preferential information provided by the DM. This necessity preference relation is usually incomplete, unless the model is completely specified from the preferential information of the DM. The necessity preference relation is cautious in the sense that its outcomes cannot be contradicted.

The goal of this paper is to study the necessity preference relation for several examples of typical preferential information, in the case where the underlying model is the Choquet integral. In particular, we want to analyze the strength of the completion mechanism obtained from the Choquet integral. This will give complementary insights (compare to the axiomatic characterizations) on the ability of the Choquet integral to represent DM preferences. For each preferential information, the layout is the following:

- Statement of the preferential information and precise description in terms of binary relations.
- Equivalence conditions on the model. We determine the necessary and sufficient conditions on the parameters of the model for this latter to satisfy the preferential information.
- Description of the necessity preference relation derived from the preferential information.

Section 2 presents a more formal description of the problem. The Choquet integral is defined in Section 3. Section 4 describes the process elicitation of the preferential information. Three decision strategies are then studied: the case of empty preference information (see Section 5), the relative importance (see Section 6) and the veto (see Section 7).

2 Description of the problem

Multi-Criteria Decision Aid aims at modeling the preferences of a Decision Maker (DM) over alternatives described by several points of view $N = \{1, \ldots, n\}$. The points of view to be taken into account in the decision making process are denoted by $X_1, \ldots, X_n$. An alternative is characterized by a value w.r.t. each point of view

ISBN: 978-989-95079-6-8

and is thus identified to a point in the Cartesian product X of the points of view: $X = X_1 \times \cdots \times X_n$.

Preference modeling aims at helping a DM to compare the options of a set $\Xi \subseteq X$ containing the alternatives of interest for him. We denote by $\succsim$ the preference relation over X that we wish to define. We are interested in the case where $\succsim$ results from an algebraic utility model. According to this model, the preference relation $\succsim$ can be represented by an overall utility $U : X \to \mathbb{R}$. Hence for all $x, y \in X$, we have

$$x \succsim y \iff U(x) \geq U(y).$$

A standard representation of U is the so-called transitive decomposable model $U(x) = F(u(x))$, where $u_i : X_i \to \mathbb{R}$ is the utility function on attribute $i \in N$, $u(x) = (u_1(x_1), \ldots, u_n(x_n))$, and $F : \mathbb{R}^N \to \mathbb{R}$ is the aggregation function. We assume here that u is known so that the options can be described by a vector in $\mathbb{R}^n$. Hence $X = \mathbb{R}^n$ and the utility functions are the identity function. The aggregation function is characterized by some parameter vector w that need to be determined. Let $\mathcal{W}$ the set of values of the parameters. The preference relation obtained from the parameter vector w is denoted by $\succsim_w$. We denote by $\succ_w$ and $\sim_w$ the asymmetric and symmetric parts of $\succsim_w$.

Any binary relation on $\mathbb{R}^n$ can be identified to a subset of $(\mathbb{R}^n)^2$ composed of the pairs (a, b) such that a is in relation with b.

The fact that the DM needs some help to compare the options in Ξ, means that the comparison of at least some of the options in Ξ is complex. Some preferential information, containing the preferences of the DM over several options in $X \backslash \Xi$, is asked to the DM. This preferential information is encoded by three partial orders $\trianglerighteq, \triangleright, \equiv$ over X. For $x, y \in X$, relation $x \trianglerighteq y$ means that the DM finds x at least as good as y, $x \triangleright y$ means that the DM finds x strictly better than y, and $x \equiv y$ means that the DM finds x similar to y. We set $\square = (\trianglerighteq, \triangleright, \equiv)$. In order to deduce a comparison of the elements of Ξ from $\square$, one necessarily needs some completion mechanism. We want to construct three binary relations $\succsim$, $\succ$ and $\sim$ on X that extend $\square$. The strength of the completion is measured through the differences $\succsim \backslash \trianglerighteq$, $\succ \backslash \triangleright$ and $\sim \backslash \equiv$ compare to $\trianglerighteq, \triangleright, \equiv$ respectively. For a very weak completion, we have $\succsim \backslash \trianglerighteq = \emptyset$, $\succ \backslash \triangleright = \emptyset$ and $\sim \backslash \equiv = \emptyset$.

Let $\mathcal{W}(\square)$ be the set of parameter vectors w for which $x \succsim_w y$ whenever $x \trianglerighteq y$, $x \succ_w y$ whenever $x \triangleright y$, and $x \sim_w y$ whenever $x \equiv y$. The completion is necessarily based on the family $\{\succsim_w\}_{w \in \mathcal{W}(\square)}$. A classical completion is the preference relation $\succsim_w$ corresponding to a particular $w \in \mathcal{W}(\square)$ solution to an optimization problem [1, 2, 3]

$$w = \text{argmax}_{w \in \mathcal{W}(\square)} G_\square(w) \tag{1}$$

where $G_\square$ is the function depending on $\square$ to be maximized. As said in the introduction, we are interested in the *necessity* preference relation $\succsim^{\mathcal{W}}_{\text{N},\square}$ that is constructed from $\square$ as follows [7, 5, 6]:

$$x \succsim^{\mathcal{W}}_{\text{N},\square} y \iff [\forall w \in \mathcal{W}(\square) \;\; x \succsim_w y]\,. \tag{2}$$

One may also define $\succ^{\mathcal{W}}_{\text{N},\square}$ and $\sim^{\mathcal{W}}_{\text{N},\square}$ by

$$x \succ^{\mathcal{W}}_{\text{N},\square} y \iff [\forall w \in \mathcal{W}(\square) \;\; x \succ_w y]$$
$$x \sim^{\mathcal{W}}_{\text{N},\square} y \iff [\forall w \in \mathcal{W}(\square) \;\; x \sim_w y]\,.$$

Note that one may have $\succsim^{\mathcal{W}}_{\text{N},\square} \neq \succ^{\mathcal{W}}_{\text{N},\trianglerighteq} \cup \sim^{\mathcal{W}}_{\text{N},\square}$.

Another interesting order relation derived from $\square$ is the so-called *possibility preference relation* defined as follows [7, 5, 6]:

$$x \succsim^{\mathcal{W}}_{\Pi,\square} y \iff [\exists w \in \mathcal{W}(\square) \;\; x \succsim_w y]\,.$$

Let us mention that these concepts of necessary and possible preference representation are originated from earlier works on preference modelling. In the context of decision rules, the necessary and possibility preference relations are the lower and upper approximations of a rough set [8]. Moreover, in qualitative decision models, these concepts correspond to the dominance and ordering queries for the Ceterus Paribus nets [9].

The choice of the representation of $\succsim_w$ (and thus of the family of aggregation functions F) is very important since it characterizes directly the completion mechanism. The choice of the family F results from a compromise: if the completion is too strong, we deduce much more than what the DM expresses and the DM may not agree with the outcome. If the completion is too weak, we basically do not deduce really more than what the DM says. Moreover, the stronger the completion mechanism, the less information $\square$ needs to be given.

As an example, the completion resulting from a simple model such a weighted sum is stronger than that obtained from a more general model such as the Choquet integral. We will illustrate this in the main part of the paper. Consider the case where F is the usual weighted sum: $F(x) = \sum_{i \in N} w_i\, x_i$ where w_i is the weight of criterion i. Thanks to the additivity properties of this model, one can deduce that $(1, \frac{1}{2}, 0) \succsim^{\mathcal{W}}_{\text{N},\square} (0, 1, \frac{1}{2})$, from the simplier comparisons $(\frac{1}{2}, \frac{1}{2}, 0) \trianglerighteq (0, \frac{1}{2}, \frac{1}{2})$, $(\frac{1}{2}, 0, 0) \trianglerighteq (0, \frac{1}{2}, 0)$ provided by the DM. The previous deduction is wrong for the Choquet integral. The weighted sum, which is a very simple aggregation model, implies a very strong completion mechanism.

The necessity preference relation $\succsim^{\mathcal{W}}_{\text{N},\square}$ corresponds to the comparisons among the options in X that can be made for sure. If the initial preferences $\square$ of the DM are fixed, the recommendation given by the facilitator shall be based only on $\succsim^{\mathcal{W}}_{\text{N},\square}$.

Now, if $\succsim^{\mathcal{W}}_{\text{N},\square}$ is not enough complete, the facilitator will probably try to enrich $\square$ with new comparisons provided by the DM. The candidate comparisons must be picked up from the possibility preference relation. Assume indeed that $\square$ is enriched with a new information "x is at least as good as y" such that $x \not\succsim^{\mathcal{W}}_{\Pi,\square} y$. The enriched preferential information is $\square' = (\trianglerighteq', \triangleright', \equiv')$ with $\trianglerighteq' = \trianglerighteq \cup \{(x, y)\}$, $\triangleright' = \triangleright$ and $\equiv' = \equiv$. Since $x \not\succsim^{\mathcal{W}}_{\Pi,\square} y$, we obtain $\mathcal{W}(\square') = \emptyset$. Hence the new preferential information to add to $\square$ must belong to $\succsim^{\mathcal{W}}_{\Pi,\square}$, $\succ^{\mathcal{W}}_{\Pi,\square}$ or $\sim^{\mathcal{W}}_{\Pi,\square}$. One clearly has [5, Proposition 4.1]

$$\succsim^{\mathcal{W}}_{\text{N},\square} \subseteq \succsim^{\mathcal{W}}_{\Pi,\square}\,.$$

ISBN: 978-989-95079-6-8

The comparisons contained in the necessity preference relation are automatically deduced from $\Box$, and do not therefore enrich $\Box$. In other words, if $(x,y) \in \succsim^{\mathcal{W}}_{\mathrm{N},\Box}$, then $\mathcal{W}(\Box') = \mathcal{W}(\Box)$, where $\Box' = (\trianglerighteq', \rhd', \equiv')$ with $\trianglerighteq' = \trianglerighteq \cup \{(x,y)\}$, $\rhd' = \rhd$ and $\equiv' = \equiv$. This proves that the new relevant preferential information to add to $\Box$ actually belongs to

$$\succsim^{\mathcal{W}}_{\Pi,\Box} \setminus \succsim^{\mathcal{W}}_{\mathrm{N},\Box}\ ,\ \succ^{\mathcal{W}}_{\Pi,\Box} \setminus \succ^{\mathcal{W}}_{\mathrm{N},\Box} \quad \text{or} \quad \sim^{\mathcal{W}}_{\Pi,\Box} \setminus \sim^{\mathcal{W}}_{\mathrm{N},\Box}\ .$$

In this paper, we will not study specifically the possibility preference relation. We will focus only on the necessity preference relation. The reason is that $\succsim^{\mathcal{W}}_{\mathrm{N},\Box}$ and $\succsim^{\mathcal{W}}_{\Pi,\Box}$ are strongly linked together. More precisely, since $\succsim_w$ is a complete preorder, we have

$$x \succsim^{\mathcal{W}}_{\Pi,\Box} y \iff y \not\succ^{\mathcal{W}}_{\mathrm{N},\Box} x.$$

In other words, $\succsim^{\mathcal{W}}_{\Pi,\Box} = X^2 \setminus (\succ^{\mathcal{W}}_{\mathrm{N},\Box})^{-1}$.

3 Notation and the Choquet integral

We hereafter restrict to the case where F is a Choquet integral.

3.1 Choquet integral

A fuzzy measure (also called *capacity*) on a set N of criteria is a set function $\mu : 2^N \to [0,1]$ such that [10]

- $\mu(\emptyset) = 0$, $\mu(N) = 1$,
- $\forall A \subseteq B \subseteq N\ ,\ \mu(A) \leq \mu(B)$

Let $\mathcal{M}$ be the set of all fuzzy measures. A fuzzy measure is said to be additive if $\mu(A \cap B) = \mu(A) + \mu(B)$ for every pair (A,B) of disjoint coalitions.

The Choquet integral of $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ defined w.r.t. a capacity μ has the following expression [11] :

$$C_\mu(x_1, \ldots, x_n) = \sum_{i=1}^{n} \left(x_{\tau(i)} - x_{\tau(i-1)}\right) \times \mu\left(\{\tau(i), \cdots, \tau(n)\}\right), \qquad (3)$$

where τ is a permutation on N such that $x_{\tau(1)} \leq x_{\tau(2)} \leq \cdots \leq x_{\tau(n)}$, and $x_{\tau(0)} := 0$. The Choquet integral has been proved to be able to model both the importance of criteria and the interaction between criteria. One fundamental property of the Choquet integral is that

$$C_\mu(1_A, 0_{N \setminus A}) = \mu(A). \qquad (4)$$

3.2 k-additive capacities

We introduce a useful linear transformation of fuzzy measures. The *Möbius transform* m of a fuzzy measure μ is the unique solution of the equation

$$\forall A \subseteq N \quad \mu(A) = \sum_{B \subseteq A} m(B), \qquad (5)$$

and is given by:

$$m^\mu(A) := \sum_{B \subseteq A} (-1)^{|A|-|B|} \mu(B). \qquad (6)$$

A fuzzy measure is defined by 2^n coefficients, which is much more than a weighted sum. The concept of k-additive fuzzy measure has a complexity in-between a fuzzy measure and a weighted sum. More precisely, a fuzzy measure μ is said to be k-additive [12] if $m^\mu(A) = 0$ whenever $|A| > k$ and there exists A with $|A| = k$ such that $m^\mu(A) \neq 0$. We denote by $\mathcal{M}^k$ the set of fuzzy measures that are at most k-additive (i.e. 1, or 2, or ..., or k additive). Note that $\mathcal{M} = \mathcal{M}^n$ and that $\mathcal{M}^1$ are the additive fuzzy measures and corresponds to the usual weighted sum. An interesting particular case is when $k = 2$.

3.3 Interpretation

A capacity contains 2^n terms, which makes its interpretation complex for a DM. In order to ease its interpretation, several global indices have been defined. The first index called *importance index* aims at measuring the degree to which a criterion is important for the decision. The importance index of criterion $i \in N$ is defined by [13]:

$$v_i := \sum_{A \subset N \setminus \{i\}} \frac{(n - |A| - 1)!|A|!}{n!} [\mu(A \cup \{i\}) - \mu(A)].$$

Note that the definition is originated from Cooperative Game Theory in which it corresponds to a fair share of a common wealth among several players.

The interaction index I_{ij} of the couple of criteria $\{i,j\}$ is defined by [14]

$$I_{ij} := \sum_{A \subset N \setminus \{i,j\}} \frac{(n - |A| - 2)!|A|!}{(n-1)!} [\mu(A \cup \{i,j\}) - \mu(A \cup \{i\}) - \mu(A \cup \{j\}) + \mu(A)].$$

The Shapley index v_i can be interpreted as a kind of average value of the contribution of criterion i alone in all coalitions. The interaction index I_{ij} can be interpreted as a kind of average value of the *added value* obtained by putting criterion i and j together. When I_{ij} is positive (resp. negative), the interaction is said to be positive (resp. negative). These two concepts correspond to a mean behavior and are anyhow not so easy to understand for a DM.

3.4 Initial preferential information

We denote by $\mathcal{M}^k(\Box)$ the set of capacities in $\mathcal{M}^k$ that satisfy the preferential information $\Box = (\trianglerighteq, \rhd, \equiv)$:

$$\mathcal{M}^k(\Box) = \Big\{ \mu \in \mathcal{M}^k\ , \forall a, b \in \mathbb{R}^n \quad a \trianglerighteq b \ \Rightarrow\ C_\mu(a) \geq C_\mu(b) \quad a \rhd b \ \Rightarrow\ C_\mu(a) > C_\mu(b) \quad a \equiv b \ \Rightarrow\ C_\mu(a) = C_\mu(b) \Big\}$$

The necessity preference relation on $\mathbb{R}^n$ is a cautious way to take into account the fact that the aggregation model is not uniquely determined from a knowledge of $\Box$, and

ISBN: 978-989-95079-6-8

is defined by:

$$\forall a, b \in \mathbb{R}^n \quad , \quad a \succsim_{\mathrm{N},\Box}^{\mathcal{M}^k} b \iff [\forall \mu \in \mathcal{M}^k(\Box) \quad C_\mu(a) \geq C_\mu(a)] \, .$$

It is easy to see that $\succsim_{\mathrm{N},\Box}^{\mathcal{M}^k}$ is usually incomplete, but is transitive and reflexive. One can define $\succ_{\mathrm{N},\Box}^{\mathcal{M}^k}$ and $\sim_{\mathrm{N},\Box}^{\mathcal{M}^k}$ as in Section 2. One has

$$\sim_{\mathrm{N},\Box}^{\mathcal{M}^k} = \succsim_{\mathrm{N},\Box}^{\mathcal{M}^k} \cap (\succsim_{\mathrm{N},\Box}^{\mathcal{M}^k})^{-1} \quad \text{and} \quad \succ_{\mathrm{N},\Box}^{\mathcal{M}^k} \subseteq \succsim_{\mathrm{N},\Box}^{\mathcal{M}^k} \setminus \sim_{\mathrm{N},\Box}^{\mathcal{M}^k} .$$

Let us define finally the incomparability as $I_{\mathrm{N},\Box}^{\mathcal{M}^k} = (\mathbb{R}^n)^2 \setminus (\succsim_{\mathrm{N},\Box}^{\mathcal{M}^k} \cup (\succsim_{\mathrm{N},\Box}^{\mathcal{M}^k})^{-1})$.

4 Elicitation process

The preference model is usually constructed iteratively through a decision aiding process [15]. This process is centered around the interaction between the DM and the decision aiding system. The information that the DM expresses regarding his preferences can be of two types:

- The first type of information can be easily produced and interpreted by the DM. It is usually admitted that the DM can naturally compare some options. In this paper, the only information of this type corresponds to the three binary relations $\Box = (\trianglerighteq, \rhd, \equiv)$. It is called the *Preferential information*.

- The second type of information (called *Decision strategies*) corresponds to abstract information concerning the preference model. It can typically concern concepts such as the mean importance, veto, relative importance,... The DM can understand them but only to some extend, but it is unexpected that he will naturally provide information of this type. If the decision analyst guesses that the DM fulfills to a decision strategy, he needs to convert it to preferential information in order to check the validity of his guess.

At the end, the only learning data that is considered from the DM corresponds to some preferential information $\Box$. It is transformed in the *internal representation* of the model. This is the set of parameter vectors fulfilling the preferential information, that is $\mathcal{M}^k(\Box)$. It is not understood by the DM. Consistency and completeness is checked in this representation. The necessity preference relation can then be constructed. This relation can be presented to the DM in different ways. It can be directly applied to the set Ξ of options of interest for the DM. It is also possible to interpret the necessity preference relation in terms of decision strategies. The interpretation may concern some indices regarding the mean importance, veto, relative importance, if all the admissible capacities $\mathcal{M}^k(\Box)$ share some common properties on one of these concepts.

The elicitation process can thus be seen as a flow diagram between the three concepts and is organized on the steps *a* to *f* in Figure 1:

- *a*: The DM can express some abstract information regarding his decision strategies. He may express for instance that a criterion is more important than another one.

- *b*: The DM can express information in the format of the order relations $\Box$. He may say for instance that an option is preferred to another one. If the DM has already given some preferential information (denoted by $\Box'$) in a previous step of the elicitation process and he wishes to enrich $\Box'$, the new preferential information shall lie in the sets $\succsim_{\Pi,\Box'}^{\mathcal{M}^k} \setminus \succsim_{\mathrm{N},\Box'}^{\mathcal{M}^k}$, $\succ_{\Pi,\Box'}^{\mathcal{M}^k} \setminus \succ_{\mathrm{N},\Box'}^{\mathcal{M}^k}$ or $\sim_{\Pi,\Box'}^{\mathcal{M}^k} \setminus \sim_{\mathrm{N},\Box'}^{\mathcal{M}^k}$.

- *c*: When the DM expresses abstract information (decision strategies), they shall be converted into preferential information through $\Box$ in order to be precisely set and taken into account by the elicitation method.

- *d*: The preferential information is transformed into constraints in the parameter set. This corresponds to the set $\mathcal{M}^k(\Box)$. This step is sometimes called disaggregation step.

- *e*: An interpretation of the set $\mathcal{M}^k(\Box)$ is presented in terms of, for instance, veto, importance, interaction indices. It can also provide the necessity and possibility preference relations. This step is sometimes called aggregation step.

- *f*: An interpretation of the results of the elicitation is displayed to the DM in some understandable way.

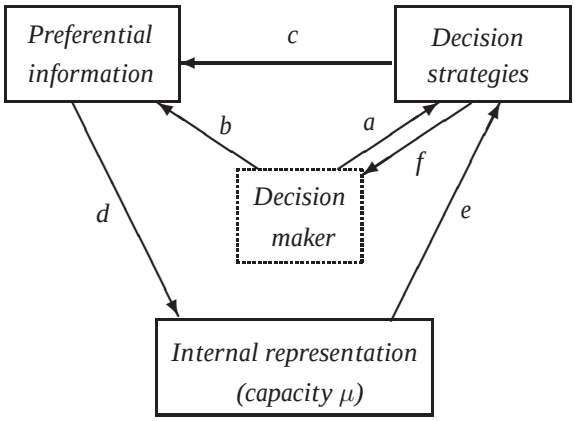


Figure 1: Eliciation process.

In the remaining of this paper, we are interested in some abstract decision strategies that the DM often wishes to express in practice. For each type of preferential information, we determine how it is converted into order relations $\Box$ (step *c*). Then, we construct the necessity preference relation corresponding to this preferential information (steps *d* and *e*).

5 Case of empty preference information

We denote by $\Box_\emptyset$ the empty preference information. When the DM says nothing, there is of course no new

ISBN: 978-989-95079-6-8

constraint on the fuzzy measure:

$$\mathcal{M}^k(\Box_\emptyset) = \mathcal{M}^k.$$

Moreover, it is easy to check that for all $k \in \{1, \ldots, n\}$

$$\succsim^{\mathcal{M}^k}_{\mathrm{N},\Box_\emptyset} = \mathcal{P}_{\mathrm{a}} := \{(a,b) \in \mathbb{R}^n \ , \ \forall i \in N \ \ a_i \geq b_i\} \ .$$

This relation extends [5, Remark 4.1] to the case of the Choquet integral. This is the standard Pareto dominance order. When no preferential information is given, the necessary preferential information is similar to the Pareto dominance order, which is what one expects.

6 Relative importance between criteria

6.1 Statement of the preferential information

We want to express a preferential information regarding the relative importance between two criteria – namely that criterion p is more important than criterion q. The first idea is to interpret this as the relation $v_p > v_q$ based on the Shapley value. This relation is considered in [1, 7]. However, as explained in Section 3.3, the Shapley value is not a simple concept for a DM. It is not clear at all whether the statement *"criterion p is more important than criterion q"* expressed by a DM can be represented by the relation $v_p > v_q$. We wish only to express the previous statement only as binary relations denoted by $\Box^{p,q}_{\mathrm{RelImp}} = (\trianglerighteq^{p,q}_{\mathrm{RelImp}}, \triangleright^{p,q}_{\mathrm{RelImp}}, \equiv^{p,q}_{\mathrm{RelImp}})$. The following definition seems to express quite accurately the previous statement.

Definition 1 ([16]) *A criterion $p \in N$ is said to be* more important *than criterion $q \in N \setminus \{p\}$ iff*

$$\left(\gamma_p, \beta_q, a_{N\setminus\{p,q\}}\right) \triangleright^{p,q}_{\mathrm{RelImp}} \left(\beta_p, \gamma_q, a_{N\setminus\{p,q\}}\right) \qquad (7)$$

for all $\gamma, \beta \in \mathbb{R}$ with $\gamma > \beta$, and all $a_{N\setminus\{p,q\}} \in \mathbb{R}^{n-2}$, where $\left(\alpha_p, \beta_q, a_{N\setminus\{p,q\}}\right)$ denotes the alternative having value α on criterion p, value β on criterion q, and value $a_{N\setminus\{p,q\}}$ on the remaining criteria.

Note that (7) makes sense only if the scales of the criteria are commensurate.

6.2 Conditions on the fuzzy measure

Let $\delta_{p,q}\mu(A) := \mu(A \cup \{p\}) - \mu(A \cup \{q\})$.

Lemma 1 *For all $k \in \{2, \ldots, n\}$, criterion p is more important than criterion q with the Choquet integral with a capacity in $\mathcal{M}^k$ iff*

$$\forall A \subseteq N \setminus \{p,q\} \quad \delta_{p,q}\mu(A) > 0. \qquad (8)$$

6.3 Completion

We have the following result.

Proposition 1 *Let $p, q \in N$ with $p \neq q$. Then for all $k \in \{2, \ldots, n\}$*

$$\begin{aligned}
&\succsim^{\mathcal{M}^k}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}} \\
&= \Big\{ \Big((\gamma_p, \beta_q, a_{N\setminus\{p,q\}}), (\beta'_p, \gamma'_q, a'_{N\setminus\{p,q\}}) \Big) \in (\mathbb{R}^n)^2 \ , \\
&\quad \gamma > \beta \ , \ \gamma' > \beta' \ , \gamma' \leq \gamma \ , \ \beta' \leq \beta \text{ and} \\
&\quad \text{for all } i \in N \setminus \{p,q\} \ \ a'_i \leq a_i \Big\} \bigcup \mathcal{P}_{\mathrm{a}} \\
&\succ^{\mathcal{M}^k}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}} \\
&= \Big\{ \Big((\gamma_p, \beta_q, a_{N\setminus\{p,q\}}), (\beta'_p, \gamma'_q, a'_{N\setminus\{p,q\}}) \Big) \in (\mathbb{R}^n)^2 \ , \\
&\quad \gamma > \beta \ , \ \gamma' > \beta' \ , \gamma' \leq \gamma \ , \ \beta' \leq \beta \ \text{and} \\
&\quad \text{for all } i \in N \setminus \{p,q\} \ \ a'_i \leq a_i \Big\} \bigcup \mathcal{P}^p_{\mathrm{a}>} \\
&\sim^{\mathcal{M}^k}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}} = \Big\{ (a,a) \ , \ \ a \in \mathbb{R}^n \Big\}
\end{aligned}$$

where $\mathcal{P}^p_{\mathrm{a}>} = \{(a,b) \in \mathcal{P}_{\mathrm{a}} \ , \ a_p > b_p\}$.

When $\gamma > \beta$, $(\gamma_p, \beta_q, a_{N\setminus\{p,q\}})$ is at least as good as $(\beta_p, \gamma_q, a_{N\setminus\{p,q\}})$ from $\Box^{p,q}_{\mathrm{RelImp}}$. From the Pareto relation $\mathcal{P}_{\mathrm{a}}$, $(\beta_p, \gamma_q, a_{N\setminus\{p,q\}})$ is at least as good as $(\beta'_p, \gamma'_q, a'_{N\setminus\{p,q\}})$ whenever $\gamma' \leq \gamma$, $\beta' \leq \beta$ and $a'_k \leq a_k$ for all $k \in N \setminus \{p,q\}$. The completion of $\Box^{p,q}_{\mathrm{RelImp}}$ by a k-additive Choquet integral (with $k \geq 2$) is just the combination of the definition of $\Box^{p,q}_{\mathrm{RelImp}}$ and the Pareto dominance order. Therefore, there is no side effect coming from the Choquet model.

Lemma 2 *Let $p, q \in N$ with $p \neq q$. Then*

$$\begin{aligned}
&\succsim^{\mathcal{M}^1}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}} \\
&= \Big\{ \Big((\gamma_p, \beta_q, a_{N\setminus\{p,q\}}), (\beta'_p, \gamma'_q, a'_{N\setminus\{p,q\}}) \Big) \in (\mathbb{R}^n)^2 \\
&\quad \gamma \geq \beta' \ , \ \gamma - \beta + \beta' - \gamma' \geq 0 \\
&\quad \text{and for all } i \in N \setminus \{p,q\} \ \ a'_i \leq a_i \Big\} \bigcup \mathcal{P}_{\mathrm{a}} \\
&\succ^{\mathcal{M}^1}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}} \\
&= \Big\{ \Big((\gamma_p, \beta_q, a_{N\setminus\{p,q\}}), (\beta'_p, \gamma'_q, a'_{N\setminus\{p,q\}}) \Big) \in (\mathbb{R}^n)^2 \\
&\quad \gamma > \beta' \ , \ \gamma - \beta + \beta' - \gamma' \geq 0 \ , \ \gamma' \leq \gamma \ , \ \beta' \leq \beta \\
&\quad \text{and for all } i \in N \setminus \{p,q\} \ \ a'_i \leq a_i \Big\} \cup \mathcal{P}^p_{\mathrm{a}>} \\
&\sim^{\mathcal{M}^1}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}} = \Big\{ (a,a) \ , \ \ a \in \mathbb{R}^n \Big\}
\end{aligned}$$

There is a strange condition $\gamma - \beta + \beta' - \gamma' \geq 0$ in the necessity preference relation of the model $\mathcal{M}^1$.

One has for all $k \in \{2, \ldots, n\}$

$$\begin{aligned}
\succsim^{\mathcal{M}^k}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}} &\subset \succsim^{\mathcal{M}^1}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}} \\
\succ^{\mathcal{M}^k}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}} &\subset \succ^{\mathcal{M}^1}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}} \\
\sim^{\mathcal{M}^k}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}} &\subset \sim^{\mathcal{M}^1}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}}
\end{aligned}$$

The information $\succsim^{\mathcal{M}^1}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}} \setminus \succsim^{\mathcal{M}^k}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}}$ and $\succ^{\mathcal{M}^1}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}} \setminus \succ^{\mathcal{M}^k}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}}$ is contained neither in Definition

ISBN: 978-989-95079-6-8

1 nor in the Pareto set $\mathcal{P}_{\rm a}$. This proves that the completion resulting from the Choquet integral fits exactly with the preferential information provided by the DM, whereas the completion resulting from the weighted sum is strong. The Choquet is thus able to represent the relative importance without adding new information not provided by the DM.

7 Veto criteria

7.1 Statement of the preferential information

As in group decision, one can define a veto criterion. The preferential information corresponding to a veto criterion p is denoted by $\Box^p_{\rm Veto} = (\trianglerighteq^p_{\rm Veto}, \triangleright^p_{\rm Veto}, \equiv^p_{\rm Veto})$.

Definition 2 *Criterion p is a* veto *criterion if for all $x, y \in \mathbb{R}^n$*

$$(x_{N\setminus\{p\}}, 0_p) \equiv^p_{\rm Veto} (y_{N\setminus\{p\}}, 0_p).$$

It is straightforward to see that the evaluation of the option $(x_{N\setminus\{p\}}, 0_p)$ is zero. This means that a bad score on a veto criterion cannot be saved by good scores on the remaining criteria.

7.2 Conditions on the fuzzy measure

Lemma 3 *Criterion $p \in N$ is a veto iff*

$$\forall S \subset N \setminus \{p\}\ , \quad \mu(S) = 0\ . \tag{9}$$

7.3 Completion

There is a unique additive fuzzy measure for which p is a veto. It is defined by $\mu(A) = 1$ if $p \in A$ and $= 0$ otherwise. Criterion p is then interpreted as a dictator. All the other criteria are merely discarded. Hence $\succsim^{\mathcal{M}^1}_{{\rm N},\Box^p_{\rm Veto}}$ is a complete order.

The following result shows that the necessity preference relation does not depend on k for $k \geq 2$.

Proposition 2 *Let P be the preferential information stating that criterion p is a veto. Then for all $k \in \{2, \ldots, n\}$*

$$\begin{aligned}
\succsim^{\mathcal{M}^k}_{{\rm N},\Box^p_{\rm Veto}} &= \Big\{(a,b) \in (\mathbb{R}^n)^2\ ,\ a_p \geq b_p \textit{ and} \\
&\quad \forall i \in N \setminus \{p\} \textit{ either } a_i \geq b_i \textit{ or } [a_i < b_i \textit{ and } a_i \geq b_p]\Big\} \\
\succ^{\mathcal{M}^k}_{{\rm N},\Box^p_{\rm Veto}} &= \Big\{(a,b) \in (\mathbb{R}^n)^2\ ,\ a_p > b_p \textit{ and} \\
&\quad \forall i \in N \setminus \{p\} \textit{ either } a_i > b_i \textit{ or } [a_i \leq b_i \textit{ and } a_i > b_p]\Big\} \\
\sim^{\mathcal{M}^k}_{{\rm N},\Box^p_{\rm Veto}} &= \Big\{(a,b) \in (\mathbb{R}^n)^2\ ,\ a_p = b_p \textit{ and} \\
&\quad \forall i \in N \setminus \{p\} \textit{ either } a_i = b_i \textit{ or } [a_i \geq a_p \textit{ and } b_i \geq a_p]\Big\}
\end{aligned}$$

The condition $a_i < b_i$ and $a_i \geq b_p$ means that the relatively bad score of a on criterion i is hidden by the score of b on the veto criterion p. Hence the necessity preference relation is completely natural. The completion mechanism of the Choquet integral is once more easily interpretable.

8 Conclusion

We have studied the completion mechanism produced by the Choquet integral on three types of preferential information: empty prior information, relative importance and veto. The result is trivial on the empty prior information since one recovers the Pareto ordering. Concerning the two other types of preferential information, we have shown that the Choquet integral does not introduce weird phenomena in the robust completion corresponding to the necessity preference relation. Moreover, the completion does not depend on order k for k-additive fuzzy measures, provided that $k \geq 2$. On the contrary, the weighted sum produced a strong completion.

References

[1] M. Grabisch, I. Kojadinovic, and P. Meyer. A review of capacity identification methods for Choquet integral based multi-attribute utility theory — applications of the Kappalab R package. *Eur. J. of Operational Research*, 186:766–785, 2008.

[2] I. Kojadinovic. Minimum variance capacity identification. *Quaterly J. of Operations Research (4OR)*, 12:23–36, 2006.

[3] Ch. Labreuche. Identification of a fuzzy measure with an l^1 entropy. In *Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, Malaga, Spain, June 22-27 2008.

[4] B. Roy. *Multicriteria methodology for decision aiding*. Kluwer Academic Publishers, Dordrecht, 1996.

[5] S. Greco, B. Matarazzo, and R. Słowinski. Ordinal regression revisited: Multiple criteria ranking with a set of additive value functions. *European Journal of Operational Research*, 191:416–436, 2008.

[6] S. Greco, R. Słowinski, J. Figueira, and V. Mousseau. Robust ordinal regression. In M. Ehrgott J. Figueira, S. Greco, editor, *New Trends in Multiple Criteria Decision Analysis*. Springer Verlag, to appear.

[7] S. Angilella, S. Greco, and B. Matarazzo. Non-additive robust ordinal regression: a multiple criteria decision model based on the choquet integral. *European Journal of Operational Research*, to appear.

[8] S. Greco, B. Matarazzo, and R. Słowinski. Decision rule approach. In M. Ehrgott J. Figueira, S. Greco, editor, *Multiple Criteria Decision Analysis - State of the Art Surveys*, pages 507–561. Springer International Series, 2005.

[9] C. Boutilier, R. Brafman, C. Domshlak, H. Hoos, and D. Poole. CP-nets: a tool for representing and reasoning with conditional Ceteris Paribus preference statements. *J. of Artificial Intelligence Research*, 21:135–191, 2004.

[10] M. Grabisch, T. Murofushi, and M. Sugeno. *Fuzzy Measures and Integrals. Theory and Applications (edited volume)*. Studies in Fuzziness. Physica Verlag, 2000.

[11] G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295, 1953.

[12] M. Grabisch. Alternative representations of discrete fuzzy measures for decision making. *Int. J. of Uncertainty, Fuzziness, and Knowledge Based Systems*, 5:587–607, 1997.

[13] L. S. Shapley. A value for n-person games. In H. W. Kuhn and A. W. Tucker, editors, *Contributions to the Theory of Games, Vol. II*, number 28 in Annals of Mathematics Studies, pages 307–317. Princeton University Press, 1953.

[14] T. Murofushi and S. Soneda. Techniques for reading fuzzy measures (III): interaction index. In *9th Fuzzy System Symposium*, pages 693–696, Sapporo, Japan, May 1993. In Japanese.

[15] A. Tsoukiàs. From decision theory to decision aiding methodology. *European Journal of Operational Research*, 187:138–161, 2008.

[16] S. Kaci and H. Prade. Constraints associated with choquet integrals and other aggregation-free ranking devices. In *Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, pages 1344–1351, Malaga, Spain, June 22-27 2008.

ISBN: 978-989-95079-6-8

Fuzzy Implications and the Weak Law of Importation

Sebastià Massanet Joan Torrens

Departament de Ciències Matemàtiques i Informàtica
Universitat de les Illes Balears
07122 Palma de Mallorca, Spain
Email: sebasmassanet@gmail.com, dmijts0@uib.es

Abstract— *Some open problems on fuzzy implications dealing with the so-called importation law are studied and totally or partially solved in this work. In particular, it is proved that such property (in fact a weaker version than the law of importation) is stronger than the exchange principle. Along this study, new characterizations of (S,N)-implications and R-implications involving the law of importation are showed.*

Keywords— Implication function, exchange principle, law of importation, (S,N)-implication, R-implication.

1 Introduction

Fuzzy implications are an essential tool in fuzzy control and approximate reasoning, as well as in many fields where these theories apply. This is because they are used not only to model fuzzy conditionals, but also to make inferences in any fuzzy rule based system (see for instance [1] or [2]) through the modus ponens and modus tollens. Moreover, fuzzy implications are also useful in many other fields like fuzzy relational equations and fuzzy mathematical morphology ([3]), fuzzy DI-subsethood measures and image processing ([4] and [5]), and data mining ([6]). Due to this great quantity of applications many authors have focused their interest in the theoretical study of fuzzy implications. See for instance the recent book [7], exclusively devoted to fuzzy implications, and the references therein.

From this theoretical study some open problems on fuzzy implications have been recently posed in some works about this topic. In this paper we want to deal with some of them involving the so-called law of importation, that is derived from the tautology in classical logic

$$(p \wedge q) \to r \equiv p \to (q \to r) \tag{1}$$

which, translated to fuzzy logic, becomes the functional equation:

$$I(T(x,y),z) = I(x,I(y,z)) \quad \text{for all } x,y,z \in [0,1] \tag{2}$$

where T is a t-norm and I is a fuzzy implication. This property has been extensively studied in [8] and [9] for many kinds of implications derived from t-norms and t-conorms. Moreover, an extension of (2) involving uninorms instead of t-norms has been also studied in [10]. From this study some open problems for fuzzy implications involving the law of importation have been posed (see [11]) that we will try to solve in next sections. During the process, new characterizations of (S,N)-implications and R-implications involving the law of importation are showed.

2 Preliminaries

We will suppose the reader to be familiar with the theory of t-norms, t-conorms and fuzzy negations (all necessary results and notations can be found in [12]). We recall here only some facts on implications.

Definition 1 *A binary operator $I : [0,1]^2 \to [0,1]$ is said to be an* implication function*, or an* implication*, if it satisfies:*

(I1) $I(x,z) \geq I(y,z)$ *when* $x \leq y$*, for all* $z \in [0,1]$.

(I2) $I(x,y) \leq I(x,z)$ *when* $y \leq z$*, for all* $x \in [0,1]$.

(I3) $I(0,0) = I(1,1) = 1$ *and* $I(1,0) = 0$.

Note that, from the definition, it follows that $I(0,x) = 1$ and $I(x,1) = 1$ for all $x \in [0,1]$ whereas the symmetrical values $I(x,0)$ and $I(1,x)$ are not derived from the definition.

Special interesting properties for implication functions are:

- The *law of importation* with a t-norm T,

$$I(T(x,y),z) = I(x,I(y,z)), \quad \text{for all} \quad x,y,z \in [0,1]. \tag{LI}$$

- The *exchange principle*,

$$I(x,I(y,z)) = I(y,I(x,z)), \quad \text{for all} \quad x,y,z \in [0,1]. \tag{EP}$$

- The *ordering property*,

$$x \leq y \iff I(x,y) = 1, \quad \text{for all} \quad x,y \in [0,1]. \tag{OP}$$

Among other models (see [3] and [7]), the most used fuzzy implications are derived from t-norms and t-conorms and they are:

- R-implications derived from a left-continuous t-norm T,

$$I_T(x,y) = \sup\{z \in [0,1] \mid T(x,z) \leq y\} \tag{3}$$

for all $x,y \in [0,1]$.

- (S,N)-implications derived from a t-conorm S and a strong negation[1] N,

$$I_{S,N}(x,y) = S(N(x),y) \qquad \text{for all} \quad x,y \in [0,1]. \tag{4}$$

[1] This kind of implications has been characterized also when N is strict or continuous, see [13]. However, (LI) has been studied only in the cases when N is strict or strong, but not when N is only continuous.

ISBN: 978-989-95079-6-8

But also QL-implications and D-implications respectively given by

$$I_{QL}(x,y) = S(N(x), T(x,y)) \tag{5}$$

for all $x, y \in [0,1]$, and

$$I_D(x,y) = S(T(N(x), N(y)), y) \tag{6}$$

for all $x, y \in [0,1]$, where T is a t-norm, S a t-conorm and N a strong negation. For all these kinds of implications the (LI) is equivalent to the (EP) and moreover, there is one and only one t-norm for which the (LI) holds (see [8] and [9]).

Finally let us recall the characterizations of (S,N) and R-implications. Several of these characterizations can be found in the literature:

Theorem 1 *([13]) For a function $I : [0,1]^2 \to [0,1]$ the following statements are equivalent:*

(i). I is an (S,N)-implication generated from some t-conorm S and some continuous (strict, strong) fuzzy negation N

(ii). I satisfies (I1), (EP) and $I(x,0) = N(x)$ is a continuous (strict, strong) fuzzy negation.

Theorem 2 *([14], [15] or [7]) Let $I : [0,1]^2 \to [0,1]$ be a function. Then I is an R-implication derived from a left-continuous t-norm, if and only if, I satisfies (I2), (OP), (EP) and I is right-continuous with respect to the second variable.*

With the assumption of continuity we have a characterization of the following subclass of R-implications, that are also (S,N)-implications, known as the Smets-Magrez Theorem, see [16] and see also [17] for the current version.

Theorem 3 *Let $I : [0,1]^2 \to [0,1]$ be a function. Then I is a continuous function satisfying (OP), (EP), if and only if, I is conjugate with the Łukasiewicz implication, that is, there exists a unique increasing bijection $\varphi : [0,1] \to [0,1]$ such that*

$$I(x,y) = \varphi^{-1}(\min\{1, 1 - \varphi(x) + \varphi(y)\}) \tag{7}$$

for all $x, y \in [0,1]$.

3 Main results

Due to the commutativity of the t-norm T (or the uninorm), it is clear that the law of importation implies the exchange principle, but in all known cases the converse also holds and the t-norm for which (LI) holds is uniquely determined. The first two open problems (posed in [11], Problem 8.1) that we want to study are just the following:

1. **(OP1)** Does the exchange principle imply that there exists a t-norm such that the law of importation holds? If not, give an example and characterize all fuzzy implications for which the answer is positive.

2. **(OP2)** Is the t-norm in the law of importation uniquely determined?

Note that for R, (S,N), QL and D-implications derived from t-norms and t-conorms the answer of both problems is positive (see [8] and [9]). With respect to (OP1) a negative answer has been done in [10] (see also [7], Section 7.3) by using implications derived from uninorms. The examples given in these works are implications satisfying (EP) such that they do not satisfy (LI) for any t-norm. However, in all these cases there exists a conjunctive uninorm for which the (LI) holds and thus, the problems above are just translated to the following more general versions:

1. **(OP1')** Does the exchange principle imply that there exists a uninorm (or a more general binary operation) such that the law of importation holds? If not, give an example and characterize all fuzzy implications for which the answer is positive.

2. **(OP2')** Is the uninorm in the law of importation uniquely determined?

Moreover, the answer of both problems is again positive for R, S, QL and D-implications derived from uninorms (see [10]).

Along the paper, we will prove that (OP1') has a negative answer in general and we will characterize some particular cases for which the answer is positive. On the other hand, with respect to (OP2) some counterexamples have been proved in [7] for implications I for which $I(x,0)$ equals the greatest or the least fuzzy negation (the same examples work for (OP2')). We will discuss however the case when $I(x,0)$ is at least a continuous fuzzy negation.

3.1 The Weak Law of Importation (WLI)

To deal with these problems let us introduce a weaker law of importation than (LI), by reducing the requirements on the function T. To maintain its relation with the exchange principle we need to take T commutative, and due to the monotonicity of implication functions it seems also adequate to take T nondecreasing, but no more conditions are necessary. Thus we introduce the following

Definition 2 *Let $I : [0,1]^2 \to [0,1]$ be a function. We say that I satisfies the* weak law of importation *if there exists a commutative and nondecreasing function $F : [0,1]^2 \to [0,1]$ such that*

$$I(F(x,y),z) = I(x, I(y,z)) \quad \textit{for all } x,y,z \in [0,1] \tag{WLI}$$

and then we say that I satisfies (WLI) with the function F.

First of all, note that (WLI) clearly implies (EP).

On the other hand, when I is an (S,N), a QL or a D-implication it satisfies the following boundary property

$$I(x,0) = N(x), \quad \text{for all} \quad x \in [0,1] \tag{BP}$$

where N is a (continuous, strict, strong) fuzzy negation.

We want to study the properties of those functions $I : [0,1]^2 \to [0,1]$ that satisfy (WLI) and (BP). These properties will be useful in order to find a counterexample to the equivalence of (EP) and (LI).

Proposition 1 *Let $I : [0,1]^2 \to [0,1]$ be a function which satisfies (WLI) with a commutative and nondecreasing function F and (BP) with a fuzzy negation N. Then*

ISBN: 978-989-95079-6-8

- *I satisfies (I3).*
- *I satisfies right contrapositive symmetry, that is,*

$$I(x, N(y)) = I(y, N(x)) \text{ for all } x, y \in [0,1]. \quad (8)$$

Moreover, if N is continuous, then

- *I satisfies (I1) and (I2), and consequently I is a fuzzy implication.*
- *I satisfies left neutrality principle, that is,*

$$I(1, y) = y \text{ for all } y \in [0,1]. \quad (9)$$

Finally, if N is strong then

- *I satisfies contrapositive symmetry, that is,*

$$I(N(y), N(x)) = I(x, y) \text{ for all } x, y \in [0,1]. \quad (10)$$

Proof: I satisfies (EP) due to the commutativity of F and (WLI). Furthermore, I satisfies also (I3) because

$$\begin{array}{ll} I(0,0) & = N(0) = 1 \\ I(1,0) & = N(1) = 0 \\ I(1,1) & = I(1, I(0,0)) = I(0, I(1,0)) = I(0,0) = 1. \end{array}$$

In addition, I satisfies right contrapositive symmetry because

$$I(x, N(y)) = I(x, I(y,0)) = I(y, I(x,0)) = I(y, N(x)).$$

From now on, N will be a continuous negation. Let us prove (I1). If $x \leq y$, using the continuity of N, we only need to show that $I(x, N(z)) \geq I(y, N(z))$ for all $z \in [0,1]$.

$$\begin{array}{ll} I(x, N(z)) & = I(x, I(z,0)) = I(F(x,z), 0) = N(F(x,z)) \\ & \geq N(F(y,z)) = I(F(y,z), 0) = I(y, I(z,0)) \\ & = I(y, N(z)) \end{array}$$

A similar argument shows (I2). If $y \leq z$, using the continuity of N, we will show that $I(x, N(y)) \geq I(x, N(z))$ for all $x, y, z \in [0,1]$.

$$\begin{array}{ll} I(x, N(y)) & = I(x, I(y,0)) = I(F(x,y), 0) = N(F(x,y)) \\ & \geq N(F(x,z)) = I(F(x,z), 0) = I(x, I(z,0)) \\ & = I(x, N(z)) \end{array}$$

Furthermore, I satisfies the left neutrality principle

$$\begin{array}{ll} I(1, N(y)) & = I(1, I(y,0)) = I(y, I(1,0)) \\ & = I(y,0) = N(y) \end{array}$$

Finally, if N is strong, we have

$$\begin{array}{ll} I(N(y), N(x)) & = I(N(y), I(x,0)) = I(x, I(N(y),0)) \\ & = I(x, N(N(y))) = I(x,y). \end{array}$$

■

It is known that in general, (EP) and (BP) with a continuous (even strong) fuzzy negation N do not imply (I1). For example, the function

$$I(x,y) = \begin{cases} 1-x & \text{if } y = 0 \\ y & \text{if } x = 1 \\ 0.5 & \text{otherwise} \end{cases} \quad (11)$$

satisfies (EP) and (BP) with $N(x) = 1 - x$ the classical strong negation, but it does not satisfy the nonincreasingness in the first variable, (I1) (see [13]). Consequently, I can not satisfy (WLI), by the previous proposition. So, we have just found a counterexample proving that, for functions $I : [0,1]^2 \to [0,1]$ in general, (EP) does not imply (WLI). In section 3.3 we will prove that a counterexample for fuzzy implications is also available.

3.2 (WLI) and (S,N)-implications

The main target of this section is the study of (WLI) on (S,N)-implications. We will show a new characterization of (S,N)-implications based on (WLI) and we will partially solve (OP2).

Proposition 2 *Let $I : [0,1]^2 \to [0,1]$ be a binary operator satisfying (WLI) with a commutative and nondecreasing function F and (BP) with a continuous (strict, strong) fuzzy negation N. Then I is an (S,N)-implication generated from a t-conorm S and the negation N.*

Proof: Due to proposition 1, it follows that I satisfies (I1) and (EP). So, by using theorem 1, I is an (S,N)-implication generated from some t-conorm S and some continuous (strict, strong) fuzzy negation N. ■

After that let us take a look to the verification of (WLI) on (S,N)-implications.

Proposition 3 *An (S,N)-implication I generated from a t-conorm S and a strict (strong) negation N satisfies (WLI) with a function $F : [0,1]^2 \to [0,1]$ if and only if F is the N-dual t-norm of S, that is*

$$F(x,y) = N^{-1}(S(N(x), N(y))) \text{ for all } x, y \in [0,1]. \quad (12)$$

Proof: Let I be an (S,N)-implication generated from a t-conorm S and a strict negation N. If I satisfies (WLI) with a function F, then for all $x, y, z \in [0,1]$ we have

$$\begin{array}{ll} I(F(x,y), z) & = S(N(F(x,y)), z) \\ I(x, I(y,z)) & = S(N(x), I(y,z)) = S(N(x), S(N(y), z)) \end{array}$$

Taking $z = 0$ in both expressions and using (WLI), we obtain $N(F(x,y)) = S(N(x), N(y))$, and consequently, $F(x,y) = N^{-1}(S(N(x), N(y)))$ for all $x, y \in [0,1]$. Conversely, if F is the N-dual t-norm of S, a straightforward computation proves that I satisfies (WLI) with F. ■

Remark 1 *Proposition 3 is the counterpart for (WLI) of an analogous result for (LI) in [8] and [13]. Note that none property of F is used in the proof. Thus, for (S,N)-implications with N strict or strong (LI) and (WLI) are equivalent and in fact, the N-dual t-norm of S is the only function $F : [0,1]^2 \to [0,1]$ (without any further assumption) for which (WLI) is satisfied.*

ISBN: 978-989-95079-6-8

We will see that the equivalence in the previous remark also holds for (S,N)-implications with N continuous (not necessarily strict), but something different occurs in this case. We will see that they also satisfy (WLI), but the function F for which they satisfy (WLI) needs not to be unique. However, among the functions F for which the (S,N)-implication satisfies (WLI), there is at least one of them that is a t-norm.

To do this, let us consider the following definition.

Definition 3 *(see [13]) Given any continuous fuzzy negation N, we consider the function $\mathfrak{R}_N : [0,1] \to [0,1]$ defined by*

$$\mathfrak{R}_N(x) = \begin{cases} N^{(-1)}(x) & \text{if } x \in (0,1] \\ 1 & \text{if } x = 0 \end{cases} \tag{13}$$

where $N^{(-1)}$ stands for the pseudo-inverse of N given by

$$N^{(-1)}(x) = \sup\{z \in [0,1] \mid N(z) > x\} \quad \text{for all } x \in [0,1]. \tag{14}$$

In [13], it is proved that $\mathfrak{R}_N$ is a strictly decreasing negation, $\mathfrak{R}_N^{(-1)} = N$, $N \circ \mathfrak{R}_N = id_{[0,1]}$ and

$$\mathfrak{R}_N \circ N|_{\text{Ran}(\mathfrak{R}_N)} = id|_{\text{Ran}(\mathfrak{R}_N)},$$

where $\text{Ran}(\mathfrak{R}_N)$ stands for the range of function $\mathfrak{R}_N$. This function $\mathfrak{R}_N$ plays the role of the "inverse" function of N and therefore, we have the following result:

Proposition 4 *An (S,N)-implication I generated from a t-conorm S and a continuous fuzzy negation N satisfies (WLI) with the function $F(x,y) = \mathfrak{R}_N(S(N(x),N(y)))$ for all $x,y \in [0,1]$.*

Proof: It is clear from its definition that F is commutative and nondecreasing. Now,

$$\begin{array}{ll} I(T(x,y),z) & = S(N(T(x,y)),z) \\ & = S(N(\mathfrak{R}_N(S(N(x),N(y)))),z) \\ (N \circ \mathfrak{R} = id_{[0,1]}) & = S(S(N(x),N(y)),z) \\ & = S(N(x),S(N(y),z)) \\ & = S(N(x),I(y,z)) = I(x,I(y,z)) \end{array}$$

■

Contrary to what happens in the case when N is strict (or strong), the function F in the proposition above is not the only one for which the (S,N)-implication satisfies (WLI). Note that the key fact in the proof of the previous proposition is that $N \circ \mathfrak{R}_N = id_{[0,1]}$ (in fact it is only necessary that $N(F(x,y)) = S(N(x),N(y))$ for all $x,y \in [0,1]$). When N is continuous but non-strict there are other functions than $\mathfrak{R}_N$ satisfying this condition, and consequently different functions F for which I satisfies the (WLI). See for instance the following example.

Example 1 *(see [13]) Consider the continuous fuzzy negation N given by*

$$N(x) = \begin{cases} -2x+1 & \text{if } x \in [0,0.25] \\ 0.5 & \text{if } x \in (0.25,0.75) \\ -2x+2 & \text{if } x \in [0.75,1]. \end{cases} \tag{15}$$

An easy calculation shows that

$$\mathfrak{R}_N(x) = \begin{cases} -0.5x+1 & \text{if } x \in [0,0.5) \\ -0.5x+0.5 & \text{if } x \in [0.5,1], \end{cases} \tag{16}$$

but taking N_1 given by

$$N_1(x) = \begin{cases} -0.5x+1 & \text{if } x \in [0,0.5] \\ -0.5x+0.5 & \text{if } x \in (0.5,1], \end{cases} \tag{17}$$

we also obtain $N \circ N_1 = id_{[0,1]}$. Consequently, given I any (S,N)-implication, derived from a t-conorm S and the negation N, defining $F_1(x,y) = N_1(S(N(x),N(y)))$ we have that I satisfies (WLI) with both functions, F (obtained from Proposition 4) and F_1.

Proposition 5 *Let S be a t-conorm and N a continuous fuzzy negation. Then the function $F(x,y) = \mathfrak{R}_N(S(N(x),N(y)))$ for all $x,y \in [0,1]$ is a t-subnorm, i.e., it is nondecreasing, commutative, associative and such that $F(x,y) \leq \min(x,y)$ for all $x,y \in [0,1]$.*

Proof: F is clearly nondecreasing and commutative. To see associativity, note that from the associativity of S we easily obtain

$$\begin{array}{ll} F(F(x,y),z) & = \mathfrak{R}_N(S(S(N(x),N(y)),N(z)) \\ & = F(x,F(y,z)). \end{array}$$

Finally, using the definition of $\mathfrak{R}_N$ we easily obtain

$$\mathfrak{R}_N(S(N(x),N(y))) \leq x \text{ for all } x \in [0,1]$$

and so $F(x,y) \leq \min(x,y)$ for all $x,y \in [0,1]$. ■

However, they are not t-norms in general because we have

$$F(1,y) = \mathfrak{R}_N(S(0,N(y))) = \mathfrak{R}_N(N(y))$$

and $\mathfrak{R}_N \circ N$ is the identity function only for values in $\text{Ran}(\mathfrak{R}_N)$. Nevertheless, one can always find a t-norm for which the (S,N)-implication satisfies (WLI) as follows.

Proposition 6 *Let I be a (S,N)-implication generated from a t-conorm S and a continuous fuzzy negation N. Then*

$$T(x,y) = \begin{cases} \mathfrak{R}_N(S(N(x),N(y))) & \text{if } \max\{x,y\} < 1 \\ \min\{x,y\} & \text{if } \max\{x,y\} = 1 \end{cases}$$

is a t-norm such that I satisfies (LI) with T.

Proof: T is clearly a t-norm since it is constructed in the usual way to obtain a t-norm from a given t-subnorm (see for instance, [12], Proposition 1.6). Thus, we only need to prove that I satisfies (LI) with T. From Proposition 4 it is sufficient to verify (LI) when $x = 1$ or $y = 1$. But, for $x = 1$ we have

$$I(T(1,y),z) = I(y,z) = I(1,I(y,z))$$

and similarly for $y = 1$. ■

Corollary 1 *Let I be an implication satisfying (BP) with a continuous fuzzy negation N. Then*

$$I \text{ satisfies (LI)} \quad \Leftrightarrow \quad I \text{ satisfies (WLI)}$$

Remark 2 *Note that the construction given in the Propositions 5 and 6 works only for the function $\mathfrak{R}_N$. For the other functions N_1 satisfying $N \circ N_1 = id_{[0,1]}$ the monotonicity of T is not guaranteed (it is enough to take $S = \max$, N and N_1 as in Example 1 and $x = 0.5 > y$, then $N_1(\max(N(0.5),N(y))) = 0.75 > 0.5$). Thus, for (S,N)-implications with N continuous we have found many functions F for which (WLI) holds, but only one t-norm for which (LI) holds. However, we are convinced that there can be found t-conorms S such that even the t-norm is not unique.*

ISBN: 978-989-95079-6-8

At this point, we propose a new characterization of (S,N)-implications based on (WLI).

Theorem 4 *For a function $I:[0,1]^2\to[0,1]$ the following statements are equivalent:*

(i). I is an (S,N)-implication generated from a t-conorm S and a continuous (strict, strong) fuzzy negation N

(ii). I satisfies (WLI) with a commutative and nondecreasing function F and (BP) with a continuous (strict, strong) fuzzy negation N.

In this case possible functions F are given by $F(x,y)=N_1(I(x,N(y)))$ for all $x,y\in[0,1]$, where N_1 is a fuzzy negation with $N\circ N_1=id_{[0,1]}$. Moreover, when N is strict (strong) then such F is unique with $N_1=N^{-1}$.

Proof: If I satisfies $i)$, then $I(x,0)=S(N(x),0)=N(x)$ and I satisfies (BP) with N. Moreover, it also satisfies (WLI) from Proposition 4. The converse is guaranteed by Proposition 2. ∎

Finally and returning to the question about the equivalence of (EP) and (LI), we have the following result.

Proposition 7 *Let I be an implication satisfying (BP) with a continuous fuzzy negation N. Then*

$$I \text{ satisfies (EP)} \Leftrightarrow I \text{ satisfies (LI)} \Leftrightarrow I \text{ satisfies (WLI)}$$

Proof: We already know that (LI) implies (EP). To see the converse, note that if I is an implication it satisfies in particular (I1) and then Theorem 1 implies that I is an (S,N)-implication generated from a t-conorm S and a continuous fuzzy negation N. Finally, Proposition 4 and Corollary 1 ensures that I satisfies (LI). The other equivalence has been already proved in Corollary 1. ∎

3.3 Counterexample for implications

To sum up, we have proved that (EP) and (LI) are equivalent for implications satisfying (BP) with a continuous fuzzy negation N. So, to find a counterexample we need to search among those implications I such that $I(x,0)=N(x)$ is non-continuous (it is always a fuzzy negation). In next proposition we give such a counterexample which proves not only that (EP) does not imply (LI), but also that it does not imply (WLI).

Proposition 8 *Let S be a nilpotent t-conorm and N a strict negation. Let $I:[0,1]^2\to[0,1]$ be the function given by*

$$I(x,y)=\begin{cases}0 & \text{if } y=0 \text{ and } x\neq 0\\ S(N(x),y) & \text{otherwise.}\end{cases} \tag{18}$$

Then I is a fuzzy implication that satisfies (EP), but there is no function F for which I satisfies (WLI).

Proof: We divide the proof in two steps.

- I satisfies (EP). When $z\neq 0$ we have $I(x,z)\neq 0$ for all $x\in[0,1]$ and then I satisfies (EP) because it is given as an (S,N)-implication.

  When $z=0$ and $x,y\neq 0$ then

  $$I(x,I(y,0))=I(y,I(x,0))=0.$$

  Finally, when $z=0$ and $\min\{x,y\}=0$ then

  $$I(x,I(y,0))=I(y,I(x,0))=1.$$

- I does not satisfy (WLI) with any F. Suppose on the contrary that I satisfies (WLI) with a function F. Then for all $x,y\in[0,1]$ and $z\neq 0$ (WLI) derives into

  $$S(N(F(x,y)),z)=S(S(N(x),N(y)),z). \tag{19}$$

  If we denote by N_S the strong negation associated to S (i.e., $N_S(x)=\varphi^{-1}(1-\varphi(x))$ where φ is the normalized additive generator of S), the left hand side of the previous equation equal 1 if and only if $N(F(x,y))\geq N_S(z)$, whereas the right hand side equals 1 if and only if $S(N(x),N(y))\geq N_S(z)$. Consequently for all $x,y\in[0,1]$ and $z\neq 0$

  $$N(F(x,y))\geq N_S(z)\Longleftrightarrow S(N(x),N(y))\geq N_S(z).$$

  Let us now prove from this equivalence that

  $$F(x,y)=N^{-1}(S(N(x),N(y)) \quad \text{for all } x,y\in[0,1]. \tag{20}$$

  Note that when $F(x,y)=0$, $N(F(x,y))\geq N_S(z)$ for all $z\in[0,1]$ and, from the equivalence before we obtain $S(N(x),N(y))=1$ and equation (20) holds. On the other hand, when $F(x,y)>0$ there exists $z\in[0,1]$ such that $N_S(z)>N(F(x,y))$ and then equation (19) can be written as

  $$\begin{aligned}&\varphi^{-1}(\varphi(N(F(x,y)))+\varphi(z))\\ &=\varphi^{-1}(\varphi(S(N(x),N(y)))+\varphi(z))\end{aligned}$$

  and consequently equation (20) also holds.

  Thus, F should be given by equation (20). However, I does not satisfy (WLI) with this F when $z=0$. To see it just take $x,y\neq 0$ such that $S(N(x),N(y))=1$ since then $F(x,y)=0$ and we obtain

  $$\begin{aligned}I(x,I(y,0)) &= I(x,0)=0\\ I(F(x,y),0) &= I(0,0)=1\end{aligned}$$

  ∎

3.4 R-implications derived from left-continuous t-norms

A characterization of those functions $I:[0,1]^2\to[0,1]$ that are R-implications derived from a left-continuous t-norm is given in Theorem 2, and the case of continuous R-implications is characterized in Theorem 3. Note that in this last case condition (I2) is not necessary to derive the conclusion. It was claimed in [17] that the same happens in the general case. Specifically it was claimed that (EP) and (OP) imply (I2). This is not true and a counterexample was recently proved in [18].

In this section, we will prove that the situation is completely different if we change (EP) by (WLI). We begin by proving that (I2) can be derived from (WLI) and (OP).

Proposition 9 *Let $I:[0,1]^2\to[0,1]$ be a function satisfying (OP) and (WLI) with a commutative and nondecreasing function F. Then I satisfies (I2), i.e, it is nondecreasing in the second variable.*

ISBN: 978-989-95079-6-8

Proof: Consider $x, y, z \in [0,1]$ with $y \leq z$ and let us divide our argument in two cases.

- If $x \leq z$, then $I(x,y) \leq I(x,z) = 1$.
- If $y \leq z < x$, then by (OP) and (WLI),

  $$1 = I(I(x,y), I(x,y)) = I(F(I(x,y),x), y)$$

  and consequently $F(I(x,y),x) \leq y \leq z$. Thus

  $$1 = I(F(I(x,y),x), z) = I(I(x,y), I(x,z))$$

  and therefore, $I(x,y) \leq I(x,z)$. ∎

Moreover, we have also the following properties.

Proposition 10 *Let $I : [0,1]^2 \to [0,1]$ be a function satisfying (OP) and (WLI) with a commutative and nondecreasing function F. Then*

i) I satisfies (I1), i.e, it is nonincreasing in the first variable.

ii) $I(x,-)$ is right-continuous for all $x \in [0,1]$.

Proof: To prove i) consider $x, y, z \in [0,1]$ with $x \leq y$. We divide our argument in two cases.

- If $x \leq z$, then $I(y,z) \leq I(x,z) = 1$.
- If $x > z$, then by (OP) and (WLI),

  $$1 = I(I(y,z), I(y,z)) = I(F(I(y,z),y), z)$$

  and consequently $F(I(y,z),y) \leq z$. Since F is nondecreasing we also have $F(I(y,z),x) \leq z$. Thus

  $$1 = I(F(I(y,z),x), z) = I(I(y,z), I(x,z))$$

  and therefore, $I(y,z) \leq I(x,z)$.

To prove ii), since I is nondecreasing in the second variable, it is enough to show that $\inf\{I(x,y_n)\} = I(x, \inf\{y_n\})$ for all $x \in [0,1]$. Say $y = \inf\{y_n\}$, it is clear that $I(x,y) \leq I(x,y_n)$ for all n and consequently $I(x,y) \leq \inf\{I(x,y_n)\}$. Moreover, for all n we have

$$1 = I(I(x,y_n), I(x,y_n)) = I(F(I(x,y_n),x), y_n)$$

which implies $F(I(x,y_n),x) \leq y_n$ for all n and thus $F(I(x,y_n),x) \leq y$. Finally, we have

$$1 = I(F(I(x,y_n),x), y)) = I(I(x,y_n), I(x,y)),$$

that is, $I(x,y_n) \leq I(x,y)$ and thus $\inf\{I(x,y_n)\} \leq I(x,y)$.∎

Due to the previous results we can reformulate the characterization of R-implications given by theorem 2.

Theorem 5 *Let $I : [0,1]^2 \to [0,1]$ be a function. Then I is an R-implication derived from a left-continuous t-norm, if and only if, I satisfies (OP) and (WLI) with a commutative and nondecreasing function F.*

Moreover, in this case F must be the t-norm from which I is obtained by residuation.

Finally, it is well know that if a function $I : [0,1]^2 \to [0,1]$ satisfies (OP), (EP) and (BP) with a continuous fuzzy negation N, then N must be strong. However, note that from the previous results more properties are derived.

Corollary 2 *Let $I : [0,1]^2 \to [0,1]$ satisfy (OP), (EP) and (BP) with a continuous fuzzy negation N. Then I satisfies (I2), I is right-continuous in the second variable and so, I is an R-implication derived from a left-continuous t-norm.*

Acknowledgment

The authors want to thank the referees for their valuable comments. This paper has been partially supported by the Spanish grant MTM2006-05540 (with FEDER support), and the Government of the Balearic Islands grant PCTIB-2005GC1-07.

References

[1] E.E. Kerre, C. Huang, and D. Ruan. *Fuzzy Set Theory and Approximate Reasoning*. Wu Han University Press, Wu Chang, 2004.

[2] S. Gottwald. *A Treatise on Many-Valued Logic*. Research Studies Press, Baldock, 2001.

[3] M. Mas, M. Monserrat, J. Torrens, and E. Trillas. A survey on fuzzy implication functions. *IEEE Transactions on Fuzzy Systems*, 15(6):1107–1121, 2007.

[4] H. Bustince, V. Mohedano, E. Barrenechea, and M. Pagola. Definition and construction of fuzzy DI-subsethood measures. *Information Sciences*, 176:3190–3231, 2006.

[5] H. Bustince, M. Pagola, and E. Barrenechea. Construction of fuzzy indices from fuzzy DI-subsethood measures: application to the global comparison of images. *Information Sciences*, 177:906–929, 2007.

[6] P. Yan and G. Chen. Discovering a cover set of ARsi with hierarchy from quantitative databases. *Information Sciences*, 173:319–336, 2005.

[7] M. Baczyński and B. Jayaram. *Fuzzy Implications*. Springer, Berlin Heidelberg, 2008.

[8] B. Jayaram. On the law of importation $(x \wedge y) \to z \equiv (x \to (y \to z))$ in fuzzy logic. *IEEE Transactions on Fuzzy Systems*, 16:130–144, 2008.

[9] M. Mas, M. Monserrat, and J. Torrens. The law of importation for QL and D-implications. *Submitted*, 2008.

[10] M. Mas, M. Monserrat, and J. Torrens. On the law of importation for some kinds of fuzzy implications derived from uninorms. *In Proceedings of IPMU'08*, Torremolinos (Málaga):1303–1310, 2008.

[11] E.P. Klement and R. Mesiar. Open problems posed at the eighth international conference in fuzzy set theory and applications (FSTA 2006, Liptovský Ján, Slovakia). *Kybernetika*, 42:225–235, 2006.

[12] E.P. Klement, R. Mesiar, and E. Pap. *Triangular norms*. Kluwer Academic Publishers, Dordrecht, 2000.

[13] M. Baczyński and B. Jayaram. On the characterization of (S,N)-implications. *Fuzzy Sets and Systems*, 158:1713–1727, 2007.

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

[15] M. Baczyński and B. Jayaram. (S,N)- and R-implications: A state-of-the-art survey. *Fuzzy Sets and Systems*, 159:1836–1859, 2008.

[16] P. Smets and P. Magrez. Implications in fuzzy logic. *International Journal of Approximate Reasoning*, 1:327–347, 1987.

[17] M. Baczyński. Residual implications revisited. notes on the smets-magrez theorem. *Fuzzy Sets and Systems*, 145:267–277, 2004.

[18] R. Smogura. A note on the article: Residual implications revisited. Notes on the Smets-Magrez theorem. *Fuzzy Sets and Systems*, 157:466–469, 2006.

ISBN: 978-989-95079-6-8

Fuzzy Transform and Smooth Functions

Luciano Stefanini
DEMQ, Department of Economics and Quantitative Methods
University of Urbino "Carlo Bo", ITALY
Email: lucste@uniurb.it

Abstract— *We describe parametric fuzzy partitions based on parametric shape functions and we show that the Perfilieva's Fuzzy Transform defined on parametric fuzzy partitions has the properties of a smoother. We also illustrate some examples.*

Keywords— Parametric Fuzzy Partition, Fuzzy Transform, Smoothing Functions, Fuzzy Numbers

1 Introduction

The fuzzy transform has been recently introduced by I. Perfilieva and the theoretical investigations and several connections between theory and applications are covered e.g. in [3], [4], [5] and [9].

In a recent paper (see [6]), we suggested the use of a parametrized class of fuzzy numbers (see [7], [8]) to define general parametric partitions for the fuzzy transform. The proposed flexible family of fuzzy partitions allows the possibility of estimating the shapes of the fuzzy partition to obtain better approximation properties.

In this paper we suggest an extension of the fuzzy partitions (section 3) and we describe the corresponding direct and inverse fuzzy transforms (section 4). Then, in section 5, we show that the inverse fuzzy transform based on extended partitions has an important smoothing property and we illustrate it by numerical examples.

2 Fuzzy Transform and Parametric Fuzzy Partitions

We briefly recall the basic definitions and properties

Definition: (Perfilieva[3]) A *fuzzy partition* of a given real compact interval $[a,b]$ is constructed by a decomposition $\mathbb{P} = \{a = x_1 < x_2 < ... < x_n = b\}$ of $[a,b]$ into $n-1$ subintervals $[x_{k-1}, x_k]$, $k = 2, ..., n$ and by a family $\mathbb{A} = \{A_1, A_2, ..., A_n\}$ of n fuzzy numbers (the *basic functions*), identified by their membership functions $A_1(x), A_2(x), ..., A_n(x)$ for $x \in [a,b]$ and with the properties (to complete this notation, we set $x_0 = a$ and $x_{n+1} = b$):

1. $A_k : [a,b] \longrightarrow [0,1]$ is continuous with $A_k(x_k) = 1$, $A_k(x) = 0$ for $x \notin [x_{k-1}, x_{k+1}]$;
2. for $k = 2, 3, ..., n$, A_k is increasing on $[x_{k-1}, x_k]$ and decreasing on $[x_k, x_{k+1}]$; A_1 is decreasing on $[a, x_2]$; A_n is increasing on $[x_{n-1}, b]$;
3. for all $x \in [a,b]$ the following partition of unity condition holds

$$\sum_{k=1}^{n} A_k(x) = 1.$$

We denote a fuzzy partition by the pair $(\mathbb{P}, \mathbb{A})$. The fuzzy numbers $A_1, A_2, ..., A_n$ giving the family of basic functions can be defined by $n-1$ increasing functions $L_2(x), ..., L_n(x)$ as follows

$$A_1(x) = 1 - L_2(x) \text{ if } x \in [a, x_2],$$

$$A_k(x) = \begin{cases} L_k(x) & \text{if} \quad x \in [x_{k-1}, x_k] \\ 1 - L_{k+1}(x) & \text{if} \quad x \in [x_k, x_{k+1}] \end{cases} \quad \text{for } k = 2, ..., n-1 \tag{1}$$

$$A_n(x) = L_n(x) \text{ if } x \in [x_{n-1}, b]$$

where $L_k(x_{k-1}) = 0$ and $L_k(x_k) = 1$.

In the standard case, the support of each basic function $A_k(x)$ is the compact interval $[x_{k-1}, x_{k+1}]$ so that, on each subinterval $[x_{k-1}, x_k]$ of the decomposition $\mathbb{P}$ only two basic functions $A_{k-1}(x)$ and $A_k(x)$ are non zero for $k = 2, ..., n$.

Definition 1: (Perfilieva[3]) Given a continuous function $f : [a,b] \longrightarrow R$ and a fuzzy partition (P, A), the direct fuzzy transform (F-transform) of f with respect to (P, A) is the n-tuple of real numbers $F = (F_1, F_2, ..., F_n)^T$ given by (notation $(\cdot)^T$ means transposition)

$$\text{For } k = 1, 2, ..., n \tag{2}$$

$$F_k = \frac{\int_a^b f(x)A_k(x)dx}{\int_a^b A_k(x)dx} = \frac{\int_{x_{k-1}}^{x_{k+1}} f(x)A_k(x)dx}{\int_{x_{k-1}}^{x_{k+1}} A_k(x)dx}$$

Definition 2: (Perfilieva[3]) Given a continuous function $f : [a,b] \longrightarrow R$, a fuzzy partition (P, A) and the direct fuzzy transform $(F_1, F_2, ..., F_n)^T$ of f with respect to (P, A), the inverse F-transform is the continuous function $\widehat{f}_F : [a,b] \longrightarrow R$ given by

$$\widehat{f}_F(x) = \sum_{k=1}^{n} F_k A_k(x) \text{ for } x \in [a,b]. \tag{3}$$

The following property (Th. 2 in Perfilieva[3]) is one of the fundamentals of the F-transform setting.

Property 1: If $f : [a,b] \longrightarrow R$ is a continuous function then, for any positive real ε, there exists a fuzzy partition $(P_\varepsilon, A_\varepsilon)$ such that the corresponding F-transform $F_\varepsilon = (F_{1,\varepsilon}, F_{2,\varepsilon}, ..., F_{n_\varepsilon,\varepsilon})^T$ and inverse fuzzy transform $\widehat{f}_{F_\varepsilon} : [a,b] \longrightarrow R$ satisfy

$$\left| f(x) - \widehat{f}_{F_\varepsilon}(x) \right| < \varepsilon \text{ for all } x \in [a,b].$$

To define a fuzzy partition $(\mathbb{P}, \mathbb{A})$ we can use a general decomposition $\mathbb{P} = \{a = x_1 < x_2 < ... < x_n = b\}$ of $[a,b]$ into $n-1$ subintervals $[x_{k-1}, x_k]$, and a general set of basic functions $\mathbb{A} = \{A_1, A_2, ..., A_n\}$ of n fuzzy numbers

ISBN: 978-989-95079-6-8

$A_1, A_2, ..., A_n$ on $[a, b]$; denote, in general, $h_k = x_k - x_{k-1}$, $k = 2, ..., n$.

If the decomposition $\mathbb{P}$ is uniform, we have $x_k = a + (k-1)h$ where $h = \frac{b-a}{n-1}$.

We can have uniform $\mathbb{P}$ and general $\mathbb{A}$.

To obtain families of basic functions $A_1, A_2, ..., A_n$ in (1), defined by $n-1$ increasing functions $L_2, ..., L_n$, we suggested to use parametric monotonic functions $p(t; \beta_0, \beta_1)$, $t \in [0, 1]$ such that, for any parameters $\beta_0, \beta_1 \geq 0$, the following Hermite-type interpolation conditions are satisfied:

$$\begin{aligned} p(0; \beta_0, \beta_1) &= 0,\ p(1; \beta_0, \beta_1) = 1, \\ p'(0; \beta_0, \beta_1) &= \beta_0,\ p'(1; \beta_0, \beta_1) = \beta_1 \end{aligned}$$

and $p'(t; \beta_0, \beta_1) \geq 0\ \forall t \in [0, 1]$ (derivatives at $t = 0$ and $t = 1$ are, respectively, right and left derivatives).

Valid examples of function $p(t; \beta_0, \beta_1)$ are reported in see [7]; to exemplify, we will use here the (2,2)-rational spline:

$$p_{rat22}(t; \beta_0, \beta_1) = \frac{t^2 + \beta_0 t(1-t)}{1 + (\beta_0 + \beta_1 - 2)t(1-t)}. \quad (4)$$

By using parametric monotonic functions $p(t; \beta_0, \beta_1)$, $t \in [0, 1]$ we obtain increasing functions $L_2, ..., L_n$ that define the basic membership functions of the fuzzy partition $A_1, A_2, ..., A_n$:

$$\begin{aligned} L_k(x) &= p(\frac{x - x_{k-1}}{x_k - x_{k-1}}; \beta_{k,0}, \beta_{k,1}) \quad (5) \\ \text{for } x &\in [x_{k-1}, x_k] \text{ and } k = 2, ..., n \end{aligned}$$

and each fuzzy partition $(\mathbb{P}, \mathbb{A})$ contains $2(n-1)$ parameters $\beta_{k,0}, \beta_{k,1}$, $k = 2, ..., n$ giving the first derivatives $L'_k(x_{k-1}) = \beta_{k,0}/(x_k - x_{k-1})$ and $L'_k(x_k) = \beta_{k,1}/(x_k - x_{k-1})$ (to be intended as right and left derivatives, respectively).

We define a *uniform fuzzy partition* $(\mathbb{P}, \mathbb{A})$ if $\mathbb{P}$ is uniform and $\beta_{k,0} = \beta_{j,0}, \beta_{k,1} = \beta_{j,1}$ for all j, k.

We define a *balanced fuzzy partition* $(\mathbb{P}, \mathbb{A})$ if $\mathbb{P}$ is uniform and $\beta_{k,0} = \beta_{k+r,0}, \beta_{k,1} = \beta_{k+r,1}$ for all k..

We obtain a uniform and balanced partition $(\mathbb{P}, \mathbb{A})$ by requiring that $\beta_{k,0} = \beta_{k,1} = \beta_{j,0} = \beta_{j,1}$ for all j, k.

3 Extended Fuzzy Partitions

On the interval $[a, b]$ with a decomposition $\mathbb{P} = \{a = x_1 < x_2 < ... < x_n = b\}$ into $n-1$ subintervals $[x_{k-1}, x_k]$, $k = 2, ..., n$ the basic functions of the fuzzy partition can be defined such that the support of each A_k is more extended than two intervals, identified by three points x_{k-1}, x_k, x_{k+1}. Consider an integer $r \geq 1$ and $2r+1$ consecutive points of $\mathbb{P}$, $x_{k-r}, ..., x_k, ..., x_{k+r}$ for all $k = 1, 2, ..., n$ (to complete the notation, we extend the points to $x_{1-r} < ... < x_0 < a$ and $b < x_{n+1} < ... < x_{n+r}$).

Definition 3: we define a *fuzzy r-partition* of $[a, b]$ to be $(\mathbb{P}, \mathbb{A}^{(r)})$ with the family of $n + 2r - 2$ continuous, normal, convex fuzzy numbers

$$\begin{aligned} \mathbb{A}^{(r)} &= \{A_k^{(r)} : [a, b] \longrightarrow [0, 1] | \\ k &= -r+2, ..., 1, 2, ..., n, ..., n+r-1\} \end{aligned}$$

where $A_k^{(r)}$ has core $\{x_k\}$ and support $[x_{k-r}, x_{k+r}]$ i.e.

a. for $k = 1, 2, ..., n$, $A_k^{(r)}$ is a continuous fuzzy number with $A_k^{(r)}(x_k) = 1$ (i.e. the core is $\{x_k\}$) and $A_k^{(r)}(x) = 0$ for $x \notin [x_{k-r}, x_{k+r}]$ (i.e. the support is $[x_{k-r}, x_{k+r}]$);

b. for $k = 1, 2, ..., n$, $A_k^{(r)}$ is increasing on $[x_{k-r}, x_k]$ and decreasing on $[x_k, x_{k+r}]$;

c. for $k = -r+2, ..., 0$, $A_k^{(r)}$ is decreasing on $[x_k, x_{k+r}]$;

d. for $k = n+1, ...n+r-1$, $A_k^{(r)}$ is increasing on $[x_{k-r}, x_k]$;

e. for all $x \in [a, b]$ the following *$r-$partition of unity* condition holds

$$\frac{1}{r} \sum_{k=-r+2}^{n+r-1} A_k^{(r)}(x) = 1.$$

If $r = 1$ we have the standard partition of unity.

We can construct parametric *fuzzy r-partition* of $[a, b]$ by considering $n + r - 2$ shape functions $p(t; \beta_0, \beta_1)$ of types (4); then by defining

$$\begin{aligned} \text{for } k &= 2, ..., n+r-1 \\ L_k^{(r)}(x) &= p(\frac{x - x_{k-r}}{x_k - x_{k-r}}; \beta_{k,0}, \beta_{k,1}) \end{aligned}$$

the basic functions are

$$\begin{aligned} \text{for } k &= 2, ..., n-1 \\ A_k^{(r)}(x) &= \begin{cases} L_k^{(r)}(x) & \text{if } x \in [x_{k-r}, x_k] \\ 1 - L_{k+r}^{(r)}(x) & \text{if } x \in [x_k, x_{k+r}] \\ 0 & \text{otherwise} \end{cases} \quad (6) \end{aligned}$$

$$\begin{aligned} \text{for } k &= -r+2, ..., 1 \\ A_k^{(r)}(x) &= \begin{cases} 1 - L_{k+r}^{(r)}(x) & \text{if } x \in [x_k, x_{k+r}] \\ 0 & \text{otherwise} \end{cases} \quad (7) \end{aligned}$$

$$\begin{aligned} \text{for } k &= n, ..., n+r-1 \\ A_k^{(r)}(x) &= \begin{cases} L_{k-r}^{(r)}(x) & \text{if } x \in [x_{k-r}, x_k] \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

It is a matter of simple algebraic manipulations to show that for all $x \in [a, b]$ the *$r-$partition of unity* condition holds; in fact for $k = 1, 2, ..., n$ and $x \in [x_{k-1}, x_k]$, only $A_{k-r}^{(r)}, ..., A_k^{(r)}, ..., A_{k+r-1}^{(r)}$ are different from zero and

$$\begin{aligned} \sum_{j=k-r}^{k+r-1} A_j^{(r)}(x) &= \sum_{j=k-r}^{k-1} (1 - L_{j+r}^{(r)}(x)) + \sum_{j=k}^{k+r-1} L_j^{(r)}(x) \\ &= \sum_{j=k}^{k+r-1} (1 - L_j^{(r)}(x)) + \sum_{j=k}^{k+r-1} L_j^{(r)}(x) \\ &= r. \end{aligned}$$

Consider for example $r = 2$; we need $n + r - 2 = n$ shape functions and (we remove the superscript $^{(r)}$ for simplicity) define $L_k(x) = p(\frac{x - x_{k-2}}{x_k - x_{k-2}}; \beta_{k,0}, \beta_{k,1})$ for $k = 2, ..., n+1$. So (in the relevant subintervals, otherwise the functions are null)

$$\begin{aligned} \text{for } k &= 2, ..., n-1 \\ A_k(x) &= \begin{cases} L_k(x) & \text{if } x \in [x_{k-2}, x_k] \\ 1 - L_{k+2}(x) & \text{if } x \in [x_k, x_{k+2}] \end{cases} \end{aligned}$$

ISBN: 978-989-95079-6-8

$$\begin{aligned} \text{for } k &= 0, 1, n, n+1 \\ A_0(x) &= 1 - L_2(x) \text{ if } x \in [x_0, x_2] \\ A_1(x) &= 1 - L_3(x) \text{ if } x \in [x_1, x_3] \\ A_n(x) &= L_n(x) \text{ if } x \in [x_{n-2}, x_n] \\ A_{n+1}(x) &= L_{n+1}(x) \text{ if } x \in [x_{n-1}, x_{n+1}]. \end{aligned}$$

The figures below illustrate three *fuzzy r-partition* $(\mathbb{P}, \mathbb{A}^{(r)})$ for $r = 1$ (standard case), $r = 2$ (each $A_k(x)$ covers four intervals) and $r = 3$ (each $A_k(x)$ covers six intervals).

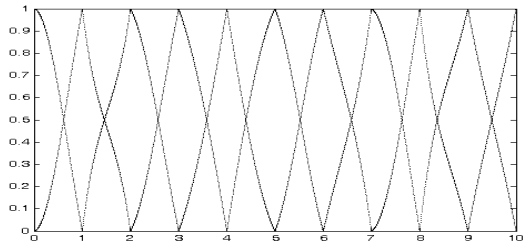


Figure 1a) *fuzzy r-partition* $(\mathbb{P},\mathbb{A}^{(r)})$ with $r = 1$

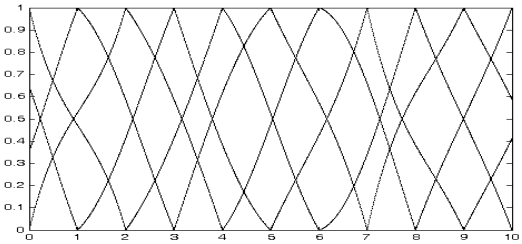


Figure 1b) *fuzzy r-partition* $(\mathbb{P},\mathbb{A}^{(r)})$ with $r = 2$

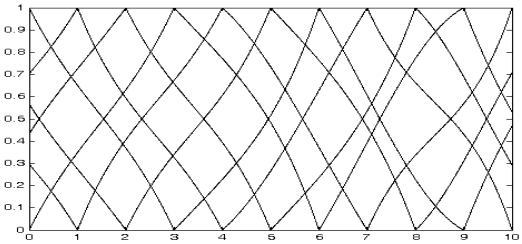


Figure 1c) *fuzzy r-partition* $(\mathbb{P},\mathbb{A}^{(r)})$ with $r = 3$

4 Fuzzy Transform with Extended Partitions

The direct F$^{(r)}$-transform (of degree r) based on the given *fuzzy r-partition* $(\mathbb{P}, \mathbb{A}^{(r)})$ can be defined by the vector $F^{(r)} = (F_1^{(r)}, F_2^{(r)}, ..., F_n^{(r)})^T$

$$F_k^{(r)} = \frac{\int_a^b f(x)A_k^{(r)}(x)dx}{\int_a^b A_k^{(r)}(x)dx}, k = 1, 2, ..., n \tag{8}$$

and the inverse F$^{(r)}$-transform by

$$\widehat{f}^{(r)}(x) = \frac{1}{r}\sum_{k=1}^{n} F_k^{(r)} A_k^{(r)}(x). \tag{9}$$

We see that each $\widehat{f}^{(r)}(x_k)$ has the structure of a moving average of the values $\{F_j^{(r)}, j = 1, ..., n\}$; in fact, at the points $x = x_k$ we have (assume $F_k^{(r)} = 0$ if $k < 1$ or $k > n$)

$$\widehat{f}^{(r)}(x_k) = \frac{1}{r}\sum_{j=k-r}^{k+r} F_j^{(r)} A_j^{(r)}(x_k) \tag{10}$$

i.e. a weighted average of $F_{k-r}^{(r)}, ..., F_k^{(r)}, ..., F_{k+r}^{(r)}$ with weights $\frac{A_{k-r}^{(r)}(x_k)}{r}, ..., \frac{1}{r}, ..., \frac{A_{k+r}^{(r)}(x_k)}{r}$, respectively.

It is possible to see that the F$^{(r)}$-transform has analogues properties to theorem 1 and theorem 2 of Perfilieva ([3]) i.e.

1. the local (for any $k = 1, 2, ..., n$) "error" function $\frac{1}{r}\int_a^b [f(x) - y]^2 A_k^{(r)}(x)dx$ is minimized by $y = F_k^{(r)}$;

2. the minimizer of $\frac{1}{r}\sum_{k=1}^{n}\int_a^b [f(x) - y_k]^2 A_k^{(r)}(x)dx$ is $y = F^{(r)} = (F_1^{(r)}, F_2^{(r)}, ..., F_n^{(r)})^T$ (notation $(\cdot)^T$ means transposition);

3. for fixed $r \geq 1$ and for any $\varepsilon > 0$ there exists a *fuzzy r-partition* $(\mathbb{P}_\varepsilon, \mathbb{A}_\varepsilon^{(r)})$ with the corresponding inverse F$^{(r)}$-transform $\widehat{f}_\varepsilon^{(r)}$ such that $\sup_x \left| f(x) - \widehat{f}_\varepsilon^{(r)}(x) \right| < \varepsilon$.

For a given family of parametric basic functions $A_k^{(r)}$ consider the linear combination (with $\theta = (\theta_1, \theta_2, ..., \theta_n)^T$)

$$C_{\mathbb{A}^{(r)}}(x; \theta, \beta) = \frac{1}{r}\sum_{j=1}^{n} \theta_j A_j^{(r)}(x; \beta) \text{ for } x \in [a, b] \tag{11}$$

where β is the vector of the non negative slope parameters $\beta_{j,0}$, $\beta_{j,1}$ giving $L_j(x; \beta_{j,0}, \beta_{j,1})$.

The following minimization defines $C_{\mathbb{A}^{(r)}}(x; \widehat{\theta}, \widehat{\beta})$ as the *least squares approximation* of $f(x)$ over interval $[a, b]$

$$\underset{\theta,\beta}{Minimize} \quad \int_a^b [f(x) - C_{\mathbb{A}^{(r)}}(x; \theta, \beta)]^2 dx \tag{12}$$

where $(\widehat{\theta}, \widehat{\beta})$ is the minimizing point.

For fixed β, the linear parameters $\widehat{\theta}$ are obtained by putting to zero the partial derivatives (for $i = 1, 2, ..., n$) and we obtain the equations

$$\int_a^b A_i^{(r)}(x; \beta) f(x)dx = \frac{1}{r}\sum_{j=1}^{n} \widehat{\theta}_j \int_a^b A_i^{(r)}(x; \beta) A_j^{(r)}(x; \beta)dx \tag{13}$$

From definition of the F$^{(r)}$-transform $F^{(r)}(\beta)$ as a function of β, we have

$$\int_a^b A_i^{(r)}(x; \beta) f(x)dx = F_i^{(r)}(\beta) \int_a^b A_i^{(r)}(x; \beta)dx \tag{14}$$

and we finally obtain the linear system of n equations

$$\begin{aligned} C^{(r)}(\beta)\widehat{\theta} &= rF^{(r)}(\beta) \quad i.e. \\ \sum_{j=1}^{n} c_{i,j}^{(r)}(\beta)\widehat{\theta}_j &= rF_i^{(r)}(\beta), \quad i = 1, 2, ..., n \end{aligned} \tag{15}$$

where the matrix $C^{(r)}(\beta)$ has elements

$$c_{i,j}^{(r)}(\beta) = \frac{1}{r}\int_a^b A_i^{(r)}(x; \beta) A_j^{(r)}(x; \beta)dx. \tag{16}$$

For fixed parameters β, we can obtain a matrix form for θ in the definition of function (11).

Given $f : [a, b] \longrightarrow \mathbb{R}$ and a fuzzy r-partition $(\mathbb{P}, \mathbb{A}^{(r)}(\beta))$ with the associated direct F$^{(r)}$-transform $F^{(r)}(\beta) = (F_1^{(r)}, F_2^{(r)}, ..., F_n^{(r)})^T$ and inverse F$^{(r)}$-transform

$$\widehat{f}_F^{(r)}(x; \beta) = \frac{1}{r}\sum_{j=1}^{n} F_j^{(r)}(\beta) A_j^{(r)}(x; \beta). \tag{17}$$

ISBN: 978-989-95079-6-8

Define the following vectors (we omit reference to β and r for simplicity of notation), obtained by evaluating $f(x)$ and $\widehat{f}_F(x)$ at points $x_1, x_2, ..., x_n$

$$f = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \\ f(x_n) \end{bmatrix}, \quad \widehat{f} = \begin{bmatrix} \widehat{f}_F(x_1) \\ \widehat{f}_F(x_2) \\ \\ \widehat{f}_F(x_n) \end{bmatrix}, \quad F = \begin{bmatrix} F_1^{(r)} \\ F_2^{(r)} \\ \\ F_n^{(r)} \end{bmatrix} \tag{18}$$

and the following $n \times n$ matrix

$$\begin{aligned} A^{(r)} &= [a_{i,j}] \text{ with} \\ a_{i,j} &= A_j^{(r)}(x_i). \end{aligned} \tag{19}$$

Matrix $A^{(r)}$ is a $(2r-1)$-band matrix and its $i-th$ row has the form ($i = 1, 2, ..., n$)

$$a_{i,j} = \begin{cases} 0 & j = 1, ..., i-r \\ A_j^{(r)}(x_i) & j = i-r+1, ..., i-1 \\ 1 & j = i \\ A_j^{(r)}(x_i) & j = i+1, ..., i+r-1 \\ 0 & j = i+r, ..., n \end{cases} \tag{20}$$

It is immediate that

$$\widehat{f} = \frac{1}{r}(A^{(r)})^T F \tag{21}$$

and $\frac{1}{r}(A^{(r)})^T$ represents a *moving average operator* acting on F to produce $\widehat{f}$.

4.1 Discrete $F^{(r)}$-Transform

Matrix A in (19) can be defined also for the discrete direct and inverse F$^{(r)}$-transforms. For m points $(t_i, y_i), i = 1, 2, ..., m$, where $y_i = f(t_i)$ and defining $A^{(r)}(\beta) = [a_{i,j}(\beta) = A_j^{(r)}(t_i, \beta)]$ for the family $\{A_j^{(r)}(t; \beta) | j = 1, 2, ..., n\}$ of basic functions, the discrete F$^{(r)}$-transform is

$$\begin{aligned} \text{For } j &= 1, 2, ..., n \\ F_j^{(r)}(\beta) &= \frac{g_j(\beta)}{s_j(\beta)} \text{ where} \\ g_j(\beta) &= \sum_{i=1}^{m} y_i A_j^{(r)}(t_i; \beta) \text{ and } s_j = \sum_{i=1}^{m} A_j^{(r)}(t_i; \beta) > 0 \end{aligned}$$

If the elements s_j are organized into a diagonal $n-$matrix

$$S = diag(s_1, s_2, ..., s_n)$$

so that

$$S^{-1} = diag(\frac{1}{s_1}, \frac{1}{s_2}, ..., \frac{1}{s_n})$$

and denote

$$\begin{aligned} y &= (y_1, ..., y_m)^T, \\ F^{(r)}(\beta) &= (F_1^{(r)}(\beta), ..., F_n^{(r)}(\beta))^T, \\ g(\beta) &= (g_1(\beta), ..., g_n(\beta))^T, \end{aligned}$$

then we have $(A^{(r)}(\beta))^T y = g(\beta)$ so that

$$F^{(r)}(\beta) = S(\beta)^{-1}(A^{(r)}(\beta))^T y.$$

Denote also, for $i = 1, 2, ..., m$ (note that for $i = r, ..., n-r+1$ we have $d_i = r$),

$$\begin{aligned} d_i(\beta) &= \sum_{j=1}^{n} A_j(t_i; \beta) > 0 \\ D &= diag(d_1, d_2, ..., d_m). \end{aligned}$$

For the discrete inverse F$^{(r)}$-transform

$$f_{F,\beta}^{(r)}(x) = \frac{\sum_{j=1}^{n} F_j^{(r)} A_j^{(r)}(x; \beta)}{\sum_{j=1}^{n} A_j^{(r)}(x; \beta)}$$

and denoting $f_{F,\beta}^{(r)} = (f_{F,\beta}^{(r)}(t_1), f_{F,\beta}^{(r)}(t_2), ..., f_{F,\beta}^{(r)}(t_m))^T$, we get

$$\begin{aligned} f_{F,\beta}^{(r)} &= D(\beta)^{-1} A^{(r)}(\beta) F^{(r)} \\ &= D(\beta)^{-1} A^{(r)}(\beta) S(\beta)^{-1} (A^{(r)}(\beta))^T y. \end{aligned} \tag{22}$$

5 Smoothing functions from $F^{(r)}$-Transform

In this section we investigate the smoothing effect of the inverse $F^{(r)}$-Transform $f^{(r)}$ on a finite set of data points (t_i, y_i), $i = 1, 2, ..., m$; the discrete direct and inverse F$^{(r)}$-transform are obtained with a partition $(\mathbb{P}, \mathbb{A}^{(r)}(\beta))$, where

$$\begin{aligned} \mathbb{P} &= \{a = x_1 < x_2 < ... < x_n = b\} \text{ and} \\ a &= \min\{t_i\}, \ b = \max\{t_i\} \\ x_i &= a + (i-1)/(n-1), i = 1, 2, ..., n. \end{aligned}$$

We first consider the parameters β as fixed for the family of basic functions $\mathbb{A}^{(r)}(\beta)$.

For a given function $f : [a, b] \longrightarrow \mathbb{R}$ a measure of smoothness can be given by its total variation

$$V(f) = \sup\{\sum_{j=1}^{k-1} |f(\alpha_{j+1}) - f(\alpha_j)|; \forall k \geq 2, \forall \alpha_j \in [a, b]\}$$

or by the following *total* variation (see [1], [2] and the references therein)

$$\widetilde{V}(f) = \sum_{j=1}^{m-1} |f(t_{j+1}) - f(t_j)|. \tag{23}$$

By increasing the integer value of r, the total variation $\widetilde{V}(f^{(r)})$ of the inverse transform functions $f^{(r)}$ tend to decrease to zero, as illustrated by the graphical representation in Fig. 2. we can obtain inverse fuzzy transform functions $f^{(r)}$ with arbitrarily small $\widetilde{V}(f^{(r)})$. A graphical representation is in Fig. 2.

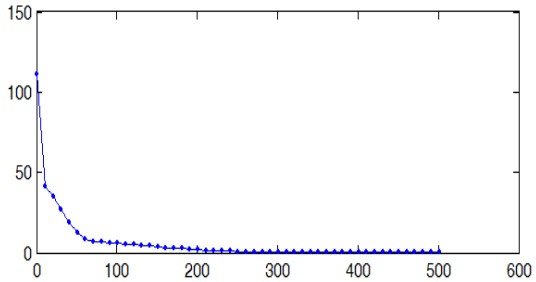


Figure 2. *Total variation $\widetilde{V}(f^{(r)})$ of inverse fuzzy transform function for increasing r.*

ISBN: 978-989-95079-6-8

It is interesting to see the combined effect of changing n and r; in figures 3a)-3c), $m = 101$ data points are generated from function $f(t) = 1+0.2t+sin(t).cos(t)$ by adding a Gaussian random noise with zero mean and standard deviation $\sigma = 2$, i.e. $y_i = f(t_i) + z_i$ and $z_i \in N(0, \sigma^2)$ (the t_i are uniform on $[-2\pi, 2\pi]$). In all the figures of this section, the continuos line is the original data, the dotted line is the inverse fuzzy transform $f^{(r)}$ and the circles are the direct $F^{(r)}$-transform.

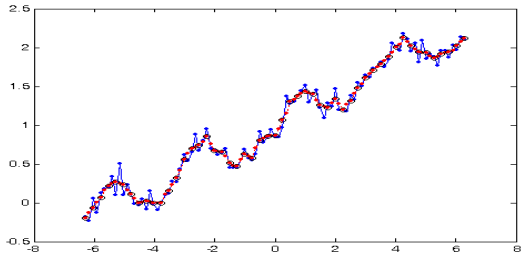


Figure 3a) Smoothing of m=101 data points, n=51, r=1.

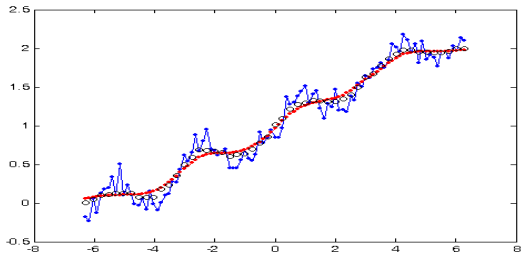


Figure 3b) Smoothing of m=101 data points, n=51, r=5.

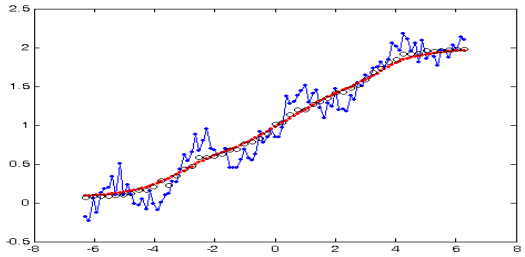


Figure 3c) Smoothing of m=101 data points, n=51, r=9.

In figures 4a)-4c) the data are generated from the same model but $n = 21$ is smaller then in previous case. When n is reduced, the smoothing effect is increased but it is more visible at a global scale: compare figures 3a) with 4a), 3b) with 4b) and 3c) with 4c).

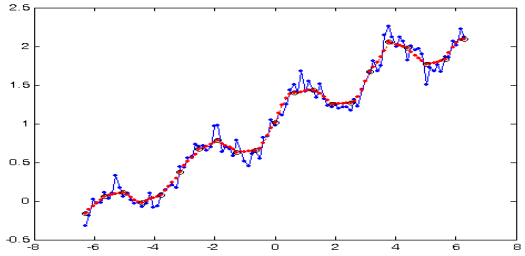


Figure 4a) Smoothing of m=101 data points, n=21, r=1.

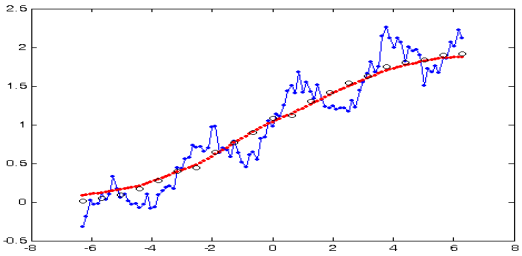


Figure 4b) Smoothing of m=101 data points, n=21, r=5.

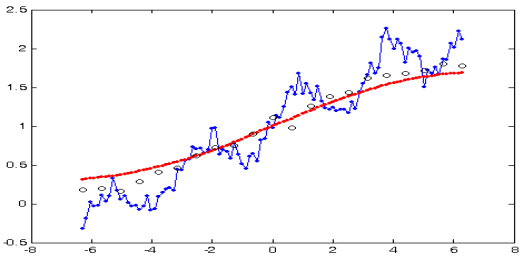


Figure 4c) Smoothing of m=101 data points, n=21, r=9.

If n is small with respect to m as in figures 5a) and 5b), the global smoothing is increased further.

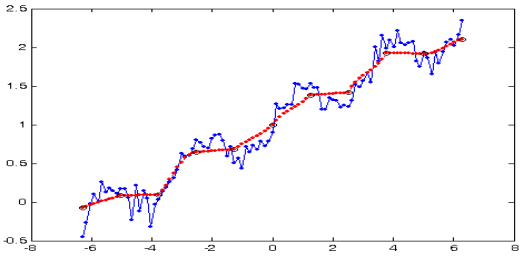


Figure 5a) Smoothing of m=101 data points, n=11, r=1.

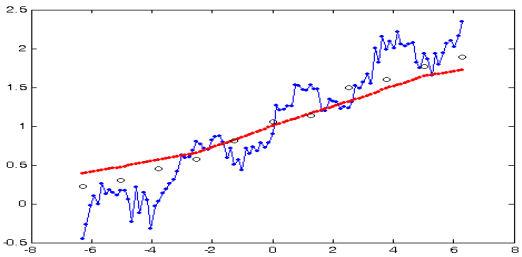


Figure 5b) Smoothing of m=101 data points, n=11, r=5.

So, according to the general properties of the $F^{(r)}$-transform, the discussion above illustrates that, in general, there are two possible sources for smoothness: for fixed m and r, reducing n will increase the smoothing effect; for fixed m and n, increasing r will increase the smoothness.

5.1 A Smoothing Procedure

Let (t_i, y_i), $i = 1, 2, ..., m$ be given observations and define $T_1 = \{i | x_1 \leq t_i \leq x_2\}$, $T_k = \{i | x_{k-1} < t_i \leq x_k\}$ for $k = 2, ..., n-1$.

The parameters β (if not fixed) can be estimated from the data by solving the following criterion, a penalized nonlinear least squares problem where the objective function is penalized by the total variation and a positive tuning parameter $\lambda > 0$ is introduced

$$\underset{\beta}{\text{Minimize}} \quad \left\| f^{(r)}_{F,\beta} - y \right\|^2 + \lambda \widetilde{V}(f^{(r)}_{F,\beta}). \qquad (24)$$

ISBN: 978-989-95079-6-8

Suppose that each T_k has a sufficient number of elements to allow the following approximation problem be well defined: determine $\beta = (\beta_{2,0}, ..., \beta_{n+r-1,0}\beta_{2,1}, ..., \beta_{n+r-1,1})$ to minimize the functional

$$\Phi^{(r)}(\beta) = \left\| L^{(r)}(\beta)F^{(r)}(\beta) - y \right\|^2 + \lambda \widetilde{V}(f^{(r)}_{F,\beta}) \quad (25)$$

where matrix $L^{(r)}(\beta)$ is

$$L^{(r)}(\beta) = D^{(r)}(\beta)^{-1}A^{(r)}(\beta) \text{ and}$$
$$F^{(r)}(\beta) = S^{(r)}(\beta)^{-1}(A^{(r)}(\beta))^T y$$

Alternatively, we can tune the total variation $\widetilde{V}(f^{(r)}_{F,\beta})$ by fixing a variation level $\overline{V} > 0$ and solving the following constrained problem

$$\Phi^{(r)}(\widehat{\beta}) = Minimize \left\| L^{(r)}(\beta)F^{(r)}(\beta) - y \right\|^2 \quad (26)$$
$$s.t. \quad \widetilde{V}(f^{(r)}_{F,\beta}) \leq \overline{V}$$

We show the procedure with the following example and with the value of $\overline{V} = \gamma \widetilde{V}(y)$ (a fraction γ of the total variation of the data); the generated data are $y_i = f(t_i) + 2z_i$ and $z_i \in N(0,1)$ (the $m = 101$ points t_i are uniform on $[0,2]$) with the function $f(t) = 5e^{-0.5t}sin^2(\pi t)$. For the generated data, $\widetilde{V}(y) = 210.25$. Figs 6a)-6c) illustrate the best smoothing obtained by solving (26) for different values of $\gamma \in \{0.5, 0.2, 0.1\}$ and $n = 21$; Figs 7a)-7b) show the best smoothing for $\gamma \in \{0.2, 0.1\}$ and $n = 51$.

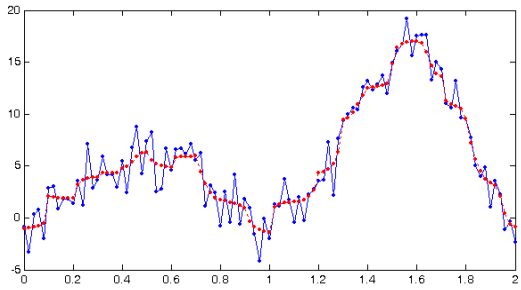


Figure 6a) Best smoothing fit with $n = 21, \gamma = 0.5$ ($r = 1$)

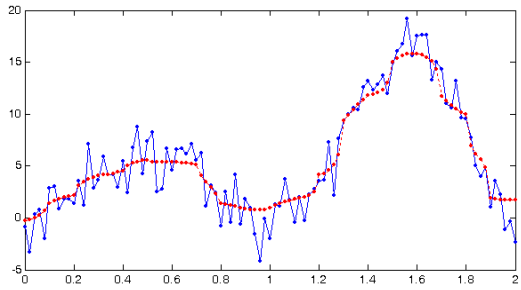


Figure 6b) Best smoothing fit with $n = 21, \gamma = 0.2$ ($r = 2$)

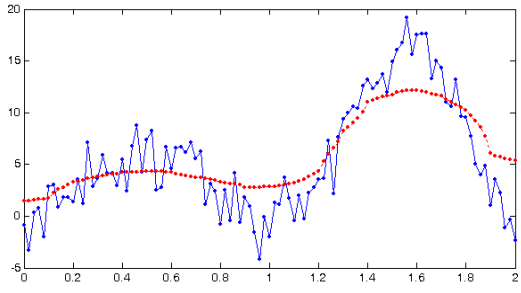


Figure 6c) Best smoothing fit with $n = 21, \gamma = 0.1$ ($r = 6$)

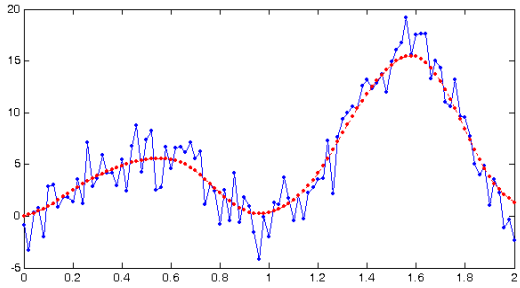


Figure 7a) Best smoothing fit with $n = 51, \gamma = 0.2$ ($r = 4$)

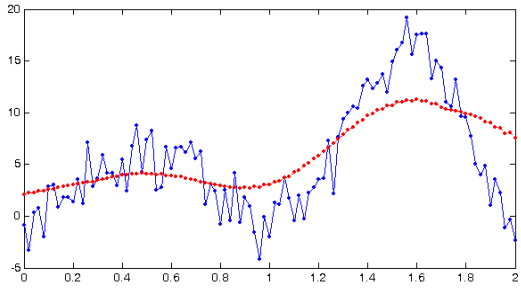


Figure 7b) Best smoothing fit with $n = 51, \gamma = 0.1$ ($r = 8$)

References

[1] P.L. Davies, A. Kovac, Local Extremes, runs, strings and multiresolution, *The Annals of Statistics*, 29 (2001) 1-65.

[2] A. Kovac, Smooth functions and local extreme values, *Computational Statistics & Data Analysis*, 51 (2007) 5155-5171.

[3] I. Perfilieva, Fuzzy Transforms: Theory and Applications, *Fuzzy Sets and Systems*, 157 (2006) 993-1023.

[4] I. Perfilieva, Fuzzy Transforms: A Challenge to conventional transforms. In P.W. Hawkes (Ed.), *Advances in Images and Electron Physics*, 147, Elsevier Academic Press, 2007, 137-196.

[5] I. Perfilieva, Fuzzy Transforms and their applications to data compression. In I. Bloch et al. (Eds), *Fuzzy Logic and Applications, LNAI 3849, Springer,* 2006, 19-31.

[6] L. Stefanini, Fuzzy Transform with Parametric LU-Fuzzy Partitions, Proceeding of the 8-th International FLINS Conference, (Madrid, Sept. 2008), in D. Ruan et Al. (Eds) "Computational Intelligence in Decision and Control", World Scientific, 2008, 399-404.

[7] L. Stefanini, L. Sorini, M.L. Guerra, Parametric representations of fuzzy numbers and application to fuzzy calculus, *Fuzzy Sets and Systems*, 157 (2006) 2423–2455.

[8] L. Stefanini, L. Sorini, M.L. Guerra, Fuzzy Numbers and Fuzzy Arithmetic. In W. Pedrycz, A. Skowron, V. Kreynovich (Eds), *Handbook of Granular Computing*, Chapter 12, J. Wiley & Sons, 2009.

[9] M. Stepnicka, R. Valasek, Fuzzy Transforms for functions of two variables. In J. Ramik, V. Novak (Eds), *Methods for decision support in environment with uncertainty*, University of Ostrava, IRAFM, 2003, 96-102.

ISBN: 978-989-95079-6-8

Towards F-transform of a Higher Degree

Irina Perfilieva, Martina Daňková

Institute for Research and Applications of Fuzzy Modeling, University of Ostrava
Ostrava, Czech Republic
Email: Irina.Perfilieva, Martina.Dankova@osu.cz

Abstract— The aim of this contribution is to show how the F-transform technique can be generalized from the case of constant components to the case of polynomial components. For this purpose, we choose complete functional spaces with inner products. After a general presentation of F^m-transform where $m \geq 0$, a detailed characterization of the F^1-transform is given. We applied a technique of numeric integration in order to simplify computation of F^1-transform components. The inverse F^m-transform, $m \geq 0$, is defined in the same way as the ordinary F-transform.

Keywords— F-transform, F^m-transform, fuzzy partition, orthogonal basis, Gaussian quadrature

1 Introduction

The goal of this paper is to provide a deeper analysis of fuzzy modeling and its contribution to general mathematics. In [1], we introduced the notion of a fuzzy transform (F-transform, for short) which explains modeling with fuzzy IF-THEN rules as a specific transformation. This enabled us to compare the success of fuzzy modeling with the success of classical transforms (Fourier, Laplace, integral, wavelet etc.). From this point of view, fuzzy transforms contribute to the theory of linear and, in particular, integral transforms. Moreover, they initiated a theory of semi-linear spaces (see [2]).

In [1], the approximation property of fuzzy transform has been described and then, in [3], it has been shown how shapes of basic functions influence the approximation quality. The F-transform has many other useful properties and great potential for various applications, such as special numerical methods, solution of ordinary and partial differential equations with fuzzy initial condition [4, 5], mining dependencies from numerical data [6], applications to signal processing, compression and decompression of images [7, 8], and fusion of images [9].

We aim at showing how the F-transform technique can be generalized in the sense that constant components considered as polynomials of 0 degree are replaced by polynomials of degree $m \geq 1$. For this purpose, we choose complete functional spaces with inner products. After presentation of the generalized approach, a detailed characterization of the F^1-transform is given. We applied a technique of numeric integration in order to simplify computation of F^1-transform components. The inverse F^m-transform, $m \geq 1$, is defined in the same way as the inverse F-transform.

The paper is organized as follows: in Section 2, we introduce the concept of F^m-transform of finite degree $m \geq 0$, and show some of its properties and its relation to the ordinary F-transform. In Section 3, F^1-transform is introduced in details. Moreover, a simplification of F^1-transform components computation is discussed. The inverse F^m-transform is discussed in Section 4.

2 F-transform of an Arbitrary Finite Degree

Let us recall that the (direct) F-transform of an integrable function is a certain vector with real components. In [1], we proposed various formulas which represent components of the F-transform and showed a relationship between a given function and its F-transform. In this contribution, we propose to generalize our view on the F-transform and consider it as a vector of components that are orthogonal projections of an original function onto a linear subspace of certain functions that have polynomial representation. In the particular, if the degree of polynomials is zero, we obtain the originally proposed F-transform which will be referred to as F^0-transform in the sequel.

2.1 Fuzzy partition

Let $[a, b]$ be an interval on the real line $\mathbb{R}$. Fuzzy sets on $[a, b]$ will be identified with their membership functions, i.e. mappings from $[a, b]$ into $[0, 1]$. The notion of fuzzy partition is a principle one for our construction so that we will repeat it from [1] and slightly change below.

Definition 1
Let $[a, b]$ be an interval on $\mathbb{R}$, $n \geq 2$ and $x_1, \ldots, x_n$ nodes such that $a = x_1 < \ldots < x_n = b$. We say that fuzzy sets $A_1, \ldots, A_n$, identified with their membership functions, constitute a fuzzy partition *of $[a, b]$ if for $k = 1, \ldots, n$, they fulfill the following conditions :*

1. *$A_k : [a, b] \longrightarrow [0, 1]$, $A_k(x_k) = 1$;*
2. *for $k = 2, \ldots, n$, $A_k(x) = 0$ if $x \in [a, x_{k-1}]$ and for $k = 1, \ldots, n-1$, $A_k(x) = 0$ if $x \in [x_{k+1}, b]$,*
3. *$A_k(x)$ is continuous;*
4. *for $k = 2, \ldots, n$, $A_k(x)$ strictly increases on $[x_{k-1}, x_k]$ and for $k = 1, \ldots, n-1$, $A_k(x)$ strictly decreases on $[x_k, x_{k+1}]$;*
5. *for all $x \in [a, b]$*

$$\sum_{k=1}^{n} A_k(x) = 1. \tag{1}$$

The membership functions $A_1, \ldots, A_n$ are called basic functions.

ISBN: 978-989-95079-6-8

Let us extend $[a,b]$ by two extra nodes $x_0 = 2a - x_2$ and $x_{n+1} = 2b - x_{n-1}$ and for $x \in [x_0, a)$ define $A_1(x) = A_1(2a - x)$, and for $(b, x_{n+1}]$ define $A_n(x) = A_n(2b - x)$. Further on we will always assume that $[a,b]$ and membership functions A_1, A_n are extended.

We say that the fuzzy partition $A_1, \ldots, A_n$, $n \geq 3$, is *h-uniform* if nodes $x_0, \ldots, x_{n+1}$ are h-equidistant, i.e. for all $k = 1, \ldots, n$, $x_k = a + h(k-1)$ where $h = (b-a)/(n-1)$, and two additional properties are met:

6. $A_k(x_k - x) = A_k(x_k + x)$, for all $x \in [0, h]$, $k = 1, \ldots, n$,
7. $A_k(x) = A_{k-1}(x - h)$, for all $k = 2, \ldots, n$ and $x \in [x_{k-1}, x_{k+1}]$.

2.2 F-transform in a space of functions with scalar product

Let us fix $[a,b]$, $n \geq 3$, nodes $x_0 < \ldots < x_{n+1}$ and fuzzy partition $A_1, \ldots, A_n$ of $[a,b]$. For every $k = 1, \ldots, n$, let us denote $L_2(A_k)$ a set of functions $f : [x_{k-1}, x_{k+1}] \longrightarrow \mathbb{R}$ for which the following integral

$$\int_{x_{k-1}}^{x_{k+1}} f(x)^2 A_k(x) dx$$

exists. Let

$$(f,g)_k = \int_{x_{k-1}}^{x_{k+1}} f(x) g(x) A_k(x) dx, \tag{2}$$

be a weighted scalar product of functions f and g. Then $L_2(A_k)$ is a *linear space of functions with scalar product*.

For every integer $m \geq 0$, let $\varphi_0^k, \varphi_1^k, \ldots, \varphi_m^k \in L_2(A_k)$, be an orthogonal system of polynomials where $\varphi_0^k = 1$ and orthogonality is considered with respect to (2). Denote $L_2^m(A_k)$ a linear subspace of $L_2(A_k)$ with the basis $\varphi_0^k, \varphi_1^k, \ldots, \varphi_m^k$.

The following trick allows us to extend arbitrary function $f : [a,b] \longrightarrow \mathbb{R}$ to function $f^{ex} : [x_0, x_{n+1}] \longrightarrow \mathbb{R}$:

$$f^{ex}(x) = \begin{cases} f(a-x) = 2f(a) - f(a+x), \text{ if } x \in [0, x_2 - a], \\ f(x), \text{ if } x \in [a,b], \\ f(b+x) = 2f(b) - f(b-x), \text{ if } x \in [0, b - x_{n-1}] \end{cases}$$

Definition 2
Let $f : [a,b] \longrightarrow \mathbb{R}$ be a given function and f^{ex} its extension to $[x_0, x_{n+1}]$. Assume that for all $k = 1, \ldots, n$, $f^{ex}|_{[x_{k-1}, x_{k+1}]} \in L_2(A_k)$. Let F_k^m be the k-th orthogonal projection of f^{ex} on $L_2^m(A_k)$, $k = 1, \ldots, n$. Then the n-tuple $[F_1^m, \ldots, F_n^m]$ of functions is the F^m-transform of f with respect to $A_1, \ldots, A_n$. Every function F_k^m, $k = 1, \ldots, n$, is called the F^m-transform component.

Remark 1
Definition 2 does not provide us with a formal representation of components $F_1^m, \ldots, F_n^m$. Let us show how they can be obtained. According to the definition above, every component F_k^m minimizes the scalar product $(f^{ex} - g, f^{ex} - g)_k$ where g is an arbitrary function from $L_2^m(A_k)$. Taking into account that $\varphi_0^k, \varphi_1^k, \ldots, \varphi_m^k$ is the basis of $L_2^m(A_k)$, we can represent g by a linear combination $x_0\varphi_k^0 + x_1\varphi_k^1 + \ldots + x_m\varphi_k^m$ of basis functions. Assume that $c_0, c_1 \ldots, c_m$ are coefficients that minimize the following integral

$$\int_{x_{k-1}}^{x_{k+1}} (f^{ex}(x) - (x_0\varphi_k^0(x) + x_1\varphi_k^1(x) + \ldots + x_m\varphi_k^m(x)))^2 A_k(x) dx. \tag{3}$$

Hence, $F_k^m = c_0\varphi_k^0 + c_1\varphi_k^1 + \ldots + c_m\varphi_k^m$.

Below, we will prove the following fact: the original F-transform (see, e.g. [1]) is actually (up to the first and the last components) the F^0-transform. This requires to show that every F-transform component F_k, $k = 2, \ldots, n-1$, that has been originally introduced by

$$F_k = \frac{\int_{x_{k-1}}^{x_{k+1}} f(x) A_k(x) dx}{\int_{x_{k-1}}^{x_{k+1}} A_k(x) dx}, \tag{4}$$

is the k-th orthogonal projection of f on $L_2^0(A_k)$. In order to prove this fact, we recall that the basis of $L_2^0(A_k)$ consists of the constant function φ_k^0 that has the value 1. Then the assertion above immediately follows from the proposition given below which has been proved in [1]:

Theorem 1
Let f be a continuous function on $[a,b]$ and $A_1, \ldots, A_n$ be basic functions which constitute a fuzzy partition of $[a,b]$. Then the k-th component F_k (4) of the F-transform gives minimum to the function

$$\Phi(y) = \int_a^b (f(x) - y)^2 A_k(x) dx$$

defined on $[f(a), f(b)]$.

3 F^1-transform

On the example of F^1-transform, we will show how an arbitrary component of F^m, $m \geq 1$, can be computed. Actually, one possibility of computation directly follows from minimization of integral (3) (see Remark 1). We will use another approach which corresponds to the definition of F^1-transform.

Throughout this section, we fix an h-uniform partition $A_1, \ldots, A_n$ of $[a,b]$ where $n \geq 3$, and assume that $f : [a,b] \longrightarrow \mathbb{R}$ is a function such that for all $k = 1, \ldots, n$, $f^{ex}|_{[x_{k-1}, x_{k+1}]} \in L_2(A_k)$. The explanation below will be given for an arbitrary k, $k = 1, \ldots, n$.

3.1 F^1-transform components

Since component F_k^1 of the F^1-transform is the k-th orthogonal projection of f^{ex} on $L_2^m(A_k)$, and $L_2^m(A_k)$ is a linear subspace of $L_2(A_k)$ with orthogonal basis φ_k^0, φ_k^1, we will begin this subsection with a characterization of orthogonal polynomials φ_k^0, φ_k^1 of degrees 0 and 1.

Lemma 1
Polynomials $\varphi_k^0, \varphi_k^1 : [x_{k-1}, x_{k+1}] \longrightarrow \mathbb{R}$ such that for all $x \in [x_{k-1}, x_{k+1}]$, $\varphi_k^0(x) = 1$, $\varphi_k^1(x) = x - x_k$ are orthogonal with weight A_k, $k = 1, \ldots, n$.

PROOF: The proof is technical and follows from the following assertions:

(i) $\int_{x_{k-1}}^{x_{k+1}} A_k(x) dx = h$,

ISBN: 978-989-95079-6-8

(ii) $\int_{x_{k-1}}^{x_{k+1}} xA_k(x)dx = hx_k$,

and properties of the uniform partition. □

Remark 2
Orthogonal polynomials of degrees 0 *and* 1, *described in Lemma 1, are invariant to shapes of* A_k.

Theorem 2
Under the assumptions above, the F^1*-transform of* f *with respect to* $A_1, \ldots, A_n$ *is the vector* $[F_1^1, \ldots, F_n^1]$ *of linear functions such that an arbitrary component* F_k^1, $k = 1, \ldots, n$, *is represented as follows:*

$$F_k^1(x) = c_k^0 + c_k^1(x - x_k), \quad x \in [x_{k-1}, x_{k+1}]$$

where

$$c_k^0 = \frac{\int_{x_{k-1}}^{x_{k+1}} f(x)A_k(x)dx}{h}, \tag{5}$$

$$c_k^1 = \frac{\int_{x_{k-1}}^{x_{k+1}} f(x)(x - x_k)A_k(x)dx}{\int_{x_{k-1}}^{x_{k+1}} (x - x_k)^2 A_k(x)dx}. \tag{6}$$

PROOF: The proof will be given for one fixed component F_k^1 where $k = 1, \ldots, n$. By Definition 2, F_k^1 is the k-th orthogonal projection of f^{ex} on $L_2^1(A_k)$ where the orthogonality is determined by (2). Therefore,

$$f^{ex}|_{[x_{k-1}, x_{k+1}]} = c_k^0 + c_k^1(x - x_k) + R_k(x)$$

where R_k is orthogonal to each basis functions φ_k^0 and φ_k^1. Therefore,

$$c_k^0 = \frac{(f^{ex}, \varphi_k^0)_k}{(\varphi_k^0, \varphi_k^0)_k},$$

$$c_k^1 = \frac{(f^{ex}, \varphi_k^1)_k}{(\varphi_k^1, \varphi_k^1)_k}$$

which after substitution gives the required expressions for c_k^0 and c_k^1. □

Corollary 1
Under the assumptions above, $c_k^0 = F_k$ *where* F_k *is the* k*-th component (4) of the ordinary* F*-transform.*

PROOF: The proof follows from expressions (4), (6) and equality $\int_{x_{k-1}}^{x_{k+1}} A_k(x)dx = h$. □

3.2 Simplification of F^1*-transform components computation*

In this subsection, we will show how Gaussian quadratures and the properties of orthogonal polynomials can be used for replacing integral $\int_{x_{k-1}}^{x_{k+1}} (x - x_k)^2 A_k(x)dx$ in the denominator of c_k^1 by its precise value.

At first, let us recall the notion of Gaussian quadratures (see e.g.[10]). The approximate equality

$$\int_{x_{k-1}}^{x_{k+1}} f(x)A_k(x)dx \approx \frac{h}{2} \sum_{i=1}^{N} d_i f(t_i) \tag{7}$$

which is precise for all polynomials of the highest possible degree is a Gaussian quadrature. We will put $N = 2$ and characterize arguments t_1, t_2 and the highest possible degree of polynomials which turn (7) into a precise equality. The following statement (see e.g.[10]) gives the required characterization: if

(i) t_1, t_2 are roots of polynomial φ_k^2 which is orthogonal to φ_k^0 and φ_k^1, and

(ii) equality

$$\int_{x_{k-1}}^{x_{k+1}} P_l(x)A_k(x)dx = h(d_1 P_l(t_1) + d_2 P_l(t_2)), \tag{8}$$

holds true for some coefficients d_1, d_2 and all polynomials P_l of degrees $l = 0, 1$,

then (8) holds true for all polynomials P_l of degrees $0 \leq l \leq 3$.

Thus, our next purpose is to find a polynomial of the degree 2, orthogonal to φ_k^0 and φ_k^1, as well as to find its roots. Due to positivity and symmetry of A_k, two lemmas below hold true.

Lemma 2
If t_1, t_2 *are symmetrical with respect to* x_k *then equalities*

$$\int_{x_{k-1}}^{x_{k+1}} A_k(x)dx = h(\frac{1}{2} + \frac{1}{2}) = h,$$

$$\int_{x_{k-1}}^{x_{k+1}} (x - x_k)A_k(x)dx = \frac{h}{2}((t_1 - x_k) + (t_2 - x_k)) = 0,$$

hold true for $l = 0, 1$.

Lemma 3
If we denote

$$I_2 = \int_{x_{k-1}}^{x_{k+1}} (x - x_k)^2 A_k(x)dx$$

then polynomial φ_k^2, *represented by*

$$\varphi_k^2(x) = (x - x_k)^2 - \frac{I_2}{h} \tag{9}$$

is orthogonal to φ_k^0 *and* φ_k^1 *on* $[x_{k-1}, x_{k+1}]$.

The roots t_1, t_2 *of* φ_k^2 *belong to* (x_{k-1}, x_{k+1}) *and are symmetrical with respect to* x_k, *i.e. for some* δ, $t_1 = x_k - \delta$ *and* $t_2 = x_k + \delta$.

The statement above together with Lemmas 2, 3 leads to the equality

$$\int_{x_{k-1}}^{x_{k+1}} (x - x_k)^2 A_k(x)dx = h\delta^2.$$

Thus, the desired simplified representation of c_k^1 is as follows:

$$c_k^1 = \frac{\int_{x_{k-1}}^{x_{k+1}} f(x)(x - x_k)A_k(x)dx}{h\delta^2}.$$

ISBN: 978-989-95079-6-8

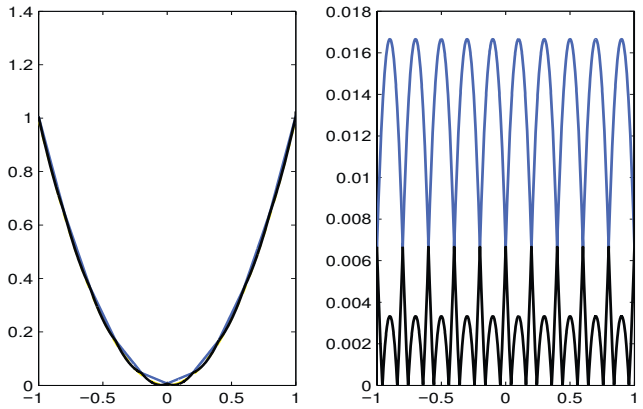


Figure 1: *Left.* The function x^2 and its inverse F^0 (gray line) and F^1 (black line) transforms. *Right.* Graphs of the error functions. Maximal errors of approximation are: 0.017 (the inverse F^0-transform) and 0.062 (the inverse F^1-transform).

4 Inverse F^m-transform

Similarly to the ordinary F-transform, the inverse F^m-transform is defined as a linear combination of the basic functions with "coefficients" given by the F^m-transform components.

Definition 3
Let $f : [a, b] \longrightarrow \mathbb{R}$ be a given function and f^{ex} its extension to $[x_0, x_{n+1}]$ such that for all $k = 1, \ldots, n$, $f^{ex}|_{[x_{k-1}, x_{k+1}]} \in L_2(A_k)$. For a certain $m \geq 0$, let $[F_1^m, \ldots, F_n^m]$ be the F^m-transform of f with respect to $A_1, \ldots, A_n$. Then the function

$$f_{F,m,n}(x) = \sum_{k=1}^{n} F_k^m A_k(x) \tag{10}$$

is called the inverse F^m-transform.

The following recurrent formula easily follows from Definition 3 and the whole structure of the F^m-transform, $m \geq 1$:

$$f_{F,m,n}(x) = f_{F,m-1,n}(x) + \sum_{k=1}^{n} c_m \varphi_k^m(x) A_k(x). \tag{11}$$

By Remark 1, the components F_k^m, $k = 1, \ldots, n$, $m \geq 0$, are the best approximation of f among all elements represented by linear combinations $a_0\varphi_k^0 + a_1\varphi_k^1 + \ldots + a_m\varphi_k^m$. Therefore, for each $k = 1, \ldots, n$, F_k^m is a better approximation of f than F_k^{m-1}. This observation together with (11) implicitly demonstrates that the quality of approximation of f by $f_{F,m,n}$ is better than that of approximation by $f_{F,m-1,n}$. The same conclusion follows also from our experiments (see Figure 1). The proof of this assertion is a matter of future investigation.

5 Conclusion

We have generalized the F-transform technique to the case where its components are polynomials. A detailed characterization of the F^1-transform with linear components was given. We have shown how a computation of F^1-transform components can be simplified if the technique of Gaussian quadratures is used. The inverse F^m-transform, $m \geq 1$, is defined in the same way as the inverse F-transform.

Acknowledgement

The paper has been supported partially by the grant IAA108270902 of GA AV and partially by the project MSM 6198898701 of the MŠMT ČR.

References

[1] I. Perfilieva. Fuzzy transforms: Theory and applications. *Fuzzy Sets and Systems*, 157:993–1023, 2006.

[2] A. Di Nola, A. Lettieri, I. Perfilieva, and V. Novák. Algebraic analysis of fuzzy systems. *Fuzzy Sets and Systems*, 158:1–22, 2007.

[3] L. Stefanini. Fuzzy transform with parametric lu-fuzzy partitions. In D. Ruan and et al., editors, *Computational Intelligence in Decision and Control*, pages 399–404. World Scientific, Paris, France, 2008.

[4] M. Štěpnička and R. Valášek. Fuzzy transforms for functions with two variables. In J. Ramík and V. Novák, editors, *Methods for Decision Support in Environment with Uncertainty*, pages 96–102. IRAFM, University of Ostrava, 2003.

[5] I. Perfilieva, H. De Meyer, B. De Baets, and D. Plskova. Cauchy problem with fuzzy initial condition and its approximate solution with the help of fuzzy transform. In *Proc. of WCCI 2008, IEEE Int. Conf. on Fuzzy Systems*, pages 2285–2290, Hong Kong, 2008.

[6] V. Novák, I. Perfilieva, A. Dvořák, G. Chen, Q. Wei, and P. Yan. Fuzzy transform in the analysis of data. *Int. Journ. of Appr. Reasoning*, 48:4–22, 2008.

[7] S. Sessa, F. Di Martino, V. Loia, and I. Perfilieva. An image coding/decoding method based on direct and inverse fuzzy transforms. *Int. Journ. of Appr. Reasoning*, 48:110—-131, 2008.

[8] I. Perfilieva. Fuzzy transforms: A challenge to conventional transforms. In P. W. Hawkes, editor, *Advances in Images and Electron Physics, 147*, pages 137–196. Elsevier Academic Press, San Diego, 2007.

[9] I. Perfilieva and M. Daňková. Image fusion on the basis of fuzzy transforms. In D. Ruan and et al., editors, *Computational Intelligence in Decision and Control*, pages 471–476. World Scientific, Paris, France, 2008.

[10] N. S. Bachvalov, N. P. Zhidkov, and G. M. Kobelkov. *Numerical Methods*. Nauka, Moscow, 1987.

ISBN: 978-989-95079-6-8

Some chaotic and mixing properties of Zadeh's extension

Jiří Kupka

Institute for Research and Applications of Fuzzy Modeling, University of Ostrava
30. dubna 22, 701 33 Ostrava, Czech Republic
Email: Jiri.Kupka@osu.cz

Abstract— *Let X be a compact metric space, let φ be a continuous self-map on X, and let $\mathbb{F}(X)$ denote the space of fuzzy sets on X equipped with the levelwise topology. In this paper we study relations between various dynamical properties of a given (crisp) dynamical system (X, φ) and its Zadeh's extension Φ on $\mathbb{F}(X)$. Among other things we study various (weak, strong, mild etc.) mixing properties and also several kinds of chaotic behaviors (Li-Yorke chaos, ω-chaos, distributional chaos, topological chaos etc.).*

Keywords— Zadeh's extension, fuzzification, chaos, mixing, transitivity, topological entropy.

1 Introduction

Throughout this paper, let (X, d_X) be a compact metric space and let $C(X)$ denote the space of continuous maps $\varphi : X \to X$. A discrete dynamical system is a pair (X, φ). For other notions and notations mentioned in this section, we refer to Section **??**. It is well known ([**?**]) that the discrete dynamical system (X, φ) naturally induces a dynamical system $(\mathbb{F}(X), \Phi)$ on the space $\mathbb{F}(X)$ of all fuzzy compact subsets of X. The map Φ is called the *fuzzification* (or *Zadeh's extension*) (see (**??**)).

It is natural to ask the following **question:** *how is the dynamical complexity of the fuzzified (resp. crisp) dynamical system related to the dynamical properties of the original (resp. fuzzy) one*. There are only a few papers devoted to this question so far – for example, [**?**], [**?**] and [**?**], where different chaotic properties of fuzzy discrete dynamical systems were considered.

In this paper, we consider the space $\mathbb{F}(X)$ of upper semi-continuous fuzzy sets with compact supports. This space is equipped with the topology induced by the levelwise metric d_∞ (see (**??**)), since this topology is stronger than the other topologies commonly used in fuzzy topological dynamics (e.g. see [**?**]). We especially deal with the subspace $\mathbb{F}^1(X) \subseteq \mathbb{F}(X)$ of all normal fuzzy sets on X (see (**??**)). The reason for this is the following: no fuzzification $\Phi : \mathbb{F}(X) \to \mathbb{F}(X)$ admits one of the simplest chaotic behaviors (namely the transitivity, see Proposition **??**) and, consequently, it does not admit more complex behavior.

This paper is a partial answer to the question mentioned above. Our results concerning the most commonly used chaotic and mixing properties can be summarized as follows:

• if P denotes either the distributional or Li-Yorke or topological or ω-chaos then

$$(\varphi \text{ has } P \Rightarrow \Phi \text{ has } P), \text{ but } (\Phi \text{ has } P \not\Rightarrow \varphi \text{ has } P), \quad (1)$$

• if P denotes either transitivity or total transitivity, then

$$(\Phi \text{ has } P \Rightarrow \varphi \text{ has } P), \text{ but } (\varphi \text{ has } P \not\Rightarrow \Phi \text{ has } P), \quad (2)$$

• if P denotes one of the following properties: exactness, sensitive dependence, weak mixing, mild mixing, or strong mixing, then

$$(\Phi \text{ has } P \Rightarrow \varphi \text{ has } P), \quad (3)$$

but the validity of the converse implication is unknown.

This paper is organized as follows: in Section **??**, we introduce notation and definitions used in this paper. Then, in Section **??**, some preliminary results are proven, showing also connections between the set-valued and fuzzified system induced by the same original system. Finally, the chaotic and mixing properties are studied in Section **??**.

2 Definitions and notation

Further we denote by $\mathbb{N}$ and $\mathbb{R}$ the set of integers and real numbers, respectively. Now we define some classic notions from topological dynamics. For a given dynamical system (X, φ) and a given point $x \in X$, we define the n-th *iteration* of the point x inductively by $\varphi^0(x) = x$, $\varphi^{n+1}(x) = \varphi(\varphi^n(x))$ for any $n \in \mathbb{N}$. Then, the sequence $\{\varphi^n(x)\}_{n\in\mathbb{N}}$ of all iterations of x is called the *trajectory* of the point x. Any limit point of the trajectory of the point x is called an *ω-limit point* of the point x, and the union $\omega_\varphi(x)$ of all ω-limit points of the point x is the *ω-limit set* of the point x. The iterations of a given set $A \subseteq X$ are defined analogously. The point $x \in X$ is called *fixed* if $\varphi(x) = x$ or *periodic* if $\varphi^k(x) = x$ for some $k \in \mathbb{N}$. We denote by $\omega(\varphi)$, $P(\varphi)$ and $Fix(\varphi)$ the set of ω-limit, periodic and fixed points, respectively.

A map $\varphi \in C(X)$ is called *transitive* if for any non-empty open subsets $U, V \subseteq X$, there exists some $k \in \mathbb{N}$ such that $\varphi^k(U) \cap V \neq \emptyset$. The map φ is *totally transitive* if the n-th iteration of φ is transitive for any $n \in \mathbb{N}$. The map φ is *weakly mixing* if the product map $\varphi \times \varphi$ is transitive. The map φ is *strongly mixing* if for any non-empty open subsets $U, V \subseteq X$ there exists some $m \in \mathbb{N}$ such that $\varphi^k(U) \cap V \neq \emptyset$ for any $k \geq m$. The map φ is *topologically exact* (or simply *exact*) if for any non-empty subset $U \subseteq X$, there exists some $k \in \mathbb{N}$, such that $\varphi^k(U) = X$.

2.1 Chaotic properties

When defining chaotic properties we follow the notation introduced in [**?**]. The notion of distributional chaos was introduced in [**?**]. For any $x, y \in X$, $t \in \mathbb{R}$ and $n \in \mathbb{N}$, set

$$\xi(x, y, t, n) = \#\{i,\ 0 \leq i < n \wedge d(\varphi^i(x), \varphi^i(y)) < t\}. \quad (4)$$

Set

$$F^\star_{xy}(t) = \limsup_{n\to\infty} \frac{1}{n}\xi(x, y, t, n) \quad (5)$$

ISBN: 978-989-95079-6-8

and

$$F_{xy}(t) = \liminf_{n\to\infty} \frac{1}{n}\xi(x,y,t,n). \quad (6)$$

Obviously, both maps $F^{\star}_{xy}$ and F_{xy} are nondecreasing, $0 \leq F_{xy}(t) \leq F^{\star}_{xy}(t) \leq 1$ for all $t \in \mathbb{R}$, $F^{\star}_{xy}(t) = 0$ if $t \leq 0$ and $F^{\star}_{xy}(t) = 1$ if $t \geq diam(X)$. The map $F^{\star}_{xy}$ ($F_{xy}(t)$) is an upper (a lower) distribution function for $x, y \in X$.

The map φ is *distributionally chaotic* of type 1 (d_1C) if $F^{\star}_{xy} \equiv 1$ and $F_{xy}(t) = 0$ for some $t > 0$. The map φ is *distributionally chaotic* of type 2 (d_2C) if $F^{\star}_{xy} \equiv 1$ and $F^{\star}_{xy}(t) > F_{xy}(t)$ for some $t > 0$. Finally, the map φ is *distributionally chaotic* of type 3 (d_3C) if $F^{\star}_{xy}(t) > F_{xy}(t)$ for all $t \in J$, where J is a nondegenerate interval.

It follows from the definition that

$$d_1C \Rightarrow d_2C \Rightarrow d_3C. \quad (7)$$

However, the converse implications are not valid (see, for instance, [**?**] and [**?**]).

Two points $x, y \in X$ form a *Li-Yorke pair* if

$$\limsup_{n\to\infty} d_X(\varphi^n(x), \varphi^n(y)) > 0 \quad (8)$$

and

$$\liminf_{n\to\infty} d_X(\varphi^n(x), \varphi^n(y)) = 0. \quad (9)$$

A set $S \subseteq X$ is a *LY-scrambled set* for the map φ if $\#S \geq 2$ and every pair from S is Li-Yorke. The map φ is *Li-Yorke chaotic* (shortly *LYC*) if there exists an uncountable LY-scrambled set.

A map $\varphi \in C(X)$ is *ω-chaotic* ([**?**]) (shortly ωC) if there exists an uncountable *ω-scrambled* set $S \subseteq X$, i.e. for any two points $x, y \in S$, the following conditions are satisfied: (i) $\omega_\varphi(x) \setminus \omega_\varphi(y)$ is uncountable, (ii) $\omega_\varphi(x) \setminus \omega_\varphi(y) \neq \emptyset$ and (iii) $\omega_\varphi(x) \cap P(\varphi) \neq \emptyset$.

If a map $\varphi : X \to X$ is transitive and $P(\varphi)$ is dense in X then φ is called *Devaney chaotic*. It should be mentioned that in the original definition of Devaney, φ *depends sensitively on initial conditions*,i.e. there exists $\delta > 0$ such that for any $x \in X$ and any open neighborhood U of x there is $y \in U$ satisfying $d_X(\varphi^k(x), \varphi^k(y)) > \delta$ for some $k \in \mathbb{N}$. But it was proved that this condition is implied by the transitivity and density of periodic points (see [**?**] and [**?**]).

The notion of positive topological entropy was firstly defined by Bowen ([**?**]). The topological entropy of a map φ is a number $h(\varphi) \in [0, \infty]$, defined by

$$h(\varphi) = \lim_{\varepsilon>0} \limsup_{n\to\infty} \#E(n, \varphi, \varepsilon), \quad (10)$$

where $E(n, \varphi, \varepsilon)$ is a $(n, \varphi, \varepsilon)$-*span* with a minimal possible number of points, i.e. a set such that for any $x \in X$ there exists a $y \in E(n, \varphi, \varepsilon)$ satisfying $d(\varphi^k(x), \varphi^k(y)) < \varepsilon$ for any j, $1 \leq j \leq n$. A map φ is *topologically chaotic* (shortly PTE) if $h(\varphi) > 0$. It is well-known that the topological entropy is monotone in the following way: for any $A, B \subseteq X$,

$$A \subseteq B \quad \Rightarrow \quad h(\varphi|_A) \leq h(\varphi|_B). \quad (11)$$

A map $\varphi \in C(X)$ has the *specification property* if for any $\varepsilon > 0$ there is a positive $M \in \mathbb{N}$ such that for any integer $k \geq 2$ and any k points $x_i \in X$, $i = 1, 2, \ldots, k$ and any $2k$ integers $a_1 \leq b_1 < a_2 \leq b_2 < \ldots < a_k \leq b_k$ with $a_i - b_{i-1} \geq M$, there exists $z \in X$ for which

$$d(\varphi^n(z), \varphi^n(x_i)) < \varepsilon \quad (12)$$

for any $n = a_i, \ldots, b_i$ and any $i = 1, 2, \ldots, k$.

The following implications are currently known among the chaotic and mixing properties mentioned above:

$$\begin{gathered}\text{specification property} \Rightarrow \text{strong mixing} \Rightarrow \\ \text{mild mixing} \Rightarrow \text{weak mixing} \Rightarrow \\ \text{total transitivity} \Rightarrow \text{transitivity}.\end{gathered} \quad (13)$$

For further details and relations among the chaotic properties, we refer to [**?**] and to the references therein.

2.2 Metric spaces of fuzzy sets

Let (X, d) denote a compact metric space, and let A, B be non-empty closed subsets of X. The *Hausdorff metric* D_X between A and B is defined, as usual, by

$$D_X(A, B) = \inf\{\varepsilon > 0 \,|\, A \subseteq U_\varepsilon(B) \wedge B \subseteq U_\varepsilon(A)\}, \quad (14)$$

where

$$U_\varepsilon(A) = \{x \in X \,|\, D(x, A) < \varepsilon\}, \quad (15)$$

and

$$D(x, A) = \inf_{a\in A} d(x, a). \quad (16)$$

By $\mathbb{K}(X)$ we denote the space of all nonempty compact subsets of X, equipped with the Hausdorff metric D_X. It is well known (c.f. [**?**]) that $(\mathbb{K}(X), D_X)$ is compact, complete and separable whenever X is compact, complete and separable.

A fuzzy set A on the space X is a function $A : X \to I$ where I denotes the closed unit interval $[0, 1]$. The *α-cuts* (or the *α-level sets*) $[A]_\alpha$ and the *support* $supp(A)$ of a given fuzzy set A are defined as usual by -

$$[A]_\alpha = \{x \in X \,|\, A(x) \geq \alpha\}, \ \alpha \in [0, 1], \quad (17)$$

and

$$supp(A) = \overline{\{x \in X \,|\, A(x) > 0\}}. \quad (18)$$

Further, we define $\mathbb{F}(X)$ as the system of all upper semi-continuous fuzzy sets $A : X \to I$ having compact supports. Moreover, let

$$\mathbb{F}^1(X) = \{A \in \mathbb{F}(X) \,|\, A(x) = 1 \text{ for some } x \in X\} \quad (19)$$

denote the system of all *normal* fuzzy sets on X. Finally, we define $\emptyset_X$ as the *empty fuzzy set* ($\emptyset_X(x) = 0$ for each $x \in X$) on the space X, and $\mathbb{F}_0(X)$ as the system of all nonempty fuzzy sets.

Let us define a *levelwise* metric d_∞ on $\mathbb{F}_0(X)$ by

$$d_\infty(A, B) = \sup_{\alpha\in(0,1]} D_X([A]_\alpha, [B]_\alpha). \quad (20)$$

This equality defines the levelwise metric correctly only for non-empty fuzzy sets $A, B \in \mathbb{F}_0(X)$ whose maximal values are identical, since the Hausdorff distance D_X is only measured between two non-empty closed subsets of the space X.

ISBN: 978-989-95079-6-8

Thus, we consider the following extension of the Hausdorff metric D_X:

$$D_X(\emptyset,\emptyset) = 0 \text{ and } D_X(\emptyset, A) = diam\,(X) \qquad (21)$$

for any $A \in \mathbb{K}(X)$. With this extension, **(??)** correctly defines the levelwise metric on $\mathbb{F}(X)$. It is obvious that

$$d_\infty(\emptyset_X,\emptyset_X) = 0 \text{ and } d_\infty(\emptyset_X, A) = diam\,(X) \qquad (22)$$

for any $A \in \mathbb{F}_0(X)$.

It should be noted that the metric d_∞ is one of the three most commonly used metrics in fuzzy topological dynamics. We also recall that the metric space $(\mathbb{F}(X), d_\infty)$ is complete but is not separable and not compact and that the levelwise topology induced by d_∞ is stronger than the remaining (*sendograph* and *endograph*) ones. For more details we refer to [**?**] and to the references therein.

2.3 Zadeh's extension

Let X be a compact metric space and $\varphi \in C(X)$. Then a *fuzzification* (or *Zadeh's extension*) of the (crisp) dynamical system (X, φ) is a map $\Phi : \mathbb{F}(X) \to \mathbb{F}(X)$ defined by

$$(\Phi(A))(x) = \sup_{y\in\varphi^{-1}(x)} \{A(y)\} \qquad (23)$$

for any $A \in \mathbb{F}(X)$ and $x \in X$.

It is shown recently by [**?**] that, if X is a compact metric space, then the fuzzification $\Phi : \mathbb{F}(X) \to \mathbb{F}(X)$ is continuous if and only if $\varphi : X \to X$ is continuous. The last statement was generalized about the case of locally compact metric spaces in [**?**] recently.

It is known that, for any $\alpha \in (0, 1]$ and any $A \in \mathbb{F}(X)$,

$$\varphi([A]_\alpha) = [\Phi(A)]_\alpha. \qquad (24)$$

Similarly, $\varphi(supp(A)) = supp(\Phi(A))$ holds.

3 Preliminary results

Inspired by the results mentioned, for instance, in [**?**], we define some basic properties of generalized extensions. For any $U \subseteq X$, we define

$$e(U) = \{B \in \mathbb{F}(X) \,|\, supp(B) \subseteq U\} \qquad (25)$$

It is obvious that $e(U) \neq \emptyset$ if and only if $U \neq \emptyset$. Moreover, we have the following assertion (Lemma **??**) that was partially proved in [**?**]:.

Lemma 1 *A subset U is a non-empty open subset of X if and only if $e(U)$ is a non-empty open subset of $\mathbb{F}(X)$.*

Proof. Since the implication "⇒" has been proven in [**?**], the converse remains to be proven. So let $e(U)$ be a non-empty open subset of $(\mathbb{F}(X), d_\infty)$. Assume by contradiction that U is not open. Take any $x \in U \setminus int(U)$ and consider a fuzzy set $\chi_{\{x\}}$. Then, for any $\varepsilon > 0$, an open ε-neighborhood $V \subseteq X$ of x intersects the exterior of U. Consequently, χ_V is ε-close to $\chi_{\{x\}}$ (by the definition of d_∞), but $\chi_V \notin e(U)$. Thus no ε-neighborhood V of $\chi_{\{x\}}$ is a subset of $e(U)$. This contradicts the fact that $e(U)$ is open in $(\mathbb{F}(X), d_\infty)$. □

Lemma 2 (Representation theorem of Negoita-Ralescu, e.g. [**?**]) *Consider a family $\{B_\alpha \,|\, \alpha \in [0, 1]\}$ of closed subsets of X satisfying the following two conditions:*
(a) $B_\beta \subseteq B_\alpha \subseteq B_0$ if $0 \leq \alpha \leq \beta$,
(b) if $\{\alpha_n\}$ is an increasing sequence in I converging to α_0 then $B_{\alpha_0} = \bigcap_{n\in\mathbb{N}} B_{\alpha_n}$.
Then there exists $B \in \mathbb{F}(X)$ such that $[B]_\alpha = B_\alpha$.

Conversely, if B is a fuzzy set on X then the system $\{B_\beta\}_{\beta\in I}$ defined by $B_\beta = [B]_\beta$ for any $\beta \in (0, 1]$ and $B_0 = supp(B)$ satisfies conditions (a) and (b).

Lemma 3 *Let U, V be two subsets of X and $\varphi \in C(X)$. Then*
(i) $e(U \cap V) = e(U) \cap e(V)$,
(ii) $\Phi(e(U)) \subseteq e(\varphi(U))$,
(iii) $\Phi(e(U)) = e(\varphi(U))$ whenever U is closed.

Proof. The statements (i) and (ii) were already proved in [**?**].

The statement (iii) still remains to be proven. The inclusion $\Phi(e(U)) \subseteq e(\varphi(U))$ in (iii) follows from (ii). Let us prove $e(\varphi(U)) \subseteq \Phi(e(U))$ if U is closed. Take any $A \in e(\varphi(U))$. We want to show that there exists $B \in e(U)$ such that $\Phi(B) = A$. Since $A \neq \emptyset_X$ and upper semi-continuous, there exists $\alpha_0 = \max_{x\in X}\{A(x) \,|\, A(x) > 0\}$. Moreover, for any $\alpha \in (0, \alpha_0]$, $[A]_\alpha$ is nonempty, closed and, consequently by the continuity of φ, $\varphi^{-1}([A]_\alpha) \cap U$ is also nonempty and closed.

By the definition of the fuzzification Φ, $\max(A) = \max(B)$ whenever $B \in \mathbb{F}(X)$ is any preimage of A. So a fuzzy set $B \in e(U)$ can be defined as follows. For any $\beta \in (0, \alpha_0]$ we put

$$[B]_\beta = \varphi^{-1}([A]_\beta) \cap U. \qquad (26)$$

We also put $[B]_0 = U$ and $[B]_\beta = \emptyset$ for any $\beta \in (\alpha_0, 1]$. Obviously the system $\{[B]_\beta\}_{\beta\in I}$ satisfies the condition (a) of Lemma **??**. We shall now show that the condition (b) of Lemma **??** is satisfied, i.e. $B \in \mathbb{F}(X)$. Then, by the definition of B, we obtain that $B \in e(U)$.

Assume that $\{\beta_n\} \subseteq I$ is an increasing sequence that converges to $\beta_0 \leq \alpha_0$. Suppose by contradiction that $[B]_{\beta_0} \neq \bigcap_{n\in\mathbb{N}}[B]_{\beta_n}$. By the monotonicity of $\{\beta_n\}$ and Lemma **??** (a), the only possibility is that $[B]_{\beta_0}$ is a proper subset of $\bigcap_{n\in\mathbb{N}}[B]_{\beta_n}$, i.e.

$$\Big(\bigcap_{n\in\mathbb{N}} [B]_{\beta_n}\Big) \setminus [B]_{\beta_0} \neq \emptyset. \qquad (27)$$

Take any $x_0 \in \bigcap[B]_{\beta_n} \setminus [B]_{\beta_0}$ and then take $x_n \in [B]_{\beta_n} \setminus [B]_{\alpha_0}$ such that $\{x_n\}$ converges to x_0. Obviously, since $x_0 \notin [B]_{\beta_0}$ we obtain from **(??)** that

$$\varphi(x_0) \notin [A]_{\beta_0}. \qquad (28)$$

On the other hand, $x_n \in [B]_{\beta_n}$ for any $n \in \mathbb{N}$, i.e.,

$$\varphi(x_n) \in [A]_{\beta_n}, \text{ for any } n \in \mathbb{N}. \qquad (29)$$

Now the continuity of φ implies that $\{\varphi(x_n)\}$ converges to $\varphi(x_0)$. Hence **(??)** and **(??)** imply that $[A]_{\beta_0}$ is a proper subset of $\bigcap_{n\in\mathbb{N}}[A]_{\beta_n}$, i.e. $A \notin \mathbb{F}(X)$ by Lemma **??** – a contradiction. Thus, we have shown that $B \in \mathbb{F}(X)$.

ISBN: 978-989-95079-6-8

Finally, by the construction of B, we have $\Phi(B) = A$. □

We will need some further notation. For any $\alpha \in (0,1]$ and $U \subseteq X$, set

$$e_\alpha(U) = \{A \in \mathbb{F}(X) \mid [A]_\alpha \neq \emptyset \text{ and } [A]_\alpha \subseteq U\} \tag{30}$$

and

$$\vartheta(U) = e_1(U) \cap e(U). \tag{31}$$

Lemma 4 *For any $\alpha \in (0,1]$, $e_\alpha(U)$ is open in $(\mathbb{F}(X), \tau_\infty)$ if and only if $U \subseteq X$ is open in X.*

Proof. Let $\alpha \in (0,1]$ be fixed. We shall show that $e_\alpha(U)$ is open in $(\mathbb{F}(X), \tau_\infty)$ if U is open. Take any $A \in e_\alpha(U)$. Since $[A]_\alpha \subseteq U$ is closed and U is open, there exists an open ε-neighborhood V of $[A]_\alpha$ lying in U for any $\varepsilon > 0$. Consequently, by the definition of d_∞, if we consider an open ε-neighborhood $V' \subseteq (\mathbb{F}(X), d_\infty)$ of A with $\varepsilon < diam(X)$, then for any $B \in V'$,

$$[B]_\alpha \subseteq V \subseteq U. \tag{32}$$

Thus, there exists an open neighborhood V' of A in $e_\alpha(U)$, i.e., $e_\alpha(U)$ is open in $(\mathbb{F}(X), \tau_\infty)$.

Let us prove the converse implication. Assume by contradiction that U is not open and take any $x \in U \setminus int(U)$. Then, by the definition of $e_\alpha(U)$, $\chi_{\{x\}}$ belongs to $e_\alpha(U)$ but no ε-neighborhood of $\chi_{\{x\}}$ is a subset of $e_\alpha(U)$, i.e. $e_\alpha(U)$ is not open - a contradiction. □

Corollary 1 *A subset $U \subseteq X$ is open in X if and only if $\vartheta(U)$ is open in $(\mathbb{F}(X), \tau_\infty)$ (and therefore also in $(\mathbb{F}^1(X), d_\infty)$).*

Lemma 5 *Let X be a compact metric space and $\varphi \in C(X)$ Then, for any $\alpha \in (0,1]$ and $U, V \subseteq X$,*
(i) $e_\alpha(U \cap V) = e_\alpha(U) \cap e_\alpha(V)$,
(ii) $\Phi(e_\alpha(U)) \subseteq e_\alpha(\varphi(U))$.

Proof. Clearly, for any $\alpha \in (0,1]$, $A \in e_\alpha(U \cap V)$ if and only if $[A]_\alpha \subseteq U \cap V$, if and only if $[A]_\alpha \subseteq U$ and $[A]_\alpha \subseteq V$, if and only if $A \in e_\alpha(U)$ and $A \in e_\alpha(V)$, if and only if $A \in e_\alpha(U) \cap e_\alpha(V)$. Thus, (i) holds.

Let us prove (ii). Consider any $A \in \Phi(e_\alpha(U))$. Then there exists $B \in e_\alpha(U)$ for which $\Phi(B) = A$. Since $[B]_\alpha \subseteq U$ it follows from **(??)** and from the continuity of φ that

$$\varphi([B]_\alpha) = [\Phi(B)]_\alpha = [A]_\alpha \subseteq \varphi(U), \tag{33}$$

i.e. $A \in e_\alpha(\varphi(U))$, and the inclusion is valid for any $\alpha \in (0,1]$. □

We are ready to modify Lemma **??** and to prove the next lemma, which is used for the further study of dynamics in the space of normal fuzzy sets on X. For completeness, we note the obvious fact that $\vartheta(U) \neq \emptyset$ if and only if $U \neq \emptyset$.

Lemma 6 *Let U, V be two subsets of X and $\varphi \in C(X)$. Then*
(i) $\vartheta(U \cap V) = \vartheta(U) \cap \vartheta(V)$,
(ii) $\Phi(\vartheta(U)) \subseteq \vartheta(\varphi(U))$,
(iii) $\Phi(\vartheta(U)) = \vartheta(\varphi(U))$ whenever U is closed.

Proof. The statements (i) and (ii) are easy consequences of Lemmas **??** and **??**. Moreover, the proof of the statement (iii) is a slight modification of the proof of Lemma **??** ($\max_{x \in X} A(x) = \alpha_0 = 1$). □

At the end of this section we mention two simple properties of the usual fuzzification. By χ_A we denote the characteristic function of a given set $A \subseteq X$.

Lemma 7 *Let X be a compact metric space, $\varphi \in C(X)$ let Φ be a fuzzification of φ. Then*

$$\Phi(\alpha\chi_A) = \alpha\chi_{\varphi(A)} \tag{34}$$

for any $\alpha \in (0,1]$.

Proof. Obvious. □

Lemma 8 *Let X be a compact metric space. Then the map $i : (\mathbb{K}(X), D_X) \to (\mathbb{F}(X), d_\infty)$ defined by $i(A) = \chi_A$ for any $A \in \mathbb{K}(X)$ is an isometrical embedding.*

Proof. It is obvious that the map i is correctly defined. It follows directly from the definition of the metric d_∞ on the space of fuzzy sets that

$$d_X(x,y) = d_\infty(\chi_{\{x\}}, \chi_{\{y\}}) \tag{35}$$

for any $x, y \in X$, and also

$$D_X(U,V) = d_\infty(\chi_U, \chi_V) \tag{36}$$

for any $U, V \in \mathbb{K}(X)$. Thus, by Lemma **??**, i is injective, continuous and isometric. □

Remark 1 *The same assertion is true when we consider either the sendograph or endograph topology instead of the levelwise one, since the equalities used in the proof are valid also in those topologies. Consequently, some results (Propositions **??** and **??**) presented in the next section are valid also in all mentioned topologies. The property mentioned in Lemma **??** was firstly mentioned in Kloeden's pioneering paper [**?**] in a little bit different setting.*

4 Chaotic and mixing properties.

Firstly we study some chaotic properties. It follows from Lemma **??** that any fuzzified system is an extension of the set-valued dynamical system $(\mathbb{K}(X), \tilde{\varphi})$ induced by the original dynamical system (X, φ). Consequently, we can use results published in [**?**] to easily provide some answers (Propositions **??** and **??**) to the question mentioned at the beginning of this paper.

Lemma 9 [**?**] *Let X be a compact metric space and $\varphi \in C(X)$.*

(i) If there is a set $S \subseteq X$ that is d_jC-scrambled ($j = 1,2,3$) (ω-scrambled, LY-scrambled, resp.) for the map φ then there exists d_jC-scrambled ($j = 1,2,3$) (ω-scrambled, LY-scrambled, resp.) set for $\tilde{\varphi}$ with the same cardinality as S.

(ii) If φ is d_jC (LYC, ωC, resp.) then the same holds for $\tilde{\varphi}$.

ISBN: 978-989-95079-6-8

Lemma 10 [**?**] *There exists a compact metric space X with a zero topological entropy map φ for which there exist no LY pairs, d_3C-scrambled sets, or ω-scrambled sets, such that $\tilde{\varphi}$ is PTE, d_1C, ωC and LYC.*

Proposition 1 *Let X be a compact metric space, $\varphi \in C(X)$ and let Φ be a fuzzification of φ. Then*

(i) If there is a set $S \subseteq X$ that is d_jC-scrambled ($j = 1, 2, 3$) (ω-scrambled, LY-scrambled, resp.) for the map φ, then there exists a d_jC-scrambled ($j = 1, 2, 3$) (resp. ω-scrambled, LY-scrambled) set S' for Φ, with the same cardinality as S.

(ii) If φ is d_jC (resp. LYC, ωC), then the same holds for Φ.

Proof. Let $S' = \{A \in \mathbb{F}(X) \mid A = \chi_{\{x\}} \text{ and } x \in S\}$. Then this proposition is a corollary of Lemmas **??**, **??** and **??**. □

Proposition 2 *Let Φ denote a fuzzification of φ. There exists a compact metric space X with a zero topological entropy map φ for which there exists no LY pairs, neither d_3C-scrambled set, nor ω-scrambled set such that Φ is PTE, d_1C, ωC and LYC.*

Proof. This proposition is an easy consequence of Lemmas **??**, **??** and **??**, and of the fact that the three versions of distributional chaos are comparable (**??**). □

The following proposition justifies why we study dynamical properties of fuzzified dynamical systems on the space of normal fuzzy sets in the rest of this paper.

Proposition 3 *Let X be a compact metric space and $\varphi \in C(X)$. Then no fuzzification $\Phi : (\mathbb{F}(X), d_\infty) \to (\mathbb{F}(X), d_\infty)$ of φ is transitive.*

Proof. Take any $A, B \in \mathbb{F}(X)$ such that $\max(A) \neq \max(B)$. By the definition of d_∞, there is an open ε-neighborhood U_ε of A (resp., an open ϑ-neighborhood V_ϑ of B) for some $\varepsilon, \vartheta > 0$ such that $\max(A') = \max(A)$ for any $A' \in U_\varepsilon$ (resp., $\max(B') = \max(B)$ for any $B' \in V_\vartheta$). Moreover, Φ preserves the maxima of fuzzy sets from U_ε and V_ϑ. Thus, by the choice of A and B, $\Phi^n(U_\varepsilon) \cap V_\vartheta = \emptyset$ (resp., $U_\varepsilon \cap \Phi^n(V_\vartheta) = \emptyset$) for any $n \in \mathbb{N}$, i.e., Φ is not transitive. □

Corollary 2 *Let X be a compact metric space and $\varphi \in C(X)$. Then a fuzzification $\Phi : (\mathbb{F}(X), d_\infty) \to (\mathbb{F}(X), d_\infty)$ of φ has none of the properties listed in (**??**) (specification property, strong mixing, mild mixing, weak mixing and total transitivity).*

Proof. The proof is an obvious consequence of Proposition **??** and (**??**). □

For completeness, it follows directly from Lemmas **??** and (**??**) that if the original dynamical system (X, φ) is topologically chaotic then also the fuzzified system $(\mathbb{F}^1(X), \Phi)$ (and hence $(\mathbb{F}(X), \Phi)$) is topologically chaotic. Now we would like to mention a recent paper [**?**] where the conditions defining Devaney chaos ([**?**]) were studied on the space of all fuzzy sets on X. We recall that we extended their results since no such fuzzification can be Devaney chaotic by Lemma **??**.

Moreover, dynamical properties of fuzzifications on the space of normal fuzzy sets on X have been never studied before. So we study them in the rest of this section.

Proposition 4 *Let X be a compact metric space, $\varphi \in C(X)$ and let Φ be the fuzzification of φ. If $\Phi : (\mathbb{F}^1(X), d_\infty) \to (\mathbb{F}^1(X), d_\infty)$ is transitive then φ is transitive, but the converse implication does not hold.*

Proof. Let the assumptions be fulfilled. Consider any open subsets $U, V \subseteq X$. We want to show that φ is transitive, i.e. there exists $n \in \mathbb{N}$ such that

$$\varphi^n(U) \cap V \neq \emptyset. \tag{37}$$

First we take open subsets $U', V' \subseteq X$ such that $U' \subseteq \overline{U'} \subseteq U$ and $V' \subseteq \overline{V'} \subseteq V$. Then, by Corollary **??**, $\vartheta(U')$ and $\vartheta(V')$ are open subsets of $(\mathbb{F}^1(X), \tau_\infty)$. Since Φ is transitive, there exists $n \in \mathbb{N}$ for which $\Phi^n(\vartheta(U')) \cap \vartheta(V') \neq \emptyset$. Then $\Phi^n(\vartheta(\overline{U'})) \cap \vartheta(\overline{V'}) \neq \emptyset$, and by using results (i) and (iii) of Lemma **??**, we obtain

$$\vartheta(\varphi^n(\overline{U'})) \cap \vartheta(\overline{V'}) \neq \emptyset \quad \Rightarrow \quad \vartheta(\varphi^n(\overline{U'}) \cap \overline{V'}) \neq \emptyset. \tag{38}$$

Therefore $\varphi^n(\overline{U'}) \cap \overline{V'} \neq \emptyset$, and consequently, (**??**) is proved.

A counterexample (an irrational rotation on the unit circle) showing that the converse implication does not hold was presented in [**?**] for the fuzzification Φ defined on $\mathbb{F}(X)$, but it can be applied also to $\Phi|_{\mathbb{F}^1(X)}$. □

Proposition 5 *Let X be a compact metric space, $\varphi \in C(X)$, and let Φ be the fuzzification of φ. If $\Phi : (\mathbb{F}^1(X), d_\infty) \to (\mathbb{F}^1(X), d_\infty)$ is totally transitive, then φ is totally transitive, but the converse implication does not hold.*

Proof. This proposition is an easy corollary of Proposition **??**, since the transitivity of any iteration Φ^n of Φ on $\mathbb{F}^1(X)$ implies the transitivity of φ^n on X for any $n \in \mathbb{N}$.

As a counterexample to the converse, we again can use the example mentioned in [**?**]. Their example was used as the counterexample to

$$\varphi \text{ transitive } \not\Rightarrow \ \Phi \text{ transitive}, \tag{39}$$

but the map φ used in [**?**] is an irrational rotation on the unit circle, and it is obvious that any irrational rotation on the unit circle is totally transitive. □

Proposition 6 *Let X be a compact metric space, $\varphi \in C(X)$ and let Φ be the fuzzification of φ. If $\Phi : (\mathbb{F}^1(X), d_\infty) \to (\mathbb{F}^1(X), d_\infty)$ is sensitively dependent, then φ is sensitively dependent.*

Proof. Let the assumptions be fulfilled. We shall show that φ is sensitively dependent for the same sensitivity constant δ as Φ. Take any $x \in X$ and any open neighborhood U of x. Then $\vartheta(U)$ is an open neighborhood of $\chi_{\{x\}}$ by Corollary **??**. Since Φ is sensitively dependent, there exist $A \in \vartheta(U)$ and $n \in \mathbb{N}$ such that

$$d_\infty(\Phi^n(\chi_{\{x\}}), \Phi^n(A)) > \delta. \tag{40}$$

ISBN: 978-989-95079-6-8

By Lemma **??**

$$d_\infty(\Phi^n(\chi_{\{x\}}), \Phi^n(A)) = d_\infty(\chi_{\{\varphi^n(x)\}}, \Phi^n(A)) > \delta. \quad (41)$$

Thus, by the definition of the metric d_∞, there exists $y_0 \in supp(A)$ such that $d_X(\varphi^n(x), \varphi^n(y_0)) > \delta$. Obviously, y_0 lies in U and this shows that φ is sensitively dependent. □

Proposition 7 *Let X be a compact metric space, $\varphi \in C(X)$, and let Φ be the fuzzification of φ. If $\Phi : (\mathbb{F}^1(X), d_\infty) \to (\mathbb{F}^1(X), d_\infty)$ is exact, then φ is exact.*

Proof. Let the assumptions be fulfilled, and let $U \subseteq X$ be an open set. Then there exists a closed $U' \subseteq U$ with nonempty interior $int(U')$. By Corollary **??**, $\vartheta(int(U'))$ is an open subset of $\mathbb{F}^1(X)$. Since Φ is exact, there exists some $k \in \mathbb{N}$ for which $\Phi^k(\vartheta(int(U'))) = \mathbb{F}^1(X)$. Therefore, by the choice of U' and by Lemma **??**,

$$\mathbb{F}^1(X) = \Phi^k(\vartheta(int(U'))) \subseteq \Phi^k(\vartheta(U')) \quad (42)$$

and

$$\Phi^k(\vartheta(U')) = \vartheta(\varphi^k(U')) \subseteq \vartheta(\varphi^k(U)). \quad (43)$$

Thus, $\vartheta(\varphi^k(U))$ covers $\mathbb{F}^1(X)$ and hence, since $\mathbb{F}^1(X)$ also contains the characteristic functions of all singletons (see Lemma **??**), $\varphi^k(U) = X$. □

Remark 2 *The map Φ on $\mathbb{F}(X)$ cannot be exact since Φ cannot be transitive (see Lemma **??**). Moreover, the validity of the converse implication to Proposition **??** is still unknown.*

Proposition 8 *Let X be a compact metric space, $\varphi \in C(X)$ and let Φ be the fuzzification of φ. If $\Phi : (\mathbb{F}^1(X), d_\infty) \to (\mathbb{F}^1(X), d_\infty)$ is strongly mixing then φ is strongly mixing.*

Proof. The proof of this proposition is a slight variation of the proof of Proposition **??**. □

Proposition 9 *Let X be a compact metric space, $\varphi \in C(X)$ and let Φ be the fuzzification of φ. If $\Phi : (\mathbb{F}^1(X), d_\infty) \to (\mathbb{F}^1(X), d_\infty)$ is mildly mixing then φ is mildly mixing.*

Proof. Let the assumptions be fulfilled. Let Y be any compact metric space and (Y, ψ) be any transitive discrete dynamical system. According to the definition of the mild mixing property, we want to show that the product system $(X \times Y, \varphi \times \psi)$ is transitive, i.e. for any open sets $U, V \subseteq X \times Y$ there exists $k \in \mathbb{N}$ for which

$$(\varphi \times \psi)^k(U) \cap V \neq \emptyset. \quad (44)$$

So let $U, V \subseteq X \times Y$ be given. Then there are open subsets $U_1, V_1 \subseteq X$, $U_2, V_2 \subseteq Y$ for which $U_1 \times U_2 \subseteq U$ and $V_1 \times V_2 \subseteq V$. We also consider open subsets $U_1', V_1' \subseteq X$ such that $\overline{U_1'} \subseteq U_1$ and $\overline{V_1'} \subseteq V_1$. By Corollary **??**, $\vartheta(U_1')$ and $\vartheta(V_1')$ are open subsets of $\mathbb{F}^1(X)$. Thus, since Φ is mildly mixing, there exists $k \in \mathbb{N}$ for which

$$(\Phi \times \psi)^k(\vartheta(U_1') \times U_2) \cap (\vartheta(V_1') \times V_2) \neq \emptyset, \quad (45)$$

i.e.

$$\psi^k(U_2) \cap V_2 \neq \emptyset \quad (46)$$

and also

$$\Phi^k(\vartheta(U_1')) \cap \vartheta(V_1') \neq \emptyset. \quad (47)$$

As in the proof of Proposition **??** ($U_1' = U'$ and $V_1' = V'$), the last inequality implies that

$$\varphi^k(U_1) \cap V_1 \neq \emptyset. \quad (48)$$

Consequently, this together with (**??**) gives

$$(\varphi \times \psi)^k(U_1 \times U_2) \cap (V_1 \times V_2) \neq \emptyset. \quad (49)$$

Finally, by the choice of U_1, U_2, V_1, V_2, (**??**) is proven. □

Proposition 10 *Let X be a compact metric space, $\varphi \in C(X)$ and let Φ be the fuzzification of φ. If $\Phi : (\mathbb{F}^1(X), d_\infty) \to (\mathbb{F}^1(X), d_\infty)$ is weakly mixing, then φ is weakly mixing.*

Proof. The proof of this proposition is a slight variation of the proof of Proposition **??**. □

References

[1] Zadeh L. *Fuzzy sets and systems, Proc. Symp. on Systems Theory*. Polytechnic Institute Press, Brooklyn, New York, 1965.

[2] Diamond P. and Pokrovskii A. Chaos, entropy and a generalized extension principle. *Fuzzy Sets and Systems*, 61:277 – 283, 1994.

[3] Kloeden P. Chaotic iterations of fuzzy sets. *Fuzzy Sets and Systems*, 42:37 – 42, 1991.

[4] Román-Flores H. and Chalco-Cano Y. Some chaotic properties of zadeh's extension. *Chaos, Solitons and Fractals*, 35:452 – 459, 2008.

[5] Kupka J. On fuzzifications of discrete dynamical systems. *Fuzzy Sets and Systems*, submitted.

[6] García Guirao J. L., Kwietniak D., Lampart M., Oprocha P., and Peris A. Chaos on hyperspaces. *Nonlinear Analysis: Theory, Methods & Applications*, accepted manuscript.

[7] Schweizer B. and Smítal J. Measures of chaos and a spectral decomposition of dynamical systems on the interval. *Trans. Amer. Math. Soc.*, 344:737 – 754, 1994.

[8] Balibrea F., Smítal J., and Štefánková M. The three versions of ditributional chaos. *Chaos, Solitons and Fractals*, 23:1581 – 1583, 2005.

[9] Smítal J. and Štefánková M. Distributional chaos for triangular maps. *Chaos, Solitons and Fractals*, 21:1125 – 1128, 2004.

[10] Li S. ω-chaos and topological entropy. *Trans. Amer. Math. Soc.*, 339(1):243 – 249, 1993.

[11] Banks J., Brooks J., Cairns G., Davis G., and Stacey P. On devaney's definition of chaos. *Amer. Math. Monthly*, 99:332 – 334, 1992.

[12] Glasner E. and Weiss B. Sensitive dependence on initial conditions. *Nonlinearity*, 6:1067 – 1075, 1993.

[13] Bowen R. Entropy for group endomorphisms and homogeneous spaces. *Trans. Amer. Math. Soc.*, 153:401 – 414, 1971.

[14] Kuratowski K. *Topology*. Vol. II, Academic Press, London, New York, 1968.

[15] Diamond P. and Kloeden P. *Metric spaces of fuzzy sets: theory and applications*. World Scientific, Singapore, 1994.

[16] Kloeden P. Fuzzy dynamical systems. *Fuzzy Sets and Systems*, 7:275 – 296, 1982.

[17] Devaney R. *An introduction to chaotic dynamical systems*. 2nd ed. Boulder: Westview Press, 2003.

ISBN: 978-989-95079-6-8

Associativity of triangular norms in light of web geometry

Milan Petrík[1,2] Peter Sarkoci[3]

1. Institute of Computer Science, Academy of Sciences of the Czech Republic,
Prague, Czech Republic
2. Center for Machine Perception, Department of Cybernetics
Faculty of Electrical Engineering, Czech Technical University
Prague, Czech Republic
3. Department of Knowledge-Based Mathematical Systems, Johannes Kepler University,
Linz, Austria
E-mails: {petrik@cs.cas.cz|petrikm@cmp.felk.cvut.cz}, peter.sarkoci@{jku.at|gmail.com}

Abstract— The aim of this paper is to promote web geometry and, especially, the Reidemeister closure condition as a powerful and intuitive tool characterizing associativity of the Archimedean triangular norms. In order to demonstrate its possible applications, we provide the full solution to the problem of convex combinations of nilpotent triangular norms.

Keywords— Archimedean triangular norm, web geometry, Reidemeister closure condition.

1 Triangular norms

The notion of *triangular norm* was originally introduced within the framework of probabilistic metric spaces [12]. Since then, triangular norms have found diverse applications in the theory of fuzzy sets, fuzzy decision making, in models of certain many-valued logics or in multivariate statistical analysis; for a reference see the books by Alsina, Frank, and Schweizer [5] and by Klement, Mesiar, and Pap [8].

A conjunctor is a function $K\colon [0,1]^2 \to [0,1]$ which is nondecreasing in both arguments, commutative, and which satisfies the *boundary condition* $T(x,1) = x$ for all $x \in [0,1]$. A triangular norm (shortly a *t-norm*, usually denoted by T) is a conjunctor which satisfies the *associativity equation* $T(T(x,y),z) = T(x,T(y,z))$ for all $x,y,z \in [0,1]$. This paper deals mainly with *Archimedean t-norms*. Let us recall that for every t-norm T, a number $n \in \mathbb{N}\cup\{0\}$, and $x \in [0,1]$ a *natural power* of x, denoted $(x)_T^{(n)}$, is defined by:

$$(x)_T^{(n)} = \begin{cases} 1 & \text{if} \quad n = 0\,, \\ T\left(x,(x)_T^{(n-1)}\right) & \text{if} \quad n > 0\,. \end{cases} \tag{1}$$

A t-norm T is said to be *Archimedean* if and only if for every pair $x,y \in \,]0,1[$, $x < y$, there exists a natural number $n \in \mathbb{N}$ such that $(y)_T^{(n)} < x$. A t-norm which is continuous and strictly increasing on the half-open square $]0,1]^2$ is said to be *strict*. A continuous t-norm T is called *nilpotent* if and only if for every $x \in \,]0,1[$ there exists a natural number $n \in \mathbb{N}$ such that $(x)_T^{(n)} = 0$. A prototypical example of a strict and a nilpotent t-norm is the product t-norm, $T_{\mathbf{P}}(x,y) = x \cdot y$, and the Łukasiewicz t-norm $T_{\mathbf{L}}(x,y) = \max\{x+y-1,0\}$,

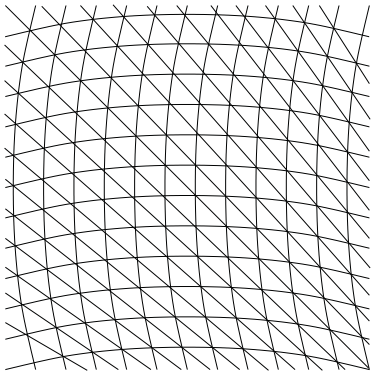


Figure 1: Example of a complete 3-web.

respectively. It is possible to show that a continuous Archimedean t-norm is either strict or nilpotent. A t-norm T is said to be *cancellative* if it satisfies $T(a,b) = T(a,c) \Rightarrow b = c$ for every $a,b,c \in [0,1]$, $a \neq 0$. A t-norm T is said to be *weakly cancellative* [9] if it satisfies $T(a,b) = T(a,c) \Rightarrow b = c$ for every $a,b,c \in [0,1]$ with $T(a,b) \neq 0$ and $T(a,c) \neq 0$. Every cancellative t-norm is weakly cancellative. Under the assumption of continuity, the set of cancelative t-norms and the set of strict t-norms coincide. Under the same assumption, the set of weakly cancelative t-norms coincides with the set of Archimedean t-norms.

2 Web geometry and local loops

In this section, *web geometry* is explained as a tool allowing to visualize algebraic identities. In particular, it is shown that it visualizes associativity of Archimedean t-norms. A detailed introduction to the subject is given in the monograph by Blashke and Bol [6]. Also the collection of papers by Aczél, Akivis, and Goldberg [1, 2, 3] can serve as an (English) introductory text.

Definition 2.1 *A* groupoid *(or a* magma*) is an algebraic structure $\mathcal{G} = (G,\circ)$ on a set G where $\circ\colon G\times G \to G$ is a binary operation. A* quasigroup *is a groupoid in which the*

ISBN: 978-989-95079-6-8

equations $a \circ x = b$ and $y \circ a = b$ have unique solutions for every a and b in G. Finally, a loop *is an algebraic structure $\mathcal{L} = (G, \circ, e)$ where $(G, \circ)$ is a quasigroup and e is an identity element, i.e. $x \circ e = x = e \circ x$ for every $x \in G$.*

The definition of a quasigroup, $\mathcal{G} = (G, \circ)$, allows to define the left and the right inverse. For this purpose we introduce a prefix notation $g(x, y) = x \circ y$.The left inverse is then defined for every $u \in G$ as ${}^{-1}g(u, y) = x$ such that $g(x, y) = u$. Similarly, the right inverse is, for every $v \in G$, $g^{-1}(x, v) = y$ such that $g(x, y) = v$. We say that g is *invertible* in x, resp. y. Defining a point $A = (a, b) \in G \times G$ we may introduce a new operation, $\bullet\colon G \times G \to G$, as

$$u \bullet v = g\left({}^{-1}g(u, b), g^{-1}(a, v)\right) . \qquad (2)$$

It is easy to show that $(G, \bullet)$ is a loop with a unit element $e = g(a, b)$; this loop is called a *local loop* of the quasigroup $(G, \circ)$ at the point $A = (a, b)$.

Definition 2.2 *Let M be a non-empty set and let $\lambda_1, \lambda_2, \lambda_3$ be three families of subsets of M. To elements of M we refer as to points and to elements of λ_α, $\alpha \in \{1, 2, 3\}$, as to lines. We say that the system $(M, \lambda_1, \lambda_2, \lambda_3)$ is a* complete 3-web *if and only if the following conditions hold:*

1. *Any point $p \in M$ is incident to just one line of each family λ_α, $\alpha \in \{1, 2, 3\}$.*
2. *Any two lines of different families are incident to exactly one point of M.*
3. *Two distinct lines from the same family λ_α are disjoint.*

Note that Item 3 of the definition is redundant and can be derived from Item 1. We kept the definition in this redundant form in order to keep some of the furher considerations simpler.

Usually M is equipped with a topology which turns the set into a manifold. In such a case the families λ_α are often required to be foliations. Figure 1 shows such an example of 3-web where M is a two dimmensional domain in plane and λ_α are foliations of co-dimension 1.

Notice that all the examples of 3-webs, shown in figures of this paper, are simplified. That means that only some particular lines from uncountable sets λ_α, $\alpha \in \{1, 2, 3\}$, are drawn.

Complete 3-webs are closely connected to quasigroups and, especially, loops. Every quasigroup defines a 3-web in the following way. Let $\mathcal{G} = (G, \circ)$ be a quasigroup defined on a manifold $M = G \times G$ and let $x, y \in G$. Let the lines of sets λ_1, λ_2, and λ_3 be given by the equations $x = a$, $y = b$, and $x \circ y = c$ respectively for fixed $a, b, c \in G$. The pair (M, λ_α), $\alpha \in \{1, 2, 3\}$, is then a 3-web.

Conversely, a 3-web (M, λ_α), $\alpha \in \{1, 2, 3\}$, on a manifold M, defines a binary operation which is invertible with respect to both operands; yet this time the result is not unique. Denote one set of the lines of the 3-web as λ_x, second as λ_y, and third as λ_z (the lines in the set λ_z are called *contour lines*), let $A \in M$ and let $g_x \in \lambda_x$ and $g_y \in \lambda_y$ be the lines passing through A; cf. Figure 2. Now we define an operation $*\colon \lambda_z \times \lambda_z \to \lambda_z$. Take $u, v \in \lambda_z$, define points $U, V \in M$ as $U = u \cap g_x$ and $V = v \cap g_y$, respectively. Let $g_u \in \lambda_y$ be the line passing through U and let $g_v \in \lambda_x$ be the line passing through V. Denote the intersection of these two lines as $W = g_u \cap g_v$; the line w passing through W is the result of the operation $u * v$.

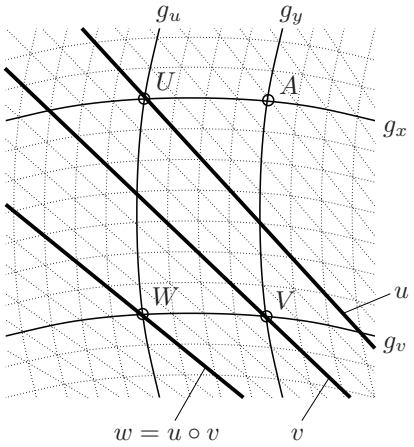


Figure 2: Operation defined on a complete 3-web.

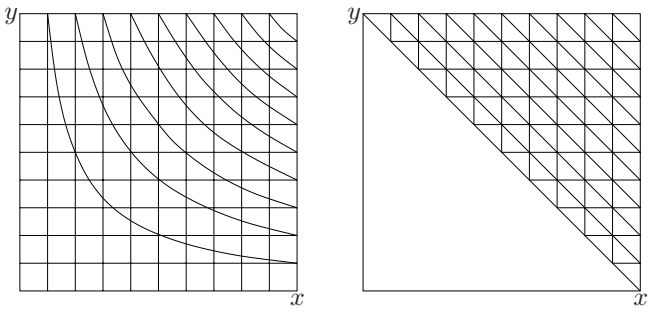


Figure 3: 3-webs given by the product t-norm, $T_{\mathbf{P}}$, and by the Łukasiewicz t-norm, $T_{\mathbf{L}}$.

It can be shown easily that the operation $*$ is invertible in both variables. Moreover, the line $e \in \lambda_z$, passing through the point A, behaves, with respect to the oparation $*$, as the unit element. Thus $(\lambda_z, *)$ is a loop. Moreover, this loop coincides, up to an isomorphism, with the local loop of $(G, \circ)$ at the point A.

T-norms, defined on the unit interval $[0, 1]$, form neither groupoids nor loops. Nevertheless, the 3-web given by a continuous Archimedean (i.e. strict or nilpotent) t-norm T satisfies all the requirements given by Definition 2.2 if the manifold M is defined as the subset of $[0, 1]^2$ where the t-norm attains non-zero values. A manifold, induced this way by a binary operation $T\colon [0, 1]^2 \to [0, 1]$, will be denoted as $\operatorname{Man} T$. Thus

$$M = \operatorname{Man} T = \left\{(x, y) \in [0, 1]^2 \mid T(x, y) > 0\right\} . \qquad (3)$$

Figure 3 shows 3-webs given by $T_{\mathbf{P}}$ and $T_{\mathbf{L}}$ as examples of a strict and a nilpotent t-norm respectively.

ISBN: 978-989-95079-6-8

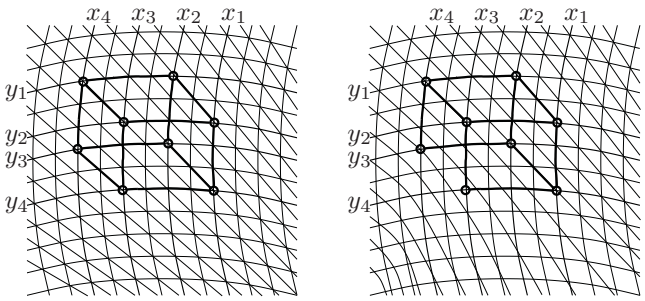


Figure 4: Example of a closed Reidemeister figure and a non-closed Reidemeister figure.

3 Reidemeister closure condition

Different types of 3-webs are characterized by closure conditions. These closure conditions have their counterparts in the related loops as algebraic properties of the loop operations. In this text we are interested in the associativity of t-norms since the associativity is the only property of a t-norm which cannot be intuitively interpreted from its graph. The 3-web counterpart of the associativity is the *Reidemeister closure condition*.

A 3-web satisfies the Reidemeister closure condition if and only if every Reidemeister figure in this web is closed; see Figure 4. Described in the terminology of contour lines, the Reidemeister closure condition is as follows. Let $x_1, x_2, x_3, x_4 \in \lambda_x$ and $y_1, y_2, y_3, y_4 \in \lambda_y$. If the points $(x_1 \cap y_2)$ and $(x_2 \cap y_1)$ lie on the same contour line (i.e. a line from the set λ_z), and if so do the pair of points $(x_1 \cap y_4)$ and $(x_2 \cap y_3)$, and the pair of points $(x_3 \cap y_2)$ and $(x_4 \cap y_1)$, then the points $(x_3 \cap y_4)$ and $(x_4 \cap y_3)$ also lie on the same contour line. The left part of Figure 4 shows a 3-web which satisfies the Reidemeister closure condition whereas the right part shows a 3-web where the condition is violated.

Let us now define a relation $\sim \subseteq M \times M$. We say that two points $A, B \in M$ are in the relation $\sim$, and we write $A \sim B$, if and only if they are both elements of the same line from the set λ_z. Using the language of $\sim$ the Reidemeister condition reads as:

$$\Big(\big((x_1 \cap y_2) \sim (x_2 \cap y_1)\big) \& \big((x_1 \cap y_4) \sim (x_2 \cap y_3)\big) \& \big((x_3 \cap y_2) \sim (x_4 \cap y_1)\big)\Big) \Rightarrow \big((x_3 \cap y_4) \sim (x_4 \cap y_3)\big). \quad (4)$$

4 Reidemeister closure condition and continuous Archimedean t-norms

We are going to show that the level set system of a continuous Archimedean t-norm always satisfies the Reidemeister closure condition. Moreover, the Reidemeister closure condition characterizes the associativity of a continuous Archimedean t-norm. For this purpose we recall that if we relax the requirement of the associativity in the definition of a t-norm, we obtain the definition of a conjunctor.

Let K be a continuous Archimedean conjunctor; for the sake of compactness we will use the infix notation: $K(x, y) = x \circ y$. Let (M, λ_α) be the corresponding 3-web defined on the manifold $M = \operatorname{Man} K = \{(x, y) \in [0, 1]^2 \mid x \circ y > 0\}$. Such a 3-web, as in the case of the continuous Archimedean t-norms, satisfies all the requirements given by Definition 2.2. In view of the previous section, the 3-web given by $\circ$ satisfies the Reidemeister closure condition if and only if the following condition holds for any $x_1, x_2, x_3, x_4, y_1, y_2, y_3, y_4 \in M$:

$$\Big(\big(x_1 \circ y_2 = x_2 \circ y_1\big) \& \big(x_1 \circ y_4 = x_2 \circ y_3\big) \& \big(x_3 \circ y_2 = x_4 \circ y_1\big)\Big) \Rightarrow \big(x_3 \circ y_4 = x_4 \circ y_3\big). \quad (5)$$

In the sequel, this condition will be denoted by $\mathbf{R}$.

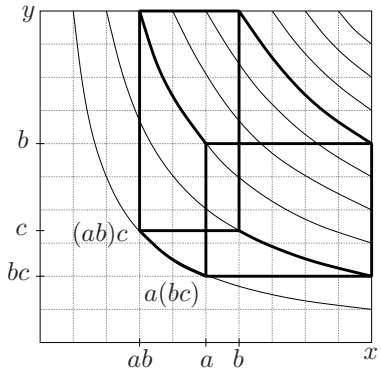


Figure 5: Reidemeister figure on the 3-web given by a continuous Archimedean t-norm.

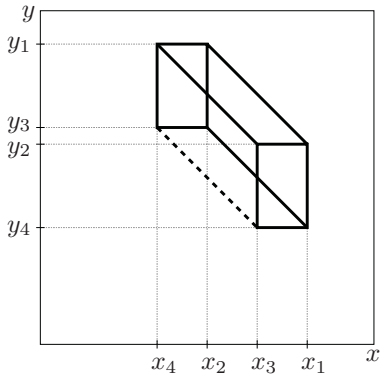


Figure 6: Illustration for the proof: an arbitrary Reidemeister figure.

Figure 5 shows that $\mathbf{R}$ implies the associativity of $\circ$ in a rather intuitive way. For any $a, b, c \in [0, 1]^2$, such that $(a \circ b) \circ c \in \operatorname{Man} \circ$ and $a \circ (b \circ c) \in \operatorname{Man} \circ$, there can be constructed a Reidemeister figure which, as can be seen in Figure 5, is closed if and only if $(a \circ b) \circ c = a \circ (b \circ c)$.

Figure 5 shows also that, in the other way round, if $\circ$ is associative then all the Reidemeister figures that "touch" the lines given by $x = 1$ and $y = 1$ are closed. In other words, $\mathbf{R}$ has to be satisfied for any $x_2, x_3, x_4, y_2, y_3, y_4 \in [0, 1]$ and, at least, for $x_1 = y_1 = 1$. Let us denote this weaker condition by $\mathbf{R_1}$. It is now clear that K is associative if and only if it

ISBN: 978-989-95079-6-8

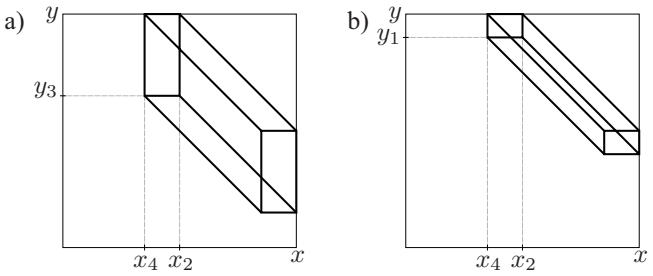


Figure 7: Illustration for the proof.

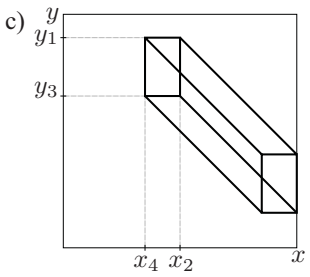


Figure 8: Illustration for the proof.

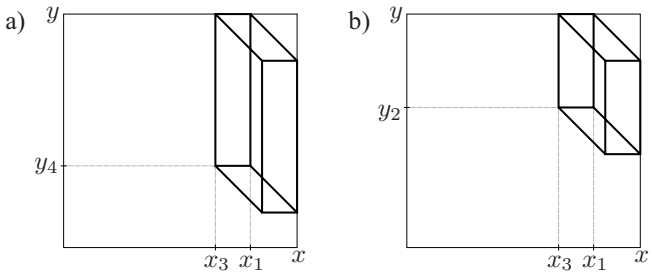


Figure 9: Illustration for the proof.

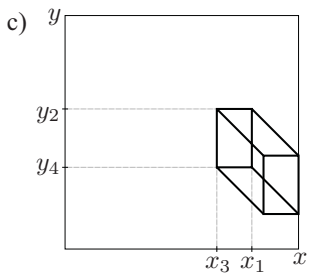


Figure 10: Illustration for the proof.

satisfies $\mathbf{R_1}$. In order to show that $\circ$ is associative if and only if it satisfies $\mathbf{R}$, we need to show that $\mathbf{R_1} \Leftrightarrow \mathbf{R}$. Obviously, $\mathbf{R} \Rightarrow \mathbf{R_1}$. The inverse implication, $\mathbf{R_1} \Rightarrow \mathbf{R}$, is given the following way.

Let us have a continuous Archimedean conjunctor K which satisfies $\mathbf{R_1}$. Let us have a Reidemeister figure drawn on its 3-web for arbitrary x_1, x_2, x_3, x_4, $y_1, y_2, y_3, y_4 \in [0, 1]$, see Figure 6. We are going to show that this figure shall be always closed. Thanks to $\mathbf{R_1}$, the Reidemeister figures, shown in Figure 7-a and Figure 7-b, are closed. Combining these two figures together it can be concluded that the Reidemeister figure in Figure 8-c is closed as well. By the same deduction, from the closedness of the Reidemeister figures in Figure 9-a and Figure 9-b it can be concluded that the Reidemeister figure in Figure 10-c is closed. Now, combining the closed Reidemeister figures in Figure 8-c and Figure 10-c, the closedness of the Reidemeister figure in Figure 6 is proven.

Corollary 4.1 *A continuous Archimedean conjunctor is associative (and, thus, a t-norm) if and only if it satisfies* $\mathbf{R}$.

5 Applications

5.1 Problem of convex combinations of t-norms

The question was introduced by Alsina, Frank, and Schweizer [4]. The approach of web geometry allows to give the following, recently published [11], answers:

Theorem 5.1 *Let T_1 and T_2 be two continuous Archimedean t-norms such that $\operatorname{Man} T_1 \neq \operatorname{Man} T_2$. Then no non-trivial convex combination F of T_1 and T_2 is a t-norm.*

Corollary 5.2 *Combining the result of Theorem 5.1 with the result given by Jenei [7], a non-trivial convex combination of two distinct nilpotent t-norms is never a t-norm.*

Corollary 5.3 *A non-trivial convex combination of a strict and a nilpotent t-norm is never a t-norm. This is an alternative proof of the result given earlier by Ouyang, Fang, and Li [10].*

Acknowledgment

The first author was supported by the Grant Agency of the Czech Republic under Project 401/09/H007.

References

[1] J. Aczél. Quasigroups, nets and nomograms. *Advances in Mathematics*, 1:383–450, 1965.

[2] M. A. Akivis and V. V. Goldberg. Algebraic aspects of web geometry. *Commentationes Mathematicae Universitatis Carolinae*, 41(2):205–236, 2000.

[3] M. A. Akivis and V. V. Goldberg. Local algebras of a differential quasigroup. *Bulletin of the American Mathematical Society*, 43(2):207–226, 2006.

[4] C. Alsina, M. J. Frank, and B. Schweizer. Problems on associative functions. *Aequationes Mathematicae*, 66(1–2):128–140, 2003.

[5] C. Alsina, M. J. Frank, and B. Schweizer. *Associative Functions: Triangular Norms and Copulas*. World Scientific, Singapore, 2006.

[6] W. Blaschke and G. Bol. *Geometrie der Gewebe, topologische Fragen der Differentialgeometrie*. Springer, Berlin, 1939.

[7] S. Jenei. On the convex combination of left-continuous t-norms. *Aequationes Mathematicae*, 72(1–2):47–59, 2006.

ISBN: 978-989-95079-6-8

[8] E. P. Klement, R. Mesiar, and E. Pap. *Triangular Norms*, vol. 8 of *Trends in Logic*. Kluwer Academic Publishers, Dordrecht, Netherlands, 2000.

[9] F.Montagna, C. Noguera, and R. Horčík. On Weakly Cancellative Fuzzy Logics. *Journal of Logic and Computation*, 16(4):423–450, August 2006.

[10] Y. Ouyang and J. Fang. Some observations about the convex combinations of continuous triangular norms. *Nonlinear Analysis*, 2007.

[11] M. Petrík, P. Sarkoci. Convex combinations of nilpotent triangular norms. *Journal of Mathematical Analysis and Applications*, 350(1):271–275, February 2009. DOI: 10.1016/j.jmaa.2008.09.060.

[12] B. Schweizer, A. Sklar. *Probabilistic Metric Spaces*. North-Holland, Amsterdam 1983; 2nd edition: Dover Publications, Mineola, NY, 2006.

ISBN: 978-989-95079-6-8

Fuzzy Arithmetic with Parametric LR Fuzzy Numbers

Luciano Stefanini, Laerte Sorini

DEMQ (Department of Economics and Quantitative Methods)
University of Urbino 'Carlo Bo', Italy.
Email: lucste@uniurb.it (L. Stefanini), laerte@uniurb.it (L. Sorini)

Abstract— *In this paper we suggest and describe a new family of parametric representations for LR fuzzy numbers and use them for fuzzy calculations and arithmetic.*

Keywords— LR Fuzzy Numbers, LU-Fuzzy Parametrization, Fuzzy Arithmetic.

1 Introduction

The arithmetical structure of fuzzy numbers have been developed in the 1980's and Dubois and Prade introduced the well known LR model and the corresponding formulas for the fuzzy operations (see the recent publication [1] and the references therein).

In general, the arithmetic operations on fuzzy numbers can be approached either by the direct use of the membership function (by the Zadeh extension principle) or by the equivalent use of the $\alpha - cuts$ representation. A solid result in fuzzy theory and practice is that calculations cannot be performed by using the same rules as in arithmetic with real numbers and in fact fuzzy calculus will not always satisfy the same properties (e.g. distributivity, invertibility and others).

A parametric representation of LR fuzzy numbers and the derivation of the corresponding arithmetic operator have been first introduced by Giachetti and Young (see [2], [3]). In this paper, we will see that it is possible to define a more flexible parametric representation of fuzzy numbers that allow a large variety of possible shapes and is very simple to implement (see [5] and [6] for a parametric $\alpha - cuts$ representation).

We suggest a parametrization for the LR fuzzy numbers, similar to the work done for the LU parametrization in [5] and [6]. In terms of the parameters representing the LR fuzzy numbers, it is possible to define the operators for the fuzzy arithmetic and we show that the errors of the approximations can be reduced to any small tolerance (by increasing the number of parameters); in fact, within the space of differentiable fuzzy numbers, the approximations form a dense subspace.

2 Fuzzy Numbers in LR Parametric Form

We will consider fuzzy quantities, i.e. fuzzy sets defined over the field $\mathbb{R}$ of real numbers and we will focus on fuzzy numbers, having a particular form of the membership function.

A general *fuzzy set* over a given set (or space) $\mathbb{X}$ of elements (the universe) is usually defined by its membership function $\mu : \mathbb{X} \longrightarrow \mathbb{T} \subseteq [0,1]$ and a fuzzy (sub)set u of $\mathbb{X}$ is uniquely characterized by the pairs $(x, \mu_u(x))$ for each $x \in \mathbb{X}$; the value $\mu_u(x) \in [0,1]$ is the membership grade of x to the fuzzy set u. So, a fuzzy set is given by

$$\begin{aligned} u &= \{(x, \mu_u(x)) | x \in \mathbb{X}\} \\ \text{given } \mu_u &: \mathbb{X} \longrightarrow [0,1] \end{aligned}$$

Fundamental concepts in fuzzy theory are the *support*, the *level-sets* (or *level-cuts*) and the *core* of a fuzzy set (or of its membership function): let μ_u be the membership function of a fuzzy set u over $\mathbb{X}$. The support of u is the (crisp) subset of points of $\mathbb{X}$ at which the membership grade $\mu_u(x)$ is positive:

$$supp(u) = \{x | x \in \mathbb{X}, \mu_u(x) > 0\}; \tag{1}$$

we always assume that $supp(u) \neq \emptyset$. For $\alpha \in]0,1]$, the $\alpha-$level cut of u (or simply the $\alpha - cut$) is defined by

$$[u]_\alpha = \{x | x \in \mathbb{X},\ \mu_u(x) \geq \alpha\} \tag{2}$$

and for $\alpha = 0$ (or $\alpha \to +0$) by the closure of the support

$$[u]_0 = cl\{x | x \in \mathbb{X},\ \mu_u(x) > 0\}.$$

The core of u is the set of elements of $\mathbb{X}$ having membership grade 1

$$core(u) = \{x | x \in \mathbb{X}, \mu_u(x) = 1\} \tag{3}$$

and we say that u is normal if $core(u) \neq \emptyset$.

Of our interest are fuzzy sets when the space $\mathbb{X}$ is $\mathbb{R}$ (unidimensional real fuzzy sets).

Well-known properties of the $level - cuts$ are:

$$[u]_\alpha \subseteq [u]_\beta \text{ for } \alpha > \beta, \tag{4}$$

$$[u]_\alpha = \bigcap_{\beta < \alpha} [u]_\beta \text{ for } \alpha \in]0,1] \tag{5}$$

and (if $x \in supp(u)$, otherwise $\mu_u(x) = 0$)

$$\mu_u(x) = \sup\{\alpha |\ \alpha \in]0,1] \text{ for which } x \in [u]_\alpha\}. \tag{6}$$

A particular class of fuzzy sets is when the support is a convex set (C is said *convex* if $(1-t)x' + tx'' \in C$ for every $x', x'' \in C$ and all $t \in [0,1]$) and the membership function is quasi-concave: we say that the membership function μ_u is quasi-concave if $\mu_u((1-t)x' + tx'') \geq \min\{\mu_u(x'), \mu_u(x'')\}$ for every $x', x'' \in supp(u)$ and $t \in [0,1]$. Equivalently, μ_u is quasi-concave if the level-cuts $[u]_\alpha$ are convex sets for all $\alpha \in [0,1]$.

A third property of the fuzzy quantities is related to the semi-continuity of the membership function and to the closedness of the level-cuts: μ_u is said to be upper semicontinuous if $\lim_{x \longrightarrow \widehat{x}} \sup \mu_u(x) = \mu_u(\widehat{x})$ at every $\widehat{x} \in supp(u)$ or, equivalently, if the level-cuts $[u]_\alpha$ are closed sets for all $\alpha \in [0,1]$.

Definition (*fuzzy number):* u is a fuzzy number if

(i.) μ_u has nonempty bounded support and is normal,

(ii.) μ_u is quasi-concave,

(iii.) μ_u is upper semicontinuous

ISBN: 978-989-95079-6-8

or, equivalently:

(a.) $[u]_\alpha$ are nonempty convex sets for all $\alpha \in [0,1]$

(b.) $[u]_\alpha$ are closed and bounded (compact) sets for all $\alpha \in [0,1]$

(c.) $[u]_\alpha$ satisfy conditions (4) and (5).

We will denote by $\mathcal{F}$ the set of fuzzy numbers. It can be structured by an addition and a scalar multiplication, defined either by the level sets or, equivalently, by the Zadeh extension principle. Let $u, v \in \mathcal{F}$ have membership functions μ_u, μ_v and $\alpha - cuts$ $[u]_\alpha$, $[v]_\alpha$, $\alpha \in [0,1]$ respectively. The addition $u+v \in \mathcal{F}$ and the scalar multiplication $ku \in \mathcal{F}$ for $k \in \mathbb{R}\backslash\{0\}$ have membership functions (extension principle)

$$\mu_{u+v}(z) = \sup\{\min\{\mu_u(x), \mu_v(y)\} | z = x+y\} \quad (7)$$

$$\mu_{ku}(x) = \mu_u(\frac{x}{k}) \quad (8)$$

and level cuts

$$[u+v]_\alpha = [u]_\alpha + [v]_\alpha \quad (9)$$
$$= \{x+y | x \in [u]_\alpha,\ y \in [v]_\alpha\}$$
$$[ku]_\alpha = k[u]_\alpha = \{kx | x \in [u]_\alpha\}. \quad (10)$$

We denote by $[u_\alpha^-, u_\alpha^+], \forall \alpha \in [0,1]$ the level-cuts of $u \in \mathcal{F}$ and by $[\widehat{u}^-, \widehat{u}^+]$ its core (level set at $\alpha = 1$). If $u_\alpha^- = \widehat{u}^-$ and $u_\alpha^+ = \widehat{u}^+, \forall \alpha \in [0,1]$ we have a crisp interval or a crisp number. We say that u is positive if $u_\alpha^- > 0$, $\forall \alpha \in [0,1]$ and that u is negative if $u_\alpha^+ < 0, \forall \alpha \in [0,1]$.

Definition (*LR-fuzzy number*): An LR-fuzzy number (or interval) u has membership function of the form

$$\mu_u(x) = \begin{cases} L(\frac{x-a}{b-a}) & \text{if } x \in [a,b] \\ 1 & \text{if } x \in [b,c] \\ R(\frac{d-x}{d-c}) & \text{if } x \in [c,d] \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

where $L, R : [0,1] \longrightarrow [0,1]$ are two non-decreasing *shape functions* such that $R(0) = L(0) = 0$ and $R(1) = L(1) = 1$. If $b = c$ we obtain a fuzzy number. The support is the compact interval $[a,d]$ and the core is $[b,c]$.

If L and R are invertible functions, then the $\alpha - cuts$ are obtained by

$$[u]_\alpha = [u_\alpha^-, u_\alpha^+] \quad (12)$$
$$\text{with}$$
$$u_\alpha^- = a + (b-a)L^{-1}(\alpha)$$
$$u_\alpha^+ = d - (d-c)R^{-1}(\alpha)$$

The usual LR-fuzzy notation is $u = \langle a,b,c,d\rangle_{L,R}$ for an interval and $u = \langle a,b,c\rangle_{L,R}$ for a number. We refer to functions $L(.)$ and $R(.)$ as the left and right branches (shape functions) of u, respectively and to the functions $u_{(.)}^-$ and $u_{(.)}^+$ as the lower and upper branches on u, respectively. Note that, in the representation (12), we have $u_0^- = a$, $u_1^- = b$, $u_1^+ = c$ and $u_0^+ = d$.

2.1 Parametric LR-fuzzy numbers

In this section we present the basic elements of a parametric representation of the shape functions based on monotonic Hermite-type interpolation. We first introduce some models for differentiable monotonic shape functions $p : [0,1] \longrightarrow [0,1]$ such that $p(0) = 0$ and $p(1) = 1$ with $p(t)$ differentiable and increasing on $[0,1]$; for any given nonnegative parameters $\beta_i \geq 0$, $i = 0,1$ we consider functions $p : [0,1] \longrightarrow [0,1]$ satisfying the four Hermite interpolation conditions

$$p(0) = 0,\ p(1) = 1$$
$$p'(0) = \beta_0,\ p'(1) = \beta_1.$$

We obtain infinite many functions simply by fixing the two parameters β_i that give the slopes (first derivatives) of the function at $t = 0$ and $t = 1$. To explicit the slope parameters we denote the interpolating function p by

$$t \longrightarrow p(t; \beta_0, \beta_1) \text{ for } t \in [0,1].$$

We recall here two of the basic forms from [5]:

∘ (2,2)-rational spline:

$$p(t; \beta_0, \beta_1) = \frac{t^2 + \beta_0 t(1-t)}{1 + (\beta_0 + \beta_1 - 2)t(1-t)}; \quad (13)$$

∘ mixed exponential spline:

$$p(t; \beta_0, \beta_1) = \frac{1}{m}[t^2(3-2t) - \beta_0(1-t)^m + \beta_0 + \beta_1 t^m] \quad (14)$$
$$\text{where } m = 1 + \beta_0 + \beta_1.$$

Both functions p in (13) and (14) are strictly increasing on [0,1] and we obtain a linear $p(t) = t, \forall t \in [0,1]$ if $\beta_0 = \beta_1 = 1$ and a quadratic $p(t) = t^2 + \beta_0 t(1-t)$ if $\beta_0 + \beta_1 = 2$ with $\beta_0 \neq 1$.

The (parametric) monotonic functions like (13) and (14) can be used as models for the shape 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, we obtain an LR-fuzzy number with the membership function

$$\mu_u(x) = \begin{cases} p(\frac{x-a}{b-a}; \beta_{0,L}, \beta_{1,L}) & \text{if } x \in [a,b] \\ 1 & \text{if } x \in [b,c] \\ p(\frac{x-d}{c-d}; \beta_{0,R}, \beta_{1,R}) & \text{if } x \in [c,d] \\ 0 & \text{otherwise} \end{cases}. \quad (15)$$

To make a uniform notation,we denote a, b, c, d as $a = u_{0,L}$, $b = u_{1,L}$, $c = u_{1,R}$, $d = u_{0,R}$ and $\beta_{0,L}^* = \frac{\beta_{0,i}}{b-a}$, $\beta_{1,L}^* = \frac{\beta_{1,i}}{b-a}$, $\beta_{0,R}^* = \frac{\beta_{0,j}}{c-d}$, $\beta_{1,R}^* = \frac{\beta_{1,j}}{c-d}$; so, the eight parameters completely define u:

$$u = (u_{0,L}, \beta_{0,L}^*, u_{0,R}, \beta_{0,R}^*; u_{1,L}, \beta_{1,L}^*, u_{1,R}, \beta_{1,R}^*) \quad (16)$$

provided that $u_{0,L} \leq u_{1,L} \leq u_{1,R} \leq u_{0,R}$ and $\beta_{0,L}, \beta_{1,L} \geq 0$, $\beta_{0,R}, \beta_{1,R} \geq 0$.

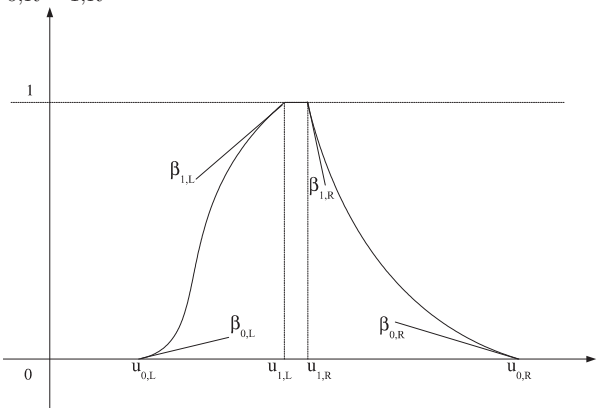


Figure 1: Typical LR parametric membership function

ISBN: 978-989-95079-6-8

Remark: As the nonnegative parameters $\beta_{i,L}$ and $\beta_{i,R}$ ($i = 1, 2$) are completely independent, each side of the membership function may have independent (monotonic) forms, e.g. concavity, convexity, inflextions etc. of left and right sides are independent, as illustrated in the figure. The shapes are simply "controlled" by the 4 parameters $\beta_{i,L}$ and $\beta_{i,R}$ ($i = 1, 2$).

Denote by $\mathbb{F}^{LR}$ the set of LR-fuzzy numbers defined by (16). The family of fuzzy numbers $\mathbb{F}^{LR}$, which include triangular and trapezoidal fuzzy numbers if all $\beta_{i,L} = \beta_{i,R} = 1$, are characterized by eight parameters and it appears that the inclusion of the slopes of left and right functions, even without generating piecewise monotonic approximations over subintervals is able to capture much more information than the linear approximation (see [5] and [6]).

The level sets included into the representation are the core and the support. Note that if we are interested to represent exactly the core, the support and the midpoint interval we require $N = 2$, $\alpha_0 = 0$, $\alpha_1 = 0.5$, $\alpha_2 = 1$, the support $[u_{0,L}, u_{0,R}]$ (corresponding to $\alpha = \alpha_0$), the midpoint interval $[u_{1,L}, u_{1,R}]$ (corresponding to $\alpha = \alpha_1$), the core $[u_{2,L}, u_{2,R}]$ (corresponding to $\alpha = \alpha_0$) and the six parameters $\beta_{i,L}, \beta_{i,R} \geq 0$ to model the first derivatives of $\mu_u(x)$ at the extremal point of the three intervals above.

More generally, we can design differentiable parametric shape functions $L()$ and $R()$ by fixing $N + 1$ distinct level sets $0 = \alpha_0 < \alpha_1 < ... < \alpha_N = 1$ and assigning, for each α_i ($i = 0, 1, ..., N$), the four parameters needed for each level.

In this general case, we need the values $u_{i,L}$, $u_{i,R}$ but, instead of the parameters $\beta_{i,L}, \beta_{i,R}$, it is convenient to give directly the first derivatives of $\mu_u(x)$ at $u_{i,L}$, $u_{i,R}$ and calculate the $\beta_{i,L}, \beta_{i,R} \geq 0$ to use in the monotonic functions like (13) and (14) and model the membership function like (15) piecewise on each subinterval $[u_{i-1,L}, u_{i,L}]$ and $[u_{i,R}, u_{i-1,R}]$.

For $i = 0, 1, ..., N$, denote directly by $\delta u_{i,L} \geq 0$ and by $\delta u_{i,R} \leq 0$ the first derivatives of $\mu_u(x)$ at the points $x = u_{i,L}$ and $x = u_{i,R}$, respectively and suppose they are given. The membership function $\mu_u(x)$ is obtained by the following simple procedure:

Assign first $\mu_u(x) = 0$ for all x;
for all $x \in [u_{N,L}, u_{N,R}]$ set $\mu_u(x) = 1$;
for $i = 1, 2, ..., N$
  for $x \in [u_{i-1,L}, u_{i,L}]$ set
    $\beta_0 = \frac{u_{ni} - u_{n-1,i}}{\alpha_n - \alpha_{n-1}} \delta u_{i-1,L}$
    $\beta_1 = \frac{u_{ni} - u_{n-1,i}}{\alpha_n - \alpha_{n-1}} \delta u_{i,L}$
    $\mu_u(x) = \alpha_{i-1} + (\alpha_i - \alpha_{i-1}) p(\frac{x - u_{n-1,i}}{u_{ni} - u_{n-1,i}}; \beta_0, \beta_1)$
  end
  for $x \in [u_{i,R}, u_{i-1,R}]$ set
    $\beta_0 = \frac{u_{nj} - u_{n-1,j}}{\alpha_n - \alpha_{n-1}} \delta u_{i-1,R}$
    $\beta_1 = \frac{u_{nj} - u_{n-1,j}}{\alpha_n - \alpha_{n-1}} \delta u_{i,R}$
    $\mu_u(x) = \alpha_{i-1} + (\alpha_i - \alpha_{i-1}) p(\frac{x - u_{n-1,j}}{u_{nj} - u_{n-1,j}}; \beta_0, \beta_1)$
  end
end

Denote by $\mathbb{F}_N^{LR}$ the fuzzy numbers obtained in the form above; as each shape function is monotonic, the left and right branches are monotonic increasing on $[u_{0,L}, u_{N,L}]$ (the left of μ_u) and decreasing on $[u_{N,R}, u_{0,R}]$ (the right of μ_u). The number of parameters is $4N + 4$ and they are simply constrained to be $u_{0,L} \leq ... \leq u_{N,L} \leq u_{N,R} \leq ... \leq u_{0,R}$ and $\delta u_{i,L} \geq 0$ and $\delta u_{i,R} \leq 0$ to ensure monotonicity. In general, we will have $u_{i-1,L} < u_{i,L}$ and $u_{i-1,R} > u_{i,R}$ but it is easy to consider case of equality so that the graph of μ_u is a vertical line (discontinuity) at $u_{i,.}$ if $u_{i-1,.} = u_{i,.}$.

In this general form, an LR fuzzy number is represented as follows:

$$u = (\alpha_i; u_{i,L}, \delta u_{i,L}, u_{i,R}, \delta u_{i,R})_{i=0,1,...,N}. \qquad (17)$$

Example:

Consider a quasi-Gaussian membership function ($m \in \mathbb{R}$, $k, \sigma \in \mathbb{R}^+$; if $k \to +\infty$ the support is unbounded)

$$\mu_{qG}(x) = \begin{cases} \exp(-\frac{(x-m)^2}{2\sigma^2}) & \text{if } m - k\sigma \leq x \leq m + k\sigma \\ 0 & \text{otherwise.} \end{cases} \qquad (18)$$

Its LR parametrization for $m = 0$, $\sigma = 2$, $k = 4$, approximated with $N = 4$ (five points), is

Table 1. LR parametrization of fuzzy number (18)

α_i	$u_{i,L}$	$\delta u_{i,L}$	$u_{i,R}$	$\delta u_{i,R}$
0.0	-8.0	0.00033	8.0	-0.00033
0.25	-3.33022	0.20814	3.33022	-0.20814
0.5	-2.35482	0.29435	2.35482	-0.29435
0.75	-1.51705	0.28445	1.51705	-0.28445
1.0	0.0	0.0	0.0	0.0

It is interesting to note that, in our parametric representation (17) of LR-fuzzy numbers, the membership function has differentiable left and right branches; if we require that $\mu_u(x)$ be differentiable in all the (internal) points of the support of u, in particular also at $x = u_{N,L}$ and $x = u_{N,R}$, we simply need to set $\delta u_{N,L} = 0$ and $\delta u_{N,R} = 0$ in the parametric representation (17). This seems to be an important advantage in all the applications (e.g. fuzzy intelligent systems, neuro-fuzzy learning models) where overall differentiability of μ_u is required.

The generated LR fuzzy numbers form a subspace of the space of fuzzy numbers. In the case of differentiable membership functions, it is immediate to understand that the union of all parametric fuzzy numbers having the form (17) for all integer $N \geq 1$, i.e.

$$\mathcal{F}^{LR} = \bigcup_{N \geq 1} \mathbb{F}_N^{LR},$$

is dense into the space of (differentiable) fuzzy numbers. In fact, each fuzzy number (17) has the property of interpolate (exactly) the values $u_{i,L}, u_{i,R}$ ($i = 0, 1, ..., N$) and it is sufficient to refine the points α_i to obtain any desired precision. In our experience, approximations with values of N from 5 to 20 have in general a very small error, of the order of $10^{-6} - 10^{-2}$.

Following [4], for $p \in [0, \infty]$, we can define a geometric distance $D_p(u, v)$ between fuzzy numbers $u, v \in \mathbb{F}_N^{LR}$, given by

$$D_p^{LR}(u, v) = \left(\sum_{i=0}^{N} |u_{i,L} - v_{i,L}|^p + |u_{i,R} - v_{i,R}|^p \right)^{1/p} + \\ + \left(\sum_{i=0}^{N} |\delta u_{i,L} - \delta v_{i,L}|^p + |\delta u_{i,R} - \delta v_{i,R}|^p \right)^{1/p}$$

and

ISBN: 978-989-95079-6-8

$$D_\infty^{LR}(u,v) = \max_{i=0,\dots,N}\{|u_{i,L}-v_{i,L}|, |u_{i,R}-v_{i,R}|\} + \\ + \max_{i=0,\dots,N}\{|\delta u_{i,L}-\delta v_{i,L}|, |\delta u_{i,R}-\delta v_{i,R}|\}.$$

If we model the LR-fuzzy numbers by a (2,2)-rational spline $p(\alpha;\beta_0,\beta_1)$ like (13) the inverse $p^{-1}(t;\beta_0,\beta_1)$ can be computed analytically as we have to solve the quadratic equation (with respect to α)

$$\alpha^2+\beta_0\alpha(1-\alpha) = t[1+(\beta_0+\beta_1-2)\alpha(1-\alpha)] \text{ i.e.}$$

$$\begin{aligned}(1+A(t))\alpha^2 - A(t)\alpha - t &= 0 \text{ where}\\ A(t) &= -\beta_0+\beta_0 t+\beta_1 t-2t.\end{aligned}$$

If $A(t)=-1$ then the equation is linear and the solution is $\alpha=t$. If $A(t)\neq -1$, then there exist two real solutions $\alpha_1 = \frac{2\sqrt{t}}{2+2A(t)}$, $\alpha_2 = \frac{2\sqrt{t}+2A(t)}{2+2A(t)}$ and we choose the one belonging to $[0,1]$.

3 Fuzzy arithmetic with LR parametrization

The fuzzy extension principle introduced by Zadeh is the basic tool for fuzzy calculus; it extends functions of real numbers to functions of fuzzy numbers and it allows the extension of arithmetic operations and calculus to fuzzy arguments. We have already defined the addition (7) and the scalar multiplication (8).

3.1 Basic arithmetic operators

Let $\circ \in \{+,-,\times,/\}$ one of the four arithmetic operations and let $u,v\in\mathbb{F}_{\mathbb{I}}$ be given fuzzy intervals (or numbers), having $\mu_u(.)$ and $\mu_v(.)$ as membership functions and level-cuts representations $u=(u^-,u^+)$, $v=(v^-,v^+)$; the extension principle for the extension of $\circ$ defines the membership function of $w=u\circ v$ by

$$\mu_{u\circ v}(z) = \sup\{\min\{\mu_u(x),\mu_v(y)\}|z=x\circ y\}. \qquad (19)$$

In terms of the $\alpha-cuts$, the four arithmetic operations and the scalar multiplication for $k\in\mathbb{R}$ are obtained by the well-known *interval arithmetic* (for all $\alpha\in[0,1]$)

Addition ($w=u+v$):
$[u+v]_\alpha = [u_\alpha^- + v_\alpha^-, u_\alpha^+ + v_\alpha^+]$,

Scalar multiplication ($w=ku$):
$[ku]_\alpha = [\min\{ku_\alpha^-, ku_\alpha^+\}, \max\{ku_\alpha^-, ku_\alpha^+\}]$,

Subtraction ($w=u-v$):
$[u-v]_\alpha = [u_\alpha^- - v_\alpha^+, u_\alpha^+ - v_\alpha^-]$,

Multiplication ($w=uv$):
$$\begin{cases}(uv)_\alpha^- = \min\{u_\alpha^-v_\alpha^-, u_\alpha^-v_\alpha^+, u_\alpha^+v_\alpha^-, u_\alpha^+v_\alpha^+\}\\ (uv)_\alpha^+ = \max\{u_\alpha^-v_\alpha^-, u_\alpha^-v_\alpha^+, u_\alpha^+v_\alpha^-, u_\alpha^+v_\alpha^+\}\end{cases},$$

Division ($w=u/v$): if $0\notin[v_0^-, v_0^+]$
$$\begin{cases}(u/v)_\alpha^- = \min\left\{\frac{u_I^-}{v_I^-}, \frac{u_I^-}{v_I^+}, \frac{u_I^+}{v_I^-}, \frac{u_I^+}{v_I^+}\right\}\\ (u/v)_\alpha^+ = \max\left\{\frac{u_I^-}{v_I^-}, \frac{u_I^-}{v_I^+}, \frac{u_I^+}{v_I^-}, \frac{u_I^+}{v_I^+}\right\}\end{cases}.$$

In recent work (see [7]) the following generalized subtraction has been defined and analyzed:

Definition: Let u,v be two fuzzy numbers; we define the generalized Hukuhara difference (gH-difference for short) of u and v as the fuzzy number w such that

$$u\ominus_g v = w \iff \begin{cases}(i) & u=v+w\\ \text{or} & (ii)\quad v=u-w\end{cases}. \qquad (20)$$

The α-cuts of both $u\ominus_g v$ can be expressed in terms of the α-cuts of u and v; but it is possible that $u\ominus_g v$ is not well defined; by the use of the parametrization (17) it is easy to test if the operation is well defined or not.

Consider two LR-fuzzy numbers u and v ($N=1$ for simplicity)

$$\begin{aligned}u &= (u_{0,L},\delta u_{0,L},u_{0,R},\delta u_{0,R}; \qquad (21)\\ &\quad u_{1,L},\delta u_{1,L},u_{1,R},\delta u_{1,R}),\\ v &= (v_{0,L},\delta v_{0,L},v_{0,R},\delta v_{0,R}; \qquad (22)\\ &\quad v_{1,L},\delta v_{1,L},v_{1,R},\delta v_{1,R}).\end{aligned}$$

Note that in the formulas below u and v are not restricted to have the same $L(.)$ and $R(.)$ shape functions and changing the slopes will change the form of the membership functions (in all the cases below, eventually assume $\frac{0}{0}=0$).

The values and the slopes at the considered α-cuts are computed exactly; in fact we use the exact standard formulae for the derivative of addition, multiplication, division and function composition. Corresponding to membership values $\alpha\neq\alpha_i$, $i=0,1,...,N$, the membership functions is approximated by using the rational (13) or the mixed (14) shape functions.

The approximate addition is the following

$$(u+v) = \left(\begin{array}{l} u_{0,L}+v_{0,L}, \frac{\delta u_{0,i}\,\delta v_{0,i}}{\delta u_{0,i}+\delta v_{0,i}},\\ u_{0,R}+v_{0,R}, \frac{\delta u_{0,j}\,\delta v_{0,j}}{\delta u_{0,j}+\delta v_{0,j}};\\ u_{1,L}+v_{1,L}, \frac{\delta u_{1,i}\,\delta v_{1,i}}{\delta u_{1,i}+\delta v_{1,i}},\\ u_{1,R}+v_{1,R}, \frac{\delta u_{1,j}\,\delta v_{1,j}}{\delta u_{1,j}+\delta v_{1,j}}\end{array}\right).$$

Nothe that if the left and right shapes of u and v are the same (e.g. linear, quadratic) then the addition is exact.

The general algorithm for approximate LR addition is
Algorithm (*LR addition*) $w=u+v$
for $i=0,1,...,N$
$w_{i,L}=u_{i,L}+v_{i,L}$, $w_{i,R}=u_{i,R}+v_{i,R}$
$\delta w_{i,L}=\frac{\delta u_{n_i}\,\delta v_{n_i}}{\delta u_{n_i}+\delta v_{n_i}}$, $\delta w_{i,R}=\frac{\delta u_{n_j}\,\delta v_{n_j}}{\delta u_{n_j}+\delta v_{n_j}}$
end

The difference $w=u-v$ is similar
Algorithm (*LR subtraction*) $w=u-v$
for $i=0,1,...,N$
$w_{i,L}=u_{i,L}-v_{i,R}$, $w_{i,R}=u_{i,R}-v_{i,L}$
$\delta w_{i,L}=\frac{\delta u_{n_i}\,\delta v_{n_j}}{\delta v_{n_j}-\delta u_{n_j}}$, $\delta w_{i,R}=\frac{\delta u_{n_j}\,\delta v_{n_i}}{\delta v_{n_i}-\delta u_{n_j}}$
end

Also the approximate multiplication $w=uv$ can be obtained easily; in the case of two positive LR-fuzzy numbers w is given by

$$w_{LR} = \left(\begin{array}{l} u_{0,L}v_{0,L}, \frac{\delta u_{0,i}\,\delta v_{0,i}}{v_{0,i}\,\delta v_{0,i}+u_{0,i}\,\delta u_{0,i}},\\ u_{0,R}v_{0,R}, \frac{\delta u_{0,j}\,\delta v_{0,j}}{v_{0,j}\,\delta v_{0,j}+u_{0,j}\,\delta u_{0,j}};\\ u_{1,L}v_{1,L}, \frac{\delta u_{1,i}\,\delta v_{1,i}}{v_{1,i}\,\delta v_{1,i}+u_{1,i}\,\delta u_{1,i}},\\ u_{1,R}v_{1,R}, \frac{\delta u_{1,j}\,\delta v_{1,j}}{v_{1,j}\,\delta v_{1,j}+u_{1,j}\,\delta u_{1,j}}\end{array}\right).$$

ISBN: 978-989-95079-6-8

The general algorithm for approximate LR multiplication is

Algorithm (*LR multiplication*) $w = uv$
(eventually, use $\frac{0}{0} = 0$)
for $i = 0, 1, ..., N$
$m_i = \min\{u_{i,L}v_{i,L}, u_{i,L}v_{i,R}, u_{i,R}v_{i,L}, u_{i,R}v_{i,R}\}$
$M_i = \max\{u_{i,L}v_{i,L}, u_{i,L}v_{i,R}, u_{i,R}v_{i,L}, u_{i,R}v_{i,R}\}$
$w_{i,L} = m_i \ , \ w_{i,R} = M_i$
if $u_{i,L}v_{i,L} = m_i$ **then** $\delta w_{i,L} = \frac{\delta u_{n_i}\, \delta v_{n_i}}{v_{n_i}\, \delta v_{n_i} + u_{n_i}\, \delta u_{n_i}}$
elseif $u_{i,L}v_{i,R} = m_i$ **then** $\delta w_{i,L} = \frac{\delta u_{n_i}\, \delta v_{n_j}}{v_{n_j}\, \delta v_{n_j} + u_{n_i}\, \delta u_{n_i}}$
elseif $u_{i,R}v_{i,L} = m_i$ **then** $\delta w_{i,L} = \frac{\delta u_{n_j}\, \delta v_{n_i}}{v_{n_i}\, \delta v_{n_i} + u_{n_j}\, \delta u_{n_j}}$
elseif $u_{i,R}v_{i,R} = m_i$ **then** $\delta w_{i,L} = \frac{\delta u_{n_j}\, \delta v_{n_j}}{v_{n_j}\, \delta v_{n_j} + u_{n_j}\, \delta u_{n_j}}$
endif
if $u_{i,L}v_{i,L} = M_i$ **then** $\delta w_{i,R} = \frac{\delta u_{n_i}\, \delta v_{n_i}}{v_{n_i}\, \delta v_{n_i} + u_{n_i}\, \delta u_{n_i}}$
elseif $u_{i,L}v_{i,R} = M_i$ **then** $\delta w_{i,R} = \frac{\delta u_{n_i}\, \delta v_{n_j}}{v_{n_j}\, \delta v_{n_j} + u_{n_i}\, \delta u_{n_i}}$
elseif $u_{i,R}v_{i,L} = M_i$ **then** $\delta w_{i,R} = \frac{\delta u_{n_j}\, \delta v_{n_i}}{v_{n_i}\, \delta v_{n_i} + u_{n_j}\, \delta u_{n_j}}$
elseif $u_{i,R}v_{i,R} = M_i$ **then** $\delta w_{i,R} = \frac{\delta u_{n_j}\, \delta v_{n_j}}{v_{n_j}\, \delta v_{n_j} + u_{n_j}\, \delta u_{n_j}}$
endif
end

The division is similar:

Algorithm (*LR division*) $w = u/v$, $0 \notin [v]_0$
(eventually, use $\frac{0}{0} = 0$)
for $i = 0, 1, ..., N$
$m_i = \min\{u_{i,L}/v_{i,L}, u_{i,L}/v_{i,R}, u_{i,R}/v_{i,L}, u_{i,R}/v_{i,R}\}$
$M_i = \max\{u_{i,L}/v_{i,L}, u_{i,L}/v_{i,R}, u_{i,R}/v_{i,L}, u_{i,R}/v_{i,R}\}$
$w_{i,L} = m_i \ , \ w_{i,R} = M_i$
if $u_{i,L}/v_{i,L} = m_i$ **then** $\delta w_{i,L} = \frac{v_{n_i}^2\, \delta u_{n_i}\, \delta v_{n_i}}{v_{n_i}\, \delta v_{n_i} - u_{n_i}\, \delta u_{n_i}}$
elseif $u_{i,L}/v_{i,R} = m_i$ **then** $\delta w_{i,L} = \frac{v_{n_j}^2\, \delta u_{n_i}\, \delta v_{n_j}}{v_{n_j}\, \delta v_{n_j} - u_{n_i}\, \delta u_{n_i}}$
elseif $u_{i,R}/v_{i,L} = m_i$ **then** $\delta w_{i,L} = \frac{v_{n_i}^2\, \delta u_{n_j}\, \delta v_{n_i}}{v_{n_i}\, \delta v_{n_i} - u_{n_j}\, \delta u_{n_j}}$
elseif $u_{i,R}/v_{i,R} = m_i$ **then** $\delta w_{i,L} = \frac{v_{n_j}^2\, \delta u_{n_j}\, \delta v_{n_j}}{v_{n_j}\, \delta v_{n_j} - u_{n_j}\, \delta u_{n_j}}$
endif
if $u_{i,L}/v_{i,L} = M_i$ **then** $\delta w_{i,R} = \frac{v_{n_i}^2\, \delta u_{n_i}\, \delta v_{n_i}}{v_{n_i}\, \delta v_{n_i} - u_{n_i}\, \delta u_{n_i}}$
elseif $u_{i,L}/v_{i,R} = M_i$ **then** $\delta w_{i,R} = \frac{v_{n_j}^2\, \delta u_{n_i}\, \delta v_{n_j}}{v_{n_j}\, \delta v_{n_j} - u_{n_i}\, \delta u_{n_i}}$
elseif $u_{i,R}/v_{i,L} = M_i$ **then** $\delta w_{i,R} = \frac{v_{n_i}^2\, \delta u_{n_j}\, \delta v_{n_i}}{v_{n_i}\, \delta v_{n_i} - u_{n_j}\, \delta u_{n_j}}$
elseif $u_{i,R}/v_{i,R} = M_i$ **then** $\delta w_{i,R} = \frac{v_{n_j}^2\, \delta u_{n_j}\, \delta v_{n_j}}{v_{n_j}\, \delta v_{n_j} - u_{n_j}\, \delta u_{n_j}}$
endif
end

The algorithm for the gH-difference $w = u \ominus_g v$ in LR form is also easy, if we recall that the α-cuts of w are the intervals (see [7]) $[\min\{u_\alpha^- - v_\alpha^-, u_\alpha^+ - v_\alpha^+\}, \max\{u_\alpha^- - v_\alpha^-, u_\alpha^+ - v_\alpha^+\}]$:

Algorithm (*LR gH-subtraction*) $w = u \ominus_g v$
(eventually, use $\frac{0}{0} = 0$)
for $i = 0, 1, ..., N$
define $d_{i,L} = u_{i,L} - v_{i,L}$ and $d_{i,R} = u_{i,R} - v_{i,R}$
if $d_{i,L} < d_{i,R}$ **then**
$w_{i,L} = d_{i,L}$, $w_{i,R} = d_{i,R}$
$\delta w_{i,L} = \frac{\delta u_{n_i}\, \delta v_{n_i}}{\delta v_{n_i} - \delta u_{n_i}}$, $\delta w_{i,R} = \frac{\delta u_{n_j}\, \delta v_{n_j}}{\delta v_{n_j} - \delta u_{n_j}}$
elseif $d_{i,L} > d_{i,R}$ **then**
$w_{i,R} = d_{i,L}$, $w_{i,L} = d_{i,R}$
$\delta w_{i,R} = \frac{\delta u_{n_i}\, \delta v_{n_i}}{\delta v_{n_i} - \delta u_{n_i}}$, $\delta w_{i,L} = \frac{\delta u_{n_j}\, \delta v_{n_j}}{\delta v_{n_j} - \delta u_{n_j}}$
else (here $d_{i,L} = d_{i,R} = d$)
$w_{i,L} = d$, $w_{i,R} = d$
$\delta w_{i,L} = 0$, $\delta w_{i,R} = 0$
endif
end
Test if the calculated $w_{i,L}, w_{i,R}$ are ordered correctly and $\delta w_{i,L} \geq 0, \delta w_{i,R} \leq 0$, then w is $u \ominus_g v$.

The operations above are exact at the nodes $\alpha = \alpha_i$, $i = 0, 1, ..., N$, and have very small global errors for all $\alpha \in [0, 1]$ (if N is sufficiently high, of the order of 5 to 20). Further, it is easy to control the error by using a sufficiently fine $\alpha-$decomposition and the results have shown that both the rational (13) and the mixed (14) models perform well.

Some parametric membership functions in the LR framework are present in many applications and the use of nonlinear shapes in increasing. Usually, one defines a given family, e.g. linear, quadratic, sigmoid, quasi gaussian, and the operations are performed within the same family.

Our proposed parametrization allows an extended set of flexible fuzzy numbers and is able to approximate all other forms with acceptable small errors and the additional advantage of producing good approximations to the results of the arithmetic operations even between LR fuzzy numbers having very different original shapes.

Furthermore, at the explicitly considered $\alpha-$cuts of the decomposition, all the values $w_{i,L}$, $w_{i,R}$ and the slopes $\delta w_{i,L}$, $\delta w_{i,R}$ $(i = 0, 1, ..., N)$ of the membership function are exact.

3.2 Computation of fuzzy-valued functions

For a continuous function $f : \mathbb{R}^n \to \mathbb{R}$, let $v = f(u_1, u_2, ..., u_n)$ denote its fuzzy extension, based on the application of the Zadeh's Extension Principle (EP for short). It is well known that the $\alpha - cuts$ $[v_\alpha^-, v_\alpha^+]$ of v are obtained by solving the following box-constrained global optimization problems ($\alpha \in [0, 1]$)

$$v_\alpha^- = \min\{f(x_1, ..., x_n) | x_k \in [u_k]_\alpha,\ k = 1, 2, ..., n\} \quad (23)$$

$$v_\alpha^+ = \max\{f(x_1, ..., x_n) | x_k \in [u_k]_\alpha,\ k = 1, 2, ..., n\} \quad (24)$$

where $[u_k]_\alpha = [u_{k,\alpha}^-, u_{k,\alpha}^+]$, $k = 1, 2, ..., n$, are the $\alpha - cuts$ of u_k.

The lower and upper values v_α^- and v_α^+ of v define equivalently (as f is assumed to be continuous) the image of the cartesian product $\underset{k=1}{\overset{n}{\times}} [u_k]_\alpha$ via f, i.e. $[v_\alpha^-, v_\alpha^+] = f([u_1]_\alpha, ..., [u_n]_\alpha)$.

Except for simple elementary cases for which the optimization problems above can be solved analytically, the direct application of (23) and (24) is difficult and computationally expensive (see [6]).

At least if f is differentiable, the advantages of the LR representation appear to be quite interesting, based on the fact that a small number of α points is in general sufficient to obtain good approximations (this is the essential gain in using the slopes to model fuzzy numbers), so reducing the number of constrained *min* (23) and *max* (24) problems to be solved directly. On the other hand, finding computationally efficient extension solvers is still an open research field in fuzzy calculations.

We now give the details of the fuzzy extension of general differentiable functions of only one variable, by the LR representation. The case of multidimensional differentiable functions can be approached in a similar way, by considering the

ISBN: 978-989-95079-6-8

partial derivatives of $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ and the chain rules for the composition of multidimensional functions.

In all the computations we will adopt the EP method, but also if other approaches are adopted, the representation still remains valid.

Let $f : \mathbb{R} \to \mathbb{R}$; its *(EP)*-extension $v = f(u)$ to a fuzzy argument $u = (u^-, u^+)$ has $\alpha - cuts$

$$[v]_\alpha = [\min\{f(x) \mid x \in [u]_\alpha\}, \max\{f(x) \mid x \in [u]_\alpha\}]. \tag{25}$$

If f is monotonic increasing we obtain $[v]_\alpha = [f(u_\alpha^-), f(u_\alpha^+)]$ while, if f is monotonic decreasing, $[v]_\alpha = [f(u_\alpha^+), f(u_\alpha^-)]$; for simplicity, in the monotonic case we assume that the derivative of f is not null over the support of u, but it is possible to design the algorithm also in the case where, for some α, $f'(u_\alpha^+) = 0$ or $f'(u_\alpha^-) = 0$.

The LR representation of $v = (\alpha_i;\ v_{i,L}, \delta v_{i,L}, v_{i,R}, \delta v_{i,R})_{i=0,1,...,N}$ is obtained by the following algorithm:

Algorithm: (1-dim monotonic extension)

Let $u = (\alpha_i; u_{i,L}, \delta u_{i,L}, u_{i,R}, \delta u_{i,R})_{i=0,1,...,N}$ be given and $f : [u_{0,L}, u_{0,R}] \to \mathbb{R}$ be differentiable and monotonic; calculate $v = f(u)$.

for $i = 0, 1, ..., N$
  if (f is increasing) **then**
    $v_{i,L} = f(u_{i,L}),\ \delta v_{i,L} = \frac{\delta u_{n,i}}{f'(u_{n,i})}$
    $v_{i,R} = f(u_{i,R}),\ \delta v_{i,R} = \frac{\delta u_{n,j}}{f'(u_{n,j})}$
  elseif (f is decreasing) **then**
    $v_{i,L} = f(u_{i,R}),\ \delta v_{i,L} = \frac{\delta u_{n,j}}{f'(u_{n,j})}$
    $v_{i,R} = f(u_{i,R}),\ \delta v_{i,R} = \frac{\delta u_{n,j}}{f'(u_{n,j})}$
  endif
end

As an example, the monotonic exponential function $f(x) = \exp(x)$ has LR-fuzzy extension

$$\exp(u) = (\alpha_i; \exp(u_{i,L}), \exp(u_i^-)\,\delta u_{i,L}/\exp(u_{i,L}), \exp(u_{i,R}), \delta u_{i,R}/\exp(u_{i,R}))_{i=0,1,...N}$$

In the non monotonic (differentiable) case, we have to solve the optimization problems in (25) for each $\alpha = \alpha_i$, $i = 0, 1, ..., N$, i.e.

$$(EP_i): \begin{cases} v_{i,L} = \min\{f(x) | x \in [u_{i,L}, u_{i,R}]\} \\ v_{i,R} = \max\{f(x) | x \in [u_{i,L}, u_{i,R}]\}. \end{cases}$$

The $\min$ (or the $\max$) can occur either at a point which is coincident with one of the extremal values of $[u_{i,L}, u_{i,R}]$ or at a point which is internal; in the last case, the derivative of f is null and $\delta v_{i,L} = +\infty$ (or $\delta v_{i,R} = -\infty$).

In the calculations, we "approximate" $\pm\infty$ by $\pm BIG$ where BIG is a big positive number.

Algorithm: (1-dim non monotonic LR extension)

Let $u = (\alpha_i; u_{i,L}, \delta u_{i,L}, u_{i,R}, \delta u_{i,R})_{i=0,1,...,N}$ be given and $f : [u_{0,L}, u_{0,R}] \to \mathbb{R}$ be differentiable and monotonic; calculate $v = f(u)$.

for $i = 0, 1, ..., N$
  solve $\min\{f(x) | x \in [u_{i,L}, u_{i,R}]\}$
  let $\widehat{x}_i = \arg\min\{f(x) | x \in [u_{i,L}, u_{i,R}]\}$
  if $\widehat{x}_i = u_{i,L}$ **then** $v_{i,L} = f(u_{i,L}),\ \delta v_{i,L} = \frac{\delta u_{n,i}}{f'(u_{n,i})}$
  elseif $\widehat{x}_i = u_{i,R}$ **then** $v_{i,L} = f(u_{i,R}), \delta v_{i,L} = \frac{\delta u_{n,j}}{f'(u_{n,j})}$
  else $v_{i,L} = f(\widehat{x}_i),\ \delta v_{i,L} = BIG$
  endif
  solve $\max\{f(x) | x \in [u_{i,L}, u_{i,R}]\}$
  let $\widehat{x}_i = \arg\max\{f(x) | x \in [u_{i,L}, u_{i,R}]\}$
  if $\widehat{x}_i = u_{i,L}$ **then** $v_{i.R} = f(u_{i,L}), \delta v_{i,R} = \frac{\delta u_{n,i}}{f'(u_{n,i})}$
  elseif $\widehat{x}_i = u_{i,R}$ **then** $v_{i,R} = f(u_{i,R}), \delta v_{i,R} = \frac{\delta u_{n,j}}{f'(u_{n,j})}$
  else $v_{i,R} = f(\widehat{x}_i),\ \delta v_{i,R} = -BIG$
  endif
end

4 Conclusion and further work

We suggest a parametrization of LR fuzzy numbers u by the use of parametric monotonic (simple) functions to model the left and the right branches of u. The obtained parametrizations form a subspace of the space of fuzzy numbers and can be possibly be refined to become a dense subspace. Within the parametrizations, we define arithmetic operators for the basic arithmetic and for the fuzzy extension of functions by the application of Zadeh's extension principle.

References

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

[2] R. E. Giachetti and R.E. Young: A Parametric Representation of Fuzzy Numbers and their Arithmetic Operators, *Fuzzy Sets and Systems,* 91, (1997) 185-202.

[3] R. E. Giachetti and R.E. Young: Analysis of the Error in the Standard Approximation used for Multiplication of Triangular and Trapezoidal fuzzy Numbers and the Development of a New Approximation, *Fuzzy Sets and Systems,* 91, (1997) 1-13.

[4] S. Heilpern, Representation and application of fuzzy numbers, *Fuzzy Sets and Systems*, 91, 1997, 259-278.

[5] 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

[6] 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.

[7] L. Stefanini, A Generalization of Hukuhara Difference, in D. Dubois et al. (Eds.), Soft Methods for Handling Variability and Imprecision, ASC 48, Springer Verlag, 2008, 203-210.

ISBN: 978-989-95079-6-8

Type-2 Fuzzy Arithmetic using Alpha-planes

Hussam Hamrawi Simon Coupland

Centre for Computational Intelligence, Faculty of Technology, De Montfort University
Leicester, UK
Email: hussamw@dmu.ac.uk, simonc@dmu.ac.uk

Abstract— This paper examines type-2 fuzzy arithmetic using interval analysis. It relies heavily on alpha-cuts and alpha-planes. Furthermore, we discuss the use of quasi type-2 fuzzy sets proposed by Mendel and Liu and define quasi type-2 fuzzy numbers. Arithmetic operations of such numbers are defined and a worked example is presented.

Keywords— Fuzzy Arithmetic, Type-2 Fuzzy Arithmetic, Type-2 Fuzzy Numbers, Quasi Type-2 Fuzzy Numbers.

1 Introduction

Zadeh introduced the concept of type-2 fuzzy sets in the first of a trio of papers in 1975 [1]. In these papers he also defined linguistic variables, interval-valued fuzzy sets and their representations using the resolution identity (later known as α-cuts). Type-2 fuzzy logic has gained much attention recently due to its ability to handle uncertainty [2], and many advances appeared in both theory and applications [3, 4, 5]. Type-2 fuzzy numbers and the associated arithmetic operations have not received the same attention, only two main contributions appear in the literature that specificly target arithmetic operations using the representation theorem and the extension principle [6, 7]. Recent work on fuzzy systems has proposed an extension to α-cuts for type-2 fuzzy sets through the notion of α-planes [8]. Further work by Mendel and Liu [9] defined a quasi type-2 fuzzy logic system as a restricted special case of a type-2 fuzzy logic system represented by its α-planes.

In this paper we examine type-2 fuzzy arithmetic through α-planes. We use arithmetic operations on type-1 and interval type-2 fuzzy sets through their α-cut representation, moreover, we define quasi type-2 fuzzy numbers and derive their arithmetic operations.

This paper is organised as follows: Section 2 reviews some basic type-2 fuzzy set definitions used in this paper; Section 3 defines type-2 fuzzy numbers and quasi type-2 fuzzy numbers; Section 4 derives the arithmetic operations for these numbers; Section 5 provides a worked example; finally Section 6 provides a conclusion.

2 Type-2 Fuzzy Sets

We present a review of the basic terminology used in this paper. A type-2 fuzzy set (**T2FS**) [2, 10] is defined by equation (1)

$$\widetilde{A} = \int_{\forall x \in X} \int_{\forall u \in J_x \subseteq [0,1]} \mu_{\widetilde{A}}(x,u)/(x,u) \qquad (1)$$

where $\int\int$ denotes the union over all admissible *domain values* x and *secondary domain values* u, and $\mu_{\widetilde{A}}(x,u)$ is a *type-2 membership function*, a T2FS is three dimensional (3D). The *Vertical Slice* (**VS**) is the two dimensional (2D) plane in the u and $\mu_{\widetilde{A}}(x,u)$ axes for a single value of $x = x^{'}$, then the VS is defined by equation (2).

$$VS(x^{'}) = \mu_{\widetilde{A}}(x^{'},u) \equiv \mu_{\widetilde{A}}(x^{'}) = \int_{u \in J_{x^{'}}} f_{x^{'}}(u)/u \qquad (2)$$

where $f_{x^{'}}(u) \in [0,1]$ is called the *secondary grade* and J_x represents the domain of the *secondary membership function* called the *secondary domain*, of course the VS is a *type-1 fuzzy set* (**T1FS**) in $[0,1]$. The *Vertical Slice Representation* (**VSR**) of T2FS is represented by the union of all the vertical slices

$$\widetilde{A} = \{(x, \mu_{\widetilde{A}}(x)) | \forall x \in X\} \qquad (3)$$

The *Footprint Of Uncertainty* (**FOU**) is derived from the union of all primary memberships

$$FOU(\widetilde{A}) = \int_{x \in X} J_x \qquad (4)$$

Another important notion is the *principal membership function* (**PrMF**) defined as the union of all the primary memberships having secondary grades equal to 1

$$Pr(\widetilde{A}) = \int_{x \in X} u/x | f_x(u) = 1 \qquad (5)$$

This notion allows us to view a T1FS as special case of T2FS in which its PrMF is obtained for only one primary membership having a secondary grade at unity [11, 12]. Based on [2] the FOU is described to be bounded by two membership functions a lower $\underline{\mu_{\widetilde{A}}(x)}$ and an upper $\overline{\mu_{\widetilde{A}}(x)}$. The FOU can be described in terms of its upper and lower membership functions (**MF**s)

$$FOU(\widetilde{A}) = \int_{x \in X} \left[\underline{\mu_{\widetilde{A}}(x)}, \overline{\mu_{\widetilde{A}}(x)}\right] \qquad (6)$$

Interval type-2 fuzzy set (**IT2FS**) is defined to be a T2FS where all its secondary grades are of unity $\forall f_x(u) = 1$. A IT2FS can be completely determined using its FOU given in equation (6). Recalling from Zadeh [1] an *α-level set*, A_α, that comprises of elements $x \in X$ of a type-1 fuzzy subset A of universe X

$$A_\alpha = \{x | \mu_A(x) \geq \alpha\} \qquad (7)$$

where $\alpha \in [0,1]$, the fuzzy set A can be represented (decomposed) as

$$A = \bigcup_{\alpha \in [0,1]} \alpha.A_\alpha \qquad (8)$$

ISBN: 978-989-95079-6-8

where $\mu_A(x) = \sup_{x \in A_\alpha} \alpha$ and (sup) denotes the supremum, this decomposition theorem [13] is called the *α-cut Representation*. The same analogy is used by Tahayori *et al.* [14]:

$$\widetilde{A}_{\widetilde{\alpha}} = \{(x,u) | f_x(u) \geq \widetilde{\alpha}\} \tag{9}$$

and the T2FS $\widetilde{A}$ can be represented (decomposed) as

$$\widetilde{A} = \bigcup_{\widetilde{\alpha} \in [0,1]} \widetilde{\alpha}.\widetilde{A}_{\widetilde{\alpha}} \tag{10}$$

where[1] $f_x(u) = \sup_{(x,u) \in \widetilde{A}_{\widetilde{\alpha}}} \widetilde{\alpha}$. Liu [8] calls equation (9) an *α-plane* due to its 2D nature and consequently equation (10) is called the *α-plane Representation*. Liu [8] and Wagner and Hagras [15] noted that when $\widetilde{\alpha} = 0$, $\widetilde{A}_{\widetilde{\alpha}}$ is actually the FOU and Liu and Mendel [9] generalise it to a footprint of uncertainty at plane $\widetilde{\alpha}$ denoted by $FOU(\widetilde{A}_{\widetilde{\alpha}})$. We are interested in the use of this representation as these α-planes can compute some of T2FS operations using IT2FS operations by applying Zadeh's extension principle [1] as follows

$$f(\widetilde{A}) = f(\bigcup_{\widetilde{\alpha} \in [0,1]} \widetilde{\alpha}.\widetilde{A}_{\widetilde{\alpha}}) = \bigcup_{\widetilde{\alpha} \in [0,1]} \widetilde{\alpha}.f(\widetilde{A}_{\widetilde{\alpha}}) \tag{11}$$

As Zadeh himself noted that this principle can be used with functions between sets, relations, etc. and it can extend any function from crisp sets to its fuzzy and obviously type-2 fuzzy counterparts. In [9] based on observations about the shape of the centroid Liu and Mendel defined the notion of *quasi type-2 fuzzy logic system* (QT2FLS). They propose approximating the T2FS using two α-planes ($\widetilde{A}_{\widetilde{\alpha}=0}$ and $\widetilde{A}_{\widetilde{\alpha}=1}$). Great use of these methods is made later in this paper.

3 Type-2 Fuzzy Numbers

A type-1 fuzzy number (T1FN) is defined as a fuzzy set that is both normal and convex [16, 17]. Normality is required in order to capture the concept of a fuzzy number being a set of real numbers close to a specific crisp number [13], in other words when all the uncertainty about a number disappears it reduces to a crisp number. Convexity is required as it allows meaningful arithmetic operations to be performed on fuzzy sets using the well established methods from interval analysis since α-level sets are closed intervals [13]. Coupland *et al.* [6, 7] define a type-2 fuzzy number (T2FN) as a type-2 fuzzy set having a numerical domain. Although no assumption for normality has been defined for these T2FN the examples used assumed normality. From another perspective within the framework of interval-valued fuzzy sets (IVFS) [18], which are equal to IT2FS [5, 19], interval-valued fuzzy numbers (IVFN) are defined to be convex and normal [20, 21]. Here it is desired to define a host of cases that may be conceived to qualify as T2FN. First normality is defined, recalling its definition within the context of T1FN.

Definition 3.1 (Normal T1FS) *A T1FS, A, is said to be normal if its height $h(A)$ is equal to 1 i.e. $\sup \mu_A(x) = 1$.*

Whenever a T1FS is not normal it is called *subnormal*. In IT2FS we differentiate between two cases of normality, one when both the upper and lower MFs are normal, and the other when only the upper T1FS is normal.

[1] Here $\widetilde{\alpha}$ is used instead of α to distinguish that these are α-cuts in the third dimension.

Definition 3.2 (Normal IT2FS) *A IT2FS, $\widetilde{A}$, is said to be normal if its upper MF is normal i.e. $\sup \overline{\mu_A(x)} = 1$.*

Note that there is a point in which a crisp number can be reached which is only depending on the upper MF, this can be seen as a less restrictive condition and is widely used in applications (e.g. Computing with words [22]).

Definition 3.3 (Perfectly Normal IT2FS) *A IT2FS, $\widetilde{A}$, is said to be perfectly[2] normal if both its upper and lower MFs are normal i.e. $\sup \overline{\mu_A(x)} = \sup \underline{\mu_A(x)} = 1$.*

Here the crisp number is reached when uncertainties about both the upper and lower MFs disappear, it is more restrictive, but conceptually more appealing since it generalises the concept of normal T1FS with only a single peak point where it is completely crisp. It can be seen as a special case of a normal IT2FS, this is clear in figure (1).

Definition 3.4 (Partially Normal T2FS) *A T2FS, $\widetilde{A}$, is said to be partially normal if its FOU is **a normal IT2FS.***

This is the weakest case in which a T2FS can qualify to be a number. There is an argument, this can not qualify to be a fuzzy number at all as the secondary membership function is clearly not about 1 at any point (i.e $f_x(u) \neq 1, \forall f_x(u)$).

Definition 3.5 (Normal T2FS) *A T2FS, $\widetilde{A}$, is said to be normal if its FOU is **a normal IT2FS** and it has a PrMF.*

Definition 3.6 (Perfectly Normal T2FS) *A T2FS, $\widetilde{A}$, is said to be perfectly normal if its FOU is a perfectly normal IT2FS and it has a PrMF which is normal (i.e. either a normal T1FS or a normal IT2FS).*

Second, recalling the general definition of a T1FN using piecewise-defined functions.

Definition 3.7 (T1FN[13]) *Let A be a fuzzy subset on real numbers. Then, A is a fuzzy number if and only if there exists a closed interval $[m_1, m_2] \neq \phi$ such that*

$$\mu_A(x) = \begin{cases} 1 & x \in [m_1, m_2] \\ l(x) & x \in [s, m_1) \\ r(x) & x \in (m_2, e] \\ 0 & x \in (-\infty, s); x \in (e, \infty) \end{cases} \tag{12}$$

where $l(x) \in [0,1]$ is monotonically increasing and continuous from the right; $r(x) \in [0,1]$ is monotonically decreasing and continuous from the left; $\langle s, m_1, m_2, e \rangle$ are the parameters that define the T1FN.

Observe that if A is subnormal it is not considered a number, but it is useful to define a *type-1 fuzzy sub-number* (T1FsN) which satisfies all the properties of a T1FN except it is subnormal, this can be defined for a fuzzy set A with height $h(A) = \sup_{\forall x} \mu_A(x) = h_A$ as follows

$$\mu_A(x) = \begin{cases} h_A & x \in [m_1, m_2] \\ l(x) & x \in [s, m_1) \\ r(x) & x \in (m_2, e] \\ 0 & x \in (-\infty, s); x \in (e, \infty) \end{cases} \tag{13}$$

[2] this term has been used by Kaufmann and Gupta [17] describing a perfectly triangular T2FN.

ISBN: 978-989-95079-6-8

Using equations (7) and (8) a T1FN can be represented as intervals using their α-cuts $A_\alpha = [a_1^\alpha, a_2^\alpha]$ as follows

$$A = \bigcup_{\forall\alpha} \alpha.[a_1^\alpha, a_2^\alpha] = \bigcup_{\forall\alpha} \alpha.[l^{-1}(\alpha), r^{-1}(\alpha)] \quad (14)$$

IT2FN can be represented by its lower and upper membership functions which themselves are T1FS. Here we use the terms above to define IT2FN. Let $FOU(\widetilde{A}) = [\underline{FOU(\widetilde{A})}, \overline{FOU(\widetilde{A})}]$ [3] to represent a perfectly normal IT2FS defined by its lower and upper membership functions, respectively. Then, using equation (14) individually for each of the memberships, it follows that

$$\underline{FOU(\widetilde{A})} = \bigcup_{\forall\alpha} \alpha.[\underline{a_1^\alpha}, \underline{a_2^\alpha}] = \bigcup_{\forall\alpha} \alpha.[\underline{l^{-1}}(\alpha), \underline{r^{-1}}(\alpha)]$$

$$\overline{FOU(\widetilde{A})} = \bigcup_{\forall\alpha} \alpha.[\overline{a_1^\alpha}, \overline{a_2^\alpha}] = \bigcup_{\forall\alpha} \alpha.[\overline{l^{-1}}(\alpha), \overline{r^{-1}}(\alpha)]$$

then

$$FOU(\widetilde{A}) = \bigcup_{\forall\alpha} \alpha.\left[[\overline{a_1^\alpha}, \underline{a_1^\alpha}], [\underline{a_2^\alpha}, \overline{a_2^\alpha}]\right] \quad (15)$$

for a normal IT2FS with lower membership function height, $h(FOU(\widetilde{A})) = h_l$, a IT2FN can be represented as

$$FOU(\widetilde{A}) = \begin{cases} \bigcup_{\forall\alpha} \alpha.\left[[\overline{a_1^\alpha}, \underline{a_1^\alpha}], [\underline{a_2^\alpha}, \overline{a_2^\alpha}]\right] & \alpha \le h_l \\ \bigcup_{\forall\alpha} \alpha.[\overline{a_1^\alpha}, \overline{a_2^\alpha}] & \alpha > h_l \end{cases} \quad (16)$$

Then a IT2FN can be defined by definition (3.8)

Definition 3.8 (IT2FN)
Let $FOU(\widetilde{A}) = [\underline{FOU(\widetilde{A})}, \overline{FOU(\widetilde{A})}]$ be an interval type-2 fuzzy subset on real numbers. Then, $FOU(\widetilde{A})$ is an interval type-2 fuzzy number if:

- *$\underline{FOU(\widetilde{A})}$ and $\overline{FOU(\widetilde{A})}$ are T1FNs, in this case it is called a perfectly normal IT2FN.*
- *$\underline{FOU(\widetilde{A})}$ is a T1FsN [4] and $\overline{FOU(\widetilde{A})}$ is a T1FN, in this case it is called a normal IT2FN.*

This definition can be seen in figure (1), observe that both IT2FSs are IT2FNs. We can also consider an interval type-2 fuzzy sub-number (IT2FsN) when bothe the upper and lower MFs are T1FsNs. Now, it is desired to define T2FN using the same terminology. From equation (10) a T2FS

$$\widetilde{A} = \bigcup_{\forall\widetilde{\alpha}} \widetilde{\alpha}.FOU(\widetilde{A}_{\widetilde{\alpha}})$$

and using equation (15) a perfectly normal T2FS

$$\widetilde{A} = \bigcup_{\forall\widetilde{\alpha}} \widetilde{\alpha}.(\bigcup_{\forall\alpha} \alpha.\left[[\overline{a_1^\alpha}, \underline{a_1^\alpha}], [\underline{a_2^\alpha}, \overline{a_2^\alpha}]\right]) \quad (17)$$

and using equation (16) a normal T2FS

$$\widetilde{A} = \begin{cases} \bigcup_{\forall\widetilde{\alpha}} \widetilde{\alpha}.(\bigcup_{\forall\alpha} \alpha.\left[[\overline{a_1^\alpha}, \underline{a_1^\alpha}], [\underline{a_2^\alpha}, \overline{a_2^\alpha}]\right]) & \alpha \le h_l \\ \bigcup_{\forall\widetilde{\alpha}} \widetilde{\alpha}.(\bigcup_{\forall\alpha} \alpha.[\overline{a_1^\alpha}, \overline{a_2^\alpha}]) & \alpha > h_l \end{cases} \quad (18)$$

A partially normal T2FS will only require $\forall\widetilde{\alpha}; \widetilde{\alpha} \in [0, \sup_{\forall(x,u)} f_x(u)]$.

[3] FOU is used here since the FOU fully describes the IT2FS.
[4] see equation (13).

Definition 3.9 (T2FN) *Let $\widetilde{A} = \bigcup_{\forall\widetilde{\alpha}} \widetilde{\alpha}.FOU(\widetilde{A}_{\widetilde{\alpha}})$ be a type-2 fuzzy subset on real numbers. Then, $\widetilde{A}$ is a type-2 fuzzy number if:*

- *$FOU(\widetilde{A}_{\widetilde{\alpha}=0})$ is a perfectly normal IT2FN, and $FOU(\widetilde{A}_{\widetilde{\alpha}=1}) = Pr(\widetilde{A})$ is normal (i.e. either a T1FN or a normal IT2FN), in this case it is called a perfectly normal T2FN.*
- *$FOU(\widetilde{A}_{\widetilde{\alpha}=0})$ is a normal IT2FN, and $FOU(\widetilde{A}_{\widetilde{\alpha}=1})$ ($Pr(\widetilde{A})$) exist, in this case it is called a normal T2FN.*
- *$FOU(\widetilde{A}_{\widetilde{\alpha}=0})$ is a normal IT2FN, in this case it is called a partially normal T2FN.*

Note that a special case of $FOU(\widetilde{A}_{\widetilde{\alpha}=1}) = Pr(\widetilde{A})$ is when all the vertical slices constructing $Pr(\widetilde{A})$ are triangular T1FN then $Pr(\widetilde{A})$ is T1FS, this is depicted in figures (2) and (3). We can also define a type-2 fuzzy sub-number (T2FsN) when its FOU is an IT2FsN. In some applications one may need to restrict a fuzzy set to a specific form, e.g. in computing with words Klir *et al.* [23, 24] defined a procedure to convert any given convex fuzzy set to a fuzzy interval that is expressed in some standard form using some specific criteria. We now examine a special form of T2FN based on some ideas from [9]. Some observations about the shape of a centroid led to the proposition of a quasi type-2 fuzzy logic system, similirly, a definition of a quasi T2FN (QT2FN) can be proposed.

Assumption 3.10 (QT2FS) *Consider the following propositions about a T2FN*

A1 The T2FN is Normal (i.e FOU is normal IT2FN and PrMF is normal).

A2 The FOU upper and lower membership functions, the PrMF, and the vertical slices are characterised by piecewise functions.

A3 All the vertical slices that construct PrMF are T1FN.

A4 As for the rest of the vertical slices, they are T1FsN in which their supremum lay on piece-wise functions between the sides of the upper membership function of the FOU and the PrMF denoted $\widetilde{l}(x)$ on the left side, and $\widetilde{r}(x)$ on the right side.

A5 All the mid points of the FOU upper and lower membership functions, and the PrMF are at the same domain value.

these assumptions allow a T2FN to be completely determined by its FOU and PrMF, just like a T1FS based on certain assumptions can be completely determined by its core and support.

Definition 3.11 (QT2FN) *Let $\widetilde{A} = \bigcup_{\forall\widetilde{\alpha}} \widetilde{\alpha}.FOU(\widetilde{A}_{\widetilde{\alpha}})$ be a type-2 fuzzy subset on real numbers. Then, $\widetilde{A}$ is a quasi type-2 fuzzy number if it is completely determined by its FOU and PrMF.*

Figures (3) and (4) are both QT2FNs.

ISBN: 978-989-95079-6-8

4 Arithmetic using Alpha-planes

In Kaufmann and Gupta [17] a comprehensive discussion on fuzzy numbers and arithmetic operations which is formulated and mostly dependent upon the interval of confidence (α-level sets). Recalling operations on intervals, let $[a_1, a_2]$ and $[b_1, b_2]$ be two interval numbers, then

$$[a_1, a_2] \oplus [b_1, b_2] = [a_1 \oplus b_1, a_2 \oplus b_2]$$

$$[a_1, a_2] \otimes [b_1, b_2] = [\min(a_1 \otimes b_1, a_1 \otimes b_2, a_2 \otimes b_1, a_2 \otimes b_2) , \max(a_1 \otimes b_1, a_1 \otimes b_2, a_2 \otimes b_1, a_2 \otimes b_2)] \quad (19)$$

where $\oplus \in \{+, -\}$, $\otimes \in \{\times, \div\}$ and $0 \notin B$ if $\otimes = \div$. Interval operations are extended to T1FN. Let $A = \bigcup_{\forall \alpha} \alpha.[a_1^\alpha, a_2^\alpha]$ and $B = \bigcup_{\forall \alpha} \alpha.[b_1^\alpha, b_2^\alpha]$ be two T1FN, then

$$A \circ B = \bigcup_{\forall \alpha} \alpha.([a_1^\alpha, a_2^\alpha] \circ [b_1^\alpha, b_2^\alpha]) \quad (20)$$

where $\circ = \{+, -, \times, \div\}$. Also these operations are extended to IT2FN. Let $FOU(\widetilde{A}) = \bigcup_{\forall \alpha} \alpha. \left[\overline{[a_1^\alpha, a_1^\alpha]}, \underline{[a_2^\alpha, a_2^\alpha]}\right]$ and $FOU(\widetilde{B}) = \bigcup_{\forall \alpha} \alpha. \left[\overline{[b_1^\alpha, b_1^\alpha]}, \underline{[b_2^\alpha, b_2^\alpha]}\right]$ be two perfectly normal IT2FN, then [17]

$$FOU(\widetilde{A}) \circ FOU(\widetilde{B}) = \bigcup_{\forall \alpha} \alpha.(\left[\overline{[a_1^\alpha, a_1^\alpha]} \circ \overline{[b_1^\alpha, b_1^\alpha]}, \underline{[a_2^\alpha, a_2^\alpha]} \circ \underline{[b_2^\alpha, b_2^\alpha]}\right]) \quad (21)$$

If $FOU(\widetilde{A})$ and $FOU(\widetilde{B})$ are normal IT2FN with lower membership function heights $h_{\widetilde{A}}$ and $h_{\widetilde{B}}$ respectively, then the following changes are made to equation (21)

- if $0 \leq \alpha \leq \min(h_{\widetilde{A}}, h_{\widetilde{B}})$ no changes are made.
- if $h_{\widetilde{A}} < \alpha \leq h_{\widetilde{B}}$ then $\underline{a_1^\alpha} = \underline{a_1^{h_{\widetilde{A}}}}$; $\underline{a_2^\alpha} = \underline{a_2^{h_{\widetilde{A}}}}$, and if $h_{\widetilde{B}} < \alpha \leq h_{\widetilde{A}}$ then $\underline{b_1^\alpha} = \underline{b_1^{h_{\widetilde{B}}}}$; $\underline{b_2^\alpha} = \underline{b_2^{h_{\widetilde{B}}}}$.
- if $\max(h_{\widetilde{A}}, h_{\widetilde{B}}) < \alpha \leq 1$ then equation (21) becomes $FOU(\widetilde{A}) \circ FOU(\widetilde{B}) = \bigcup_{\forall \alpha} \alpha.([\overline{a_1^\alpha, a_2^\alpha}] \circ [\overline{b_1^\alpha, b_2^\alpha}])$.

Wu and Mendel [25] noticed that such methods that result in discontinuous or nonconvex sets are neither desirable nor technically correct. Similarly such results appear for normal IT2FN arithmetic operations. According to Wu and Mendel [25] the following changes are made to equation (21)

$$FOU(\widetilde{A}) \circ FOU(\widetilde{B}) = \begin{cases} \bigcup_{\forall \alpha} \alpha.(\left[\overline{[a_1^\alpha, a_1^\alpha]} \circ \overline{[b_1^\alpha, b_1^\alpha]}, \underline{[a_2^\alpha, a_2^\alpha]} \circ \underline{[b_2^\alpha, b_2^\alpha]}\right]) \\ \quad , \text{if } 0 \leq \alpha \leq \min(h_{\widetilde{A}}, h_{\widetilde{B}}) \\ \bigcup_{\forall \alpha} \alpha.([\overline{a_1^\alpha, a_2^\alpha}] \circ [\overline{b_1^\alpha, b_2^\alpha}]) \\ \quad , \text{if } \min(h_{\widetilde{A}}, h_{\widetilde{B}}) < \alpha \leq 1 \end{cases} \quad (22)$$

Clearly T2FN arithmetic can be extended, let $\widetilde{A} = \bigcup_{\forall \widetilde{\alpha}} \widetilde{\alpha}.FOU(\widetilde{A}_{\widetilde{\alpha}})$ and $\widetilde{B} = \bigcup_{\forall \widetilde{\alpha}} \widetilde{\alpha}.FOU(\widetilde{B}_{\widetilde{\alpha}})$ be two T2FNs, then

$$\widetilde{A} \circ \widetilde{B} = \bigcup_{\forall \widetilde{\alpha}} \widetilde{\alpha}.FOU(\widetilde{A}_{\widetilde{\alpha}}) \circ \bigcup_{\forall \widetilde{\alpha}} \widetilde{\alpha}.FOU(\widetilde{B}_{\widetilde{\alpha}})$$

using the extension principle in equation (11) it follows that

$$\widetilde{A} \circ \widetilde{B} = \bigcup_{\forall \widetilde{\alpha}} \widetilde{\alpha}.(FOU(\widetilde{A}_{\widetilde{\alpha}}) \circ FOU(\widetilde{B}_{\widetilde{\alpha}})) \quad (23)$$

then we can use equation (21) for perfectly normal T2FNs or equation (22) for normal T2FNs. Next we examine QT2FN, let $\widetilde{A}^Q$ and $\widetilde{B}^Q$ be QT2FN completely determined by the FOU and PrMF, $\langle FOU(\widetilde{A}^Q_{\widetilde{\alpha}=0}), Pr(\widetilde{A}^Q)\rangle$ and $\langle FOU(\widetilde{B}^Q_{\widetilde{\alpha}=0}), Pr(\widetilde{B}^Q)\rangle$, respectively. Then the result of basic arithmetic operations between them is a QT2FN

$$\widetilde{C}^Q = \widetilde{A}^Q \circ \widetilde{B}^Q \quad (24)$$

completely determined by $\langle FOU(\widetilde{C}^Q_{\widetilde{\alpha}=0}), Pr(\widetilde{C}^Q)\rangle$ where $FOU(\widetilde{C}^Q_{\widetilde{\alpha}=0}) = FOU(\widetilde{A}^Q_{\widetilde{\alpha}=0}) \circ FOU(\widetilde{B}^Q_{\widetilde{\alpha}=0})$ and $Pr(\widetilde{C}^Q) = Pr(\widetilde{A}^Q) \circ Pr(\widetilde{B}^Q)$. This operation can only be performed for addition and subtraction as the functions $\widetilde{l}(x)$ and $\widetilde{r}(x)$ are preserved. In the case of multiplication and devision an approximation for these functions should be used. It has to be mentioned that methods to approximate a T1FN to some standard form is used in the literature (see Grzegorzewski [26]). Providing approximation to QT2FN is still an open question.

5 Worked Example

In this example we only consider QT2FN as it gives sufficient insight on T2FN. Let us consider the following triangular QT2FN $\widetilde{3}^Q = \langle FOU(\widetilde{3}^Q_{\widetilde{\alpha}=0}), Pr(\widetilde{3}^Q)\rangle$ depicted in figure (5) with parameters[5] $FOU(\widetilde{3}^Q_{\widetilde{\alpha}=0}) = \langle 1.5, 2.25, 3, 3.45, 4.75\rangle$, $h_{\widetilde{3}} = 0.6$, and $Pr(\widetilde{3}^Q) = \langle 1.75, 3, 4.25\rangle$. And the QT2FN $\widetilde{12}^Q = \langle FOU(\widetilde{12}^Q_{\widetilde{\alpha}=0}), Pr(\widetilde{12}^Q)\rangle$ depicted in figure (6) with parameters $FOU(\widetilde{12}^Q_{\widetilde{\alpha}=0}) = \langle 10.25, 11.5, 12, 12.5, 14\rangle$, $h_{\widetilde{12}} = 0.7$, and $Pr(\widetilde{12}^Q) = \langle 10.75, 12, 13.5\rangle$. When computing the addition $\widetilde{3}^Q + \widetilde{12}^Q$, first, we determine a suitable number of α-cuts along u for both FOU and PrMF[6]. In our case and for the sake of clarity we discretised u into 25 α-cuts. Then, we apply equation (24) to the decomposed IT2FS. This gives the result depicted in figure (7) with parameters $FOU(\widetilde{15}^Q_{\widetilde{\alpha}=0}) = \langle 11.75, 13.75, 15, 15.95, 18.75\rangle$, $h_{\widetilde{15}} = 0.6$, and $Pr(\widetilde{15}^Q) = \langle 12.5, 15, 17.75\rangle$. This result we would expect, $\widetilde{3}^Q + \widetilde{12}^Q = \widetilde{15}^Q$.

6 Conclusions

In this paper we presented methods to perform type-2 fuzzy arithmetic operations using well known arithmetic operations on intervals. In order to better analyse our methods we defined a set of terms that describe different normality formations of IT2FS and T2FS, some of which, have already been used in the literature without a clear definition. Our methods relied heavily on interval analysis through α-cuts and α-planes that allow T2FSs be represented as a collection of intervals. Furthermore, we examined the use of quasi type-2 fuzzy sets proposed by Mendel and Liu and defined quasi type-2 fuzzy numbers as a special case of type-2 fuzzy numbers. Finally, we derived arithmetic operations for quasi type-2 fuzzy numbers and provided an illustration by a worked example.

[5] The parameters used here are $FOU(\widetilde{A}^Q_{\widetilde{\alpha}=0}) = \langle s_1, s_2, m_1, e_2, e_1\rangle$ derived from equation (12) where $\overline{FOU(\widetilde{A}^Q_{\widetilde{\alpha}=0})} = \langle s_1, m_1, e_1\rangle$ and $\underline{FOU(\widetilde{A}^Q_{\widetilde{\alpha}=0})} = \langle s_2, m_1, e_2\rangle$.

[6] In the case of T2FN we first discretise along $f_x(u)$ in order to determine a suitable number of α-planes, then, we discretise along u for each of the α-planes.

ISBN: 978-989-95079-6-8

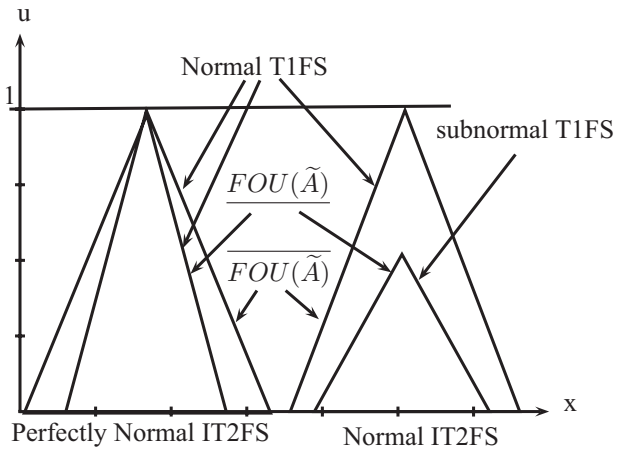


Figure 1: Different types of normal fuzzy sets

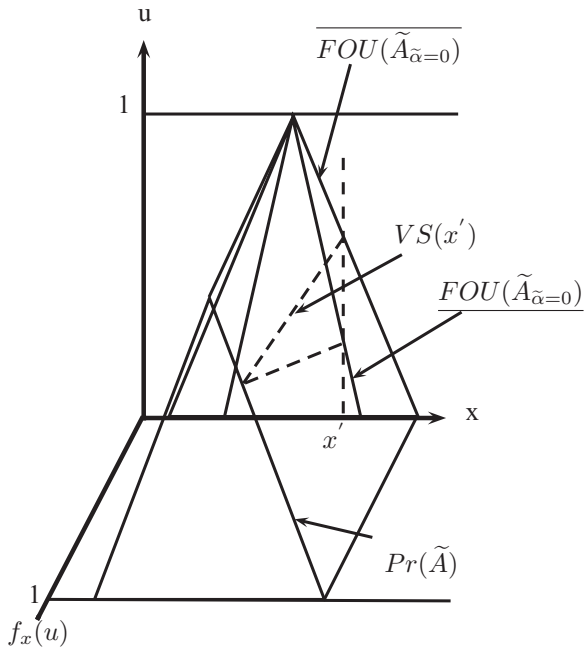


Figure 2: 3D representation of a perfectly normal T2FN, it is also a QT2FN.

We believe that this paper will form the basis for further work on type-2 fuzzy numbers and thier applications. Quasi type-2 fuzzy numbers may be viewed as a next step in the progress between IT2FN and T2FN. Arithmetic operations are already used in aggregation and averaging operations, measures that apply certain mathematical functions, and fusion functions between one system and another.

Further work with regards to the operations defined here will include approximation of standard forms of Quasi type-2 fuzzy numbers, and a thorough comparison between these methods and previous methods in terms of computational complexity, although, it is almost trivial that our methods are computationally sound as they are based on interval analysis.

Acknowledgment

The authors would like to thank the anonymous reviewers for their constructive comments which we believe have helped to improve this paper, and also would like to thank Mr. Simon Miller for his helpful discussions.

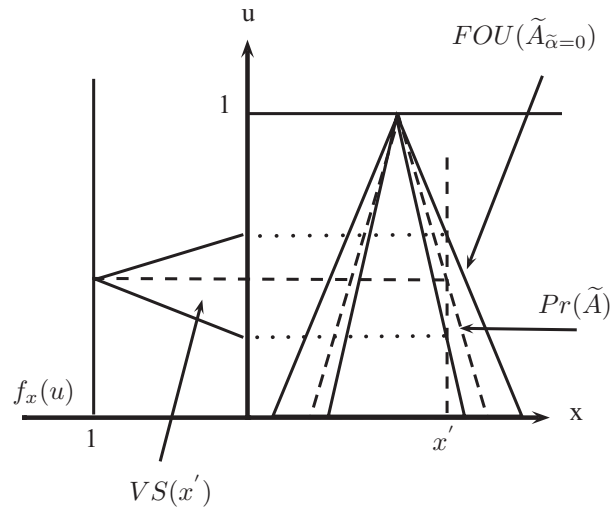


Figure 3: 2D representation of a perfectly normal T2FN of figure (2) with triangular vertical slices.

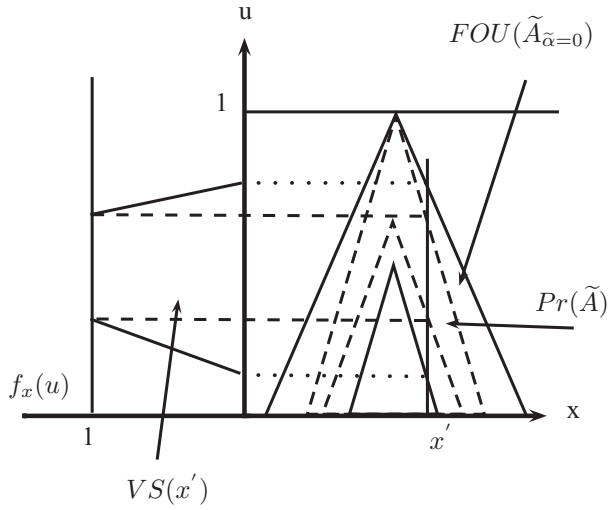


Figure 4: 2D representation of a normal T2FN with trapizoidal vertical slices, note that $Pr(\widetilde{A})$ is normal IT2FN.

References

[1] L. A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning-1. *Information Sciences*, 8:199 – 249, 1975.

[2] J.M. Mendel. *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. Prentice Hall, Upper Saddle River, NJ, 2001.

[3] R.I. John. Type-2 fuzzy sets:an appraisal of theory and applications. *Uncertainty, Fuzziness and Knowledge-Based Systems*, 6:563–576, 1998.

[4] R.I. John and S. Coupland. Type-2 fuzzy logic a historical view. *IEEE Computational Intelligence Magazine*, 2(1):57–62, February 2007.

[5] J. M. Mendel. Advances in type-2 fuzzy sets and systems. *Information Sciences*, 177(1):84–110, January 2007.

[6] S. Coupland and R.I. John. An approach to type-2 fuzzy arithmetic. In *Proc. UK Workshop on Computational Intelligence*, pages 107 – 114, 2003.

ISBN: 978-989-95079-6-8

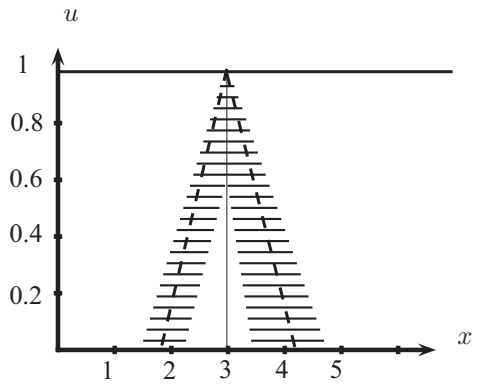


Figure 5: $\widetilde{3}^Q$ with discretised α-cuts, and the dashed line is $Pr(\widetilde{3}^Q)$

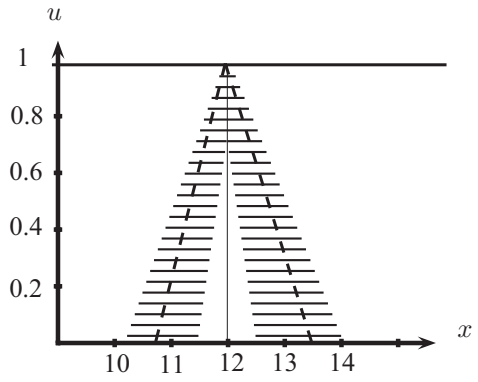


Figure 6: $\widetilde{12}^Q$ with discretised α-cuts, and the dashed line is $Pr(\widetilde{12}^Q)$

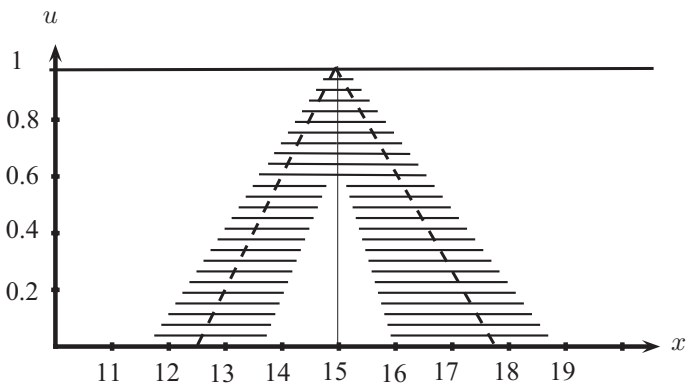


Figure 7: $\widetilde{15}^Q$ with discretised α-cuts, and the dashed line is $Pr(\widetilde{15}^Q)$

[7] W. Blewitt, S. Coupland, and S. M. Zhou. A novel approach to type-2 fuzzy addition. In *Proc. FUZZ-IEEE 2007*, pages 1456 – 1461, London, 2007.

[8] F. Liu. An efficient centroid type reduction strategy for general type-2 fuzzy logic systems. Walter J. Karplus Summer Research Grant Report, IEEE Computational Intelligence Society, 2006.

[9] J.M. Mendel and F. Liu. On new quasi-type-2 fuzzy logic systems. In *FUZZ-IEEE 2008*, pages 354–360, Hong Kong, 2008.

[10] J. M. Mendel and R.I. John. Type-2 fuzzy sets made simple. *IEEE Transaction on Fuzzy Systems*, 10(2):117–127, 2002.

[11] N.N. Karnik and J.M. Mendel. Operations on type-2 fuzzy set. *Fuzzy Sets and Systems*, 122(2):327–348, 2001.

[12] L.A. Lucas, T.M. Centeno, and M.R. Delgado. General type-2 fuzzy classifiers to land cover classification. In *Proceedings of the 2008 ACM symposium on Applied computing*, pages 1743–1747. ACM New York, NY, USA, 2008.

[13] G.J. Klir and B. Yuan. *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Prentice Hall, Upper Saddle River, NJ, 1995.

[14] H. Tahayori, A. Tettamanzi, and G.D. Antoni. Approximated type-2 fuzzy set operations. In *Proc. FUZZ-IEEE 2006*, pages 9042 – 9049, Vancouver, Canada, July 2006.

[15] C. Wagner and H. Hagras. zSlices–Towards bridging the gap between Interval and General Type-2 Fuzzy Logic. In *FUZZ-IEEE 2008.*, pages 489–497, Hong Kong, 2008.

[16] D. Dubois and H. Prade. *Fuzzy Sets and Systems: Theory and Applications*, volume 144. Academic Press INC, NY, USA, 1980.

[17] A. Kaufmann and M. Gupta. *Introduction to Fuzzy Arithmetic Theory and Applications*. Van Nostran Reinhold Co. Inc., 1985.

[18] M.B. Gorzalczany. A method of inference in approximate reasoning based on interval-valued fuzzy sets. *Fuzzy Sets and Systems*, 21:1–17, January 1987.

[19] J.M. Mendel, R.I. John, and F. Liu. Interval type-2 fuzzy logic systems made simple. *IEEE Transaction on Fuzzy Systems*, 14(6):808–821, December 2006.

[20] D. H. Hong and S. Lee. Some algebraic properties and a distance measure for interval-valued fuzzy numbers. *Information Sciences*, 148(1-4):1–10, December 2002.

[21] G. Wang and X. Li. The applications of interval-valued fuzzy numbers and interval-distribution numbers. *Fuzzy Sets and Systems*, 98(3):331–335, 1998.

[22] J.M. Mendel and D. Wu. Perceptual Reasoning: A New Computing with Words Engine. In *Granular Computing, 2007. GRC 2007. IEEE International Conference on*, pages 446–446, 2007.

[23] Olga Martin and George J. Klir. On the problem of retranslation in computing with preceptions. *International Journal of General Systems*, 35(6):655–674, 2006.

[24] G.J. Klir and K. Sentz. *Aspects of Automatic Text Analysis*, chapter On the issue of linguistic approximation. Springer, 2006.

[25] D. Wu and J. M. Mendel. Corrections to "Aggregation Using the Linguistic Weighted Average and Interval Type-2 Fuzzy Sets". *IEEE Transactions on Fuzzy Systems*, Accepted for future publication.

[26] P. Grzegorzewski. Trapezoidal approximations of fuzzy numbers preserving the expected interval–Algorithms and properties. *Fuzzy Sets and Systems*, 159:1354–1364, 2008.

ISBN: 978-989-95079-6-8

Operations for Real-Valued Bags and Bag Relations

Sadaaki Miyamoto
Department of Risk Engineering, University of Tsukuba
Email: miyamoto@risk.tuskuba.ac.jp

Abstract— *A generalization of bags is proposed and fundamental operations are described. That is, a real-valued bag is proposed which is a straightforward generalization of the traditional bags, but is necessary for complementation, and s-norm and t-norm operations. We mainly describe how complementation and s-norms are related. As a result a duality theorem between s-norm and t-norm is obtained. R-bag relations with max-s norm composition and max-t norm composition are also proposed. Further generalization that includes fuzzy bags and real-valued bags having membership of a region in a plane is moreover discussed.*

Keywords— bag relation, max-s composition, real-valued bag, set-valued bag, t-norm and conorm .

1 Introduction

Bags which are also called *multisets* have long been studied by computer scientists [5, 6]. In soft computing, Yager [19] have proposed fuzzy bags, and its theory and applications have been studied by several researchers [3, 4, 15, 16, 17, 18, 20]. In particular, the author [9, 10, 12] has redefined basic operations of the union and intersection for fuzzy bags. The definition anew by the authors is necessary in order to keep consistency between operations for ordinary fuzzy sets and those for fuzzy bags [12]. However, there are many problems that are left and should be studied.

One class of such problems is fundamental operations of bags including complementation and bag s-norms (and s-norms). Another class is bag relations that should have composition operations. In this paper we study these two problem classes. While studying these problems, we should generalize the traditional concept of bags, as described in this paper. We therefore introduce a generalization called R-bag which appears straightforward, but this generalization is necessary to consider complementation and bag t-norms. As a result, a duality theorem is derived using the complementation. Moreover a Minkwoski type bag s-norms are derived using a generating function.

Bag relations and their compositions for R-bags are then studied by introducing a max-s algebra and a max-t algebra. Here s and t stands for s-norm and t-norm.

Another generalization that includes R-bag and fuzzy bag is moreover mentioned and how the above consideration is extended to this generalized bag is suggested.

There are a number of propositions herein, but to prove propositions given in this paper is not difficult. Hence short notes on proofs are given in the appendix.

Although the purpose of this paper is to show theoretical properties of generalized bags, bags have many application possibilities, which are briefly mentioned in the conclusion.

2 Bags and R-Bags

The two terms of bags and multisets are used interchangeably, and we use *bags* throughout this paper.

2.1 Overview of crisp bags

Assume that the universal set $X = \{x_1, \ldots, x_n\}$ is finite for simplicity. A crisp bag M of $X = \{x_1, \ldots, x_n\}$ is characterized by a function $C_M(\cdot)$ (called count of M) whereby a natural number including zero $\boldsymbol{N} = \{0, 1, 2, \ldots\}$ corresponds to each $x \in X$ ($C_M : X \to \boldsymbol{N}$).

For a crisp bag, different expressions such as

$$M = \{k_1/x_1, \ldots, k_n/x_n\}$$

and

$$M = \{\overbrace{x_1, \ldots, x_1}^{k_1}, \ldots, \overbrace{x_n, \ldots, x_n}^{k_n}\}$$

are used. In such a way, an element of X may appear more than once in a bag. In the above example x_i appears k_i times in M ($i = 1, \ldots, n$).

Consider an example where $X = \{a, b, c, d\}$ and $C_M(a) = 2$, $C_M(b) = 1$, $C_M(c) = 3$, $C_M(d) = 0$. In other words, $M = \{a, a, b, c, c, c\}$. This means that a, b, c, and d are included 2, 1, 3, and 0 times, respectively, in M. We can write $M = \{2/a, 1/b, 3/c\}$ by ignoring an element of zero occurrence.

Basic relations and operations for crisp bags are as follows.

1. (inclusion):
   $M \subseteq N \Leftrightarrow C_M(x) \leq C_N(x), \quad \forall x \in X.$
2. (equality):
   $M = N \Leftrightarrow C_M(x) = C_N(x), \quad \forall x \in X.$
3. (union):
   $C_{M \cup N}(x) = \max\{C_M(x), C_N(x)\}.$
4. (intersection):
   $C_{M \cap N}(x) = \min\{C_M(x), C_N(x)\}.$
5. (addition or sum):
   $C_{M \oplus N}(x) = C_M(x) + C_N(x).$
6. (subtraction):
   $C_{M \ominus N}(x) = \max\{0, C_M(x) - C_N(x)\}.$
7. (scalar multiplication):
   $C_{\alpha M} = \alpha C_M(x)$, where α is a positive integer.
8. (Cartesian product):
   Let P is a bag of Y.
   $C_{M \times P}(x, y) = C_M(x) C_P(y).$

ISBN: 978-989-95079-6-8

We sometimes use $\vee$ and $\wedge$ for $\max$ and $\min$, respectively. We also note that the relations and operations are similar to those for fuzzy sets. However, bags have the addition which fuzzy sets do not have, and the Cartesian product for bags is different from that for fuzzy sets. Moreover, a complementation is not defined in general.

2.2 R-bags

An immediate generalization of crisp bag is to assume that $C_M(\cdot)$ has a nonnegative real-value instead of a natural number. For example, we admit $C_M(a) = 3.14$ or $C_N(a) = \pi$. Moreover we admit the value of infinity $+\infty$ for a technical reason. We thus assume $C_M : X \to [0, +\infty]$.

The above relations and operations 1–8 are used without any change except that α in the scalar multiplication can now be a positive real constant. For simplicity, a real-valued bag is called an R-bag here and the collection of all R-bags of X is denoted by $\mathcal{RB}(X)$.

Complementation of R-bags

A function $\mathcal{N} : [0, +\infty] \to [0, +\infty]$ with the next properties is used to define a complementation operation:

(i) $\mathcal{N}(0) = +\infty, \quad \mathcal{N}(+\infty) = 0.$

(ii) $\mathcal{N}(x)$ is strictly monotonically decreasing on $(0, +\infty)$.

A typical example is $\mathcal{N}(x) = const/x$ with $const > 0$.

Using this function, a natural operation for the complement is:

9.(complement):
$C_{\bar{M}}(x) = \mathcal{N}(C_M(x))$.

This operation justifies the generalization into R-bags. That is, even when we start crisp bags, the result of complementation is generally real-valued.

We immediately have the next two propositions of which the proof is easy and omitted.

Proposition 1 *For arbitrary $M, N \in \mathcal{RB}(X)$, the following properties hold:*

$$\overline{(\overline{M})} = M \tag{1}$$

$$\overline{M \cup N} = \bar{M} \cap \bar{N}, \qquad \overline{M \cap N} = \bar{M} \cup \bar{N}. \tag{2}$$

Proposition 2 *If we introduce an empty bag $\emptyset$ and the maximum bag* **Infinity** *in $\mathcal{RB}$ by*

$$C_{\emptyset}(x) = 0, \quad \forall x \in X, \tag{3}$$

$$C_{\mathbf{Infinity}}(x) = +\infty, \quad \forall x \in X. \tag{4}$$

Then we have

$$\bar{\emptyset} = \mathbf{Infinity}, \quad \overline{\mathbf{Infinity}} = \emptyset. \tag{5}$$

s-norms and t-norms of R-bags

There have been studies on t-norms for crisp bags [7, 1], but generalization into R-bags admits a broader class of s-norms and t-norms. For this purpose we introduce two functions $t(a, b)$ and $s(a, b)$ as those in fuzzy sets, but the definitions are different.

Definition 1 *Two functions $t : [0, +\infty] \times [0, +\infty] \to [0, +\infty]$ and $s : [0, +\infty] \times [0, +\infty] \to [0, +\infty]$ having the following properties (I)–(IV) are called a t-norm and an s-norm for bags, respectively. An s-norm is also called a t-conorm for bags.*

(I)[monotonicity] For $a \le c$, $b \le d$,

$$t(a, b) \le t(c, d),$$
$$s(a, b) \le s(c, d).$$

(II)[symmetry]

$$t(a, b) = t(c, d), \quad s(a, b) = s(b, a).$$

(III)[associativity]

$$t(t(a, b), c) = t(a, t(b, c)),$$
$$s(s(a, b), c) = s(a, s(b, c)).$$

(IV)[boundary condition]

$$t(0, 0) = 0, \quad t(a, +\infty) = t(+\infty, a) = a,$$
$$s(+\infty, +\infty) = +\infty, \quad s(a, 0) = s(0, a) = a.$$

A purpose to introduce such norms for bags is to generalize the intersection and union operations. First we note that $s(a, b) = a+b$ and $s(a, b) = \max\{a, b\}$ satisfy the above conditions. The norm $t(a, b) = \min\{a, b\}$ also satisfies (I)–(IV). Thus, s-norms and t-norms can represent the three operations of addition, union, and intersection.

We moreover introduce a generating function $g(x)$ for s-norm.

Definition 2 *A function $g : [0, +\infty] \to [0, +\infty]$ is called a generating function for s-norm if it satisfies the next (i)–(iii):*
(i) it is strictly monotonically increasing,
(ii) $g(0) = 0, \quad g(+\infty) = +\infty$,
(iii) $g(x + y) \ge g(x) + g(y), \quad \forall x, y \in [0, +\infty]$.
We have

Proposition 3 *Let*

$$s(a, b) = g^{-1}(g(a) + g(b)). \tag{6}$$

Then $s(a, b)$ is an s-norm.

An example of the generation function is

$$g(x) = x^p \quad (p \ge 1). \tag{7}$$

Moreover, note the following.

Proposition 4 *Let $s(a, b)$ is an s-norm and $\mathcal{N}$ is the complementation operator. Then*

$$t(a, b) = \mathcal{N}(s(\mathcal{N}(a), \mathcal{N}(b))) \tag{8}$$

is a t-norm. Suppose $t(a, b)$ is a t-norm, then

$$s(a, b) = \mathcal{N}(t(\mathcal{N}(a), \mathcal{N}(b))) \tag{9}$$

is an s-norm.

If a pair of t-norm and s-norm has the above property, we say (s, t) has the *duality* of norm and conorm. The duality has the next property.

ISBN: 978-989-95079-6-8

Proposition 5 *Suppose $s_0(a,b)$ is an s-norm and $t_0(a,b)$ is derived from $s_0(a,b)$ by the operation (8). Let*

$$s(a,b) = \mathcal{N}(t_0(\mathcal{N}(a),\mathcal{N}(b)))$$

Then $s(a,b) = s_0(a,b)$.

Suppose also that $t_0(a,b)$ is a t-norm and $s_0(a,b)$ is derived from $t_0(a,b)$ by the operation (8). Let

$$t(a,b) = \mathcal{N}(s_0(\mathcal{N}(a),\mathcal{N}(b)))$$

Then $t(a,b) = t_0(a,b)$. In summary, applying (8) and (9) (resp. (9) and (8)) successively, we have the original norm.

Note that s-norm and t-norm are applied to define bag operations $M\mathcal{S}N$ and $M\mathcal{T}N$, respectively, for $M, N \in \mathcal{RB}(X)$.

$$C_{M\mathcal{S}N}(x) = s(C_M(x), C_N(x)). \tag{10}$$
$$C_{M\mathcal{T}N}(x) = t(C_M(x), C_N(x)). \tag{11}$$

Let us consider examples of s-norms and t-norms.

Example 1 The standard operators

$$s(a,b) = \max\{a,b\} \tag{12}$$
$$t(a,b) = \min\{a,b\} \tag{13}$$

are an s-norm and a t-norm, respectively. This pair has the duality stated in Proposition 5. Note, however, that s-norm (12) does not have a generating function that satisfies (6).

Second example uses the generating function.

Example 2 Let $g(x)$ be given by (7). Then we have

$$s(a,b) = (a^p + b^p)^{\frac{1}{p}}, \tag{14}$$
$$t(a,b) = (a^{-p} + b^{-p})^{-\frac{1}{p}}. \tag{15}$$

are an s-norm and a t-norm, respectively. This pair has the duality stated in Proposition 5 when $\mathcal{N} = const/x$ is used.

The second example has interesting properties. First, $s(a,b) = a+b$ is a particular case of (14) for $p = 1$. Moreover $s(a,b) = \max\{a,b\}$ and $t(a,b) = \min\{a,b\}$ are obtained from (14) and (15) when $p \to +\infty$.

New findings are included in the above example. First, the dual t-norm derived from $s(a,b) = a + b$ is the harmonic mean:

$$t(a,b) = (a^{-1} + b^{-1})^{-1} = \frac{ab}{a+b}.$$

(Note that the arithmetic mean $\frac{a+b}{2}$ and the geometric mean $\sqrt{ab}$ are not t-norms). Second, $s(a,b) = (a^2 + b^2)^{\frac{1}{2}}$ with $p = 2$ which is similar to the Euclidean norm is an s-norm, and the dual t-norm is

$$t(a,b) = \frac{1}{\sqrt{\frac{1}{a^2} + \frac{1}{b^2}}}.$$

2.3 R-bags and fuzzy sets

Let us study what relations exist between fuzzy sets and R-bags. For this purpose we introduce another notation: the collection of all fuzzy sets of X is denoted by $\mathcal{FS}(X)$.

We introduce another function $h\colon [0,+\infty] \to [0,1]$ that is strictly monotonically increasing in $[0,+\infty)$ and

$$h(0) = 0, \quad h(+\infty) = 1.$$

An example of such a function is

$$h_0(x) = 1 - \exp(-K_0 x), \quad x \in [0,+\infty)$$

with constant $K_0 > 0$ and with the additional definition $h_0(+\infty) = 1$. Another example is

$$h_1(x) = 1 - \frac{K_1}{K_1 + K_2 x}, \quad x \in [0,+\infty)$$

with positive constants K_1, K_2 and with $h_1(+\infty) = 1$.

Since such a function has an inverse function $h^{-1}(y)$, we can define one to one correspondence between a fuzzy set and an R-bag:

$$\mu_{h(M)}(x) = h(C_M(x)). \tag{16}$$

For $M \in \mathcal{RB}(X)$, $h(M) \in \mathcal{FS}(X)$, and $h^{-1}(h(M)) = M$. Note that $h^{-1}(y)$ is also a strictly monotonically increasing.

We also have

$$h(\emptyset) = \emptyset, \quad h(\mathbf{Infinity}) = X;$$
$$h^{-1}(\emptyset) = \emptyset, \quad h^{-1}(X) = \mathbf{Infinity}.$$

Using these functions, we can generate operations for R-bags from those for fuzzy sets.

Proposition 6 *Let*

$$\mathcal{N}(x) = h^{-1}(1 - h(x)). \tag{17}$$

Then $\mathcal{N}(x)$ satisfies the properties of complementation operation. Hence for arbitrary $M \in \mathcal{RB}(X)$, $\bar{M}$ derived by

$$C_{\bar{M}}(x) = \mathcal{N}(C_M(x))$$

using (17) is a complement of M.

Note that this operation is derived from the complementation of fuzzy sets.

For the s-norm and t-norm, we have similar results. Let $s_F(a,b)$ and $t_F(a,b)$ are respectively an s-norm and a t-norm for *fuzzy sets*. We have the following.

Proposition 7 *For $a, b \in [0,+\infty]$, let*

$$s(a,b) = h^{-1}(s_F(h(a), h(b))) \tag{18}$$
$$t(a,b) = h^{-1}(t_F(h(a), h(b))) \tag{19}$$

Then $s(a,b)$ and $t(a,b)$ are an s-norm and a t-norm for R-bags.

we also have

Proposition 8 *Suppose $s(a,b)$ and $t(a,b)$ are an s-norm and a t-norm for R-bags. For $a, b \in [0,1]$, let*

$$s_F(a,b) = h(s(h^{-1}(a), h^{-1}(b))) \tag{20}$$
$$t_F(a,b) = h(t(h^{-1}(a), h^{-1}(b))) \tag{21}$$

Then $s_F(a,b)$ and $t_F(a,b)$ are an s-norm and a t-norm for fuzzy sets.

3 Bag Relations

What we consider in this section is a bag-relations and their composition. As a result we have max-s and max-t algebras for R-bags. The discussion in this section generalizes the max-plus algebra [2].

ISBN: 978-989-95079-6-8

3.1 max-s and max-t algebra

Let us use a particular notation of $\boxplus$ and $\boxdot$ for

$$a \boxplus b = \max\{a, b\}, \quad a \boxdot b = s(a, b) \tag{22}$$

where $s(a, b)$ is an s-norm for R-bags. We call this algebra as *max-s algebra*.

It is easy to see that the following properties hold.

$$a \boxplus b = b \boxplus a, \tag{23}$$
$$a \boxplus (b \boxplus c) = (a \boxplus b) \boxplus c, \tag{24}$$
$$a \boxplus 0 = a, \tag{25}$$
$$a \boxdot b = b \boxdot a, \tag{26}$$
$$a \boxdot (b \boxdot c) = (a \boxdot b) \boxdot c, \tag{27}$$
$$a \boxdot 0 = a. \tag{28}$$

Alternatively, we can define $\boxplus$ and $\boxdot$ for

$$a \boxplus b = \max\{a, b\}, \quad a \boxdot b = t(a, b) \tag{29}$$

where $t(a, b)$ is a t-norm for R-bags. We call this algebra as *max-t algebra*. We see that (23)–(27) hold, while (28) should be replaced by

$$a \boxdot +\infty = a. \tag{30}$$

We moreover note the next lemma.

Lemma 1 *Let a, b, c be real numbers. Then*

$$a \boxdot (b \boxplus c) = (a \boxdot b) \boxplus (a \boxdot c). \tag{31}$$

Thus the commutative, associative, and distributive properties hold for the max-s algebra. Noting Example 2, we observe that the max-t algebra includes the max-min algebra and the max-s algebra includes the max-plus algebra as particular cases.

3.2 Bag relations

We now define R-bag relations. The definition is simple. An R-bag relation R on $X \times Y$ is an R-bag R of $X \times Y$. The count function is denoted by $R(x, y)$ instead of $C_R(x, y)$ for simplicity. Sometimes an R-bag relation is also called bag relation for simplicity.

The reason why we call such a bag $R(x, y)$ a bag relation is that we can define composition operation.

Let X, Y, Z be three universes. Assume R is a bag relation of $X \times Y$ and S is a bag relation of $Y \times Z$. Then the max-s composition $R \circ S$ is defined as follows.

$$(R \circ S)(x, z) = \boxplus_{y \in Y}\{R(x, y) \boxdot S(y, z)\} \tag{32}$$

Note that

$$\boxplus_{y \in \{a_1, \ldots, a_L\}} = a_1 \boxplus a_2 \boxplus \cdots \boxplus a_L.$$

The max-t composition is defined by the same equation (32) except that $\boxdot$ uses a t-norm.

Note also that the addition is straightforward

$$(R_1 \boxplus R_2)(x, y) = R_1(x, y) \boxplus R_2(x, y), \tag{33}$$

for bag relations on $X \times Y$.

We have the following proposition.

Proposition 9 *The composition satisfies the associative property*

$$(R \circ S) \circ T = R \circ (S \circ T). \tag{34}$$

and the distributive property

$$(R_1 \boxplus R_2) \circ S = (R_1 \circ S) \boxplus (R_2 \circ S), \tag{35}$$
$$R \circ (S_1 \boxplus S_2) = (R \circ S_1) \boxplus (R \circ S_2). \tag{36}$$

We introduce the unit relations for the max-s and max-t compositions. Define O_X and Ω_X on $X \times X$:

$$O_X(x, y) = 0, \quad \forall x, y \in X, \tag{37}$$
$$\Omega_X(x, y) = +\infty, \quad \forall x, y \in X. \tag{38}$$

We define O_{XY} and Ω_{XY} on $X \times Y$ in the same way:

$$O_{XY}(x, y) = 0, \quad \forall (x, y) \in X \times Y, \tag{39}$$
$$\Omega_{XY}(x, y) = +\infty, \quad \forall (x, y) \in X \times Y. \tag{40}$$

Frequently we omit the subscripts like O and Ω when we have no ambiguity.

We then have

Proposition 10 *Assume that the max-s algebra is used. For arbitrary bag relation R on $X \times Y$,*

$$R \boxplus O = O \boxplus R = R, \tag{41}$$
$$R \circ O = O \circ R = R. \tag{42}$$

Proposition 11 *Assume that the max-t algebra is used. For arbitrary bag relation R on $X \times Y$,*

$$R \boxplus O = O \boxplus R = R, \tag{43}$$
$$R \circ \Omega = \Omega \circ R = R. \tag{44}$$

4 A Note on Further Generalization

The above generalization to R-bags does not include fuzzy bags [19, 10]. It appears that R-bags and fuzzy bags are inconsistent. However, there is further generalization discussed in [12] which uses a closed region in $\boldsymbol{R}^2$ as the count function for generalized bags. In this section we briefly note how the above discussion is extended into generalized bags which encompasses R-bags and fuzzy bags.

A generalized bag A has a count function as a region

$$\mathbf{N}(x, A) = C_A(x), \quad \forall x \in X \tag{45}$$

where $\mathbf{N}(x, A)$ is a closed set of $[0, +\infty)^2$. Let $\nu(y, z; x, A)$ is the characteristic function of the region $\mathbf{N}(x, A)$:

$$\nu(y, z; x, A) = \begin{cases} 1, & (y, z) \in \mathbf{N}(x, A), \\ 0, & (y, z) \notin \mathbf{N}(x, A). \end{cases}$$

This region should have the following properties:

(N1) For any $y \in [0, +\infty)$,

$$\eta(y) = \{z \in [0, +\infty) : \nu(y, z; x, A) = 1\}$$

is either an empty set or a closed interval $[0, \mu(y)]$ where $\mu(y) = \max\{y' : y' \in \eta(y)\}$.

ISBN: 978-989-95079-6-8

(N2) For all $y \in [0, +\infty)$, $\mu(y)$ is a monotonically decreasing function and $0 \leq \mu(y) \leq 1$.

(N3) For any $z \in [0, +\infty)$,

$$\theta(z) = \{y \in [0, +\infty) : \nu(y, z; x, A) = 1\}$$

is either an empty set or a closed interval $[0, \zeta(z)]$ where $\zeta(z) = \max\{z' : z' \in \theta(z)\}$.

(N4) For all $z \in [0, +\infty)$, $\zeta(z)$ is a monotonically decreasing function and $0 \leq \zeta(z) < +\infty$.

Roughly, a generalized bag is characterized by the pair $(\mu(y), \zeta(y))$ for each $x \in X$.

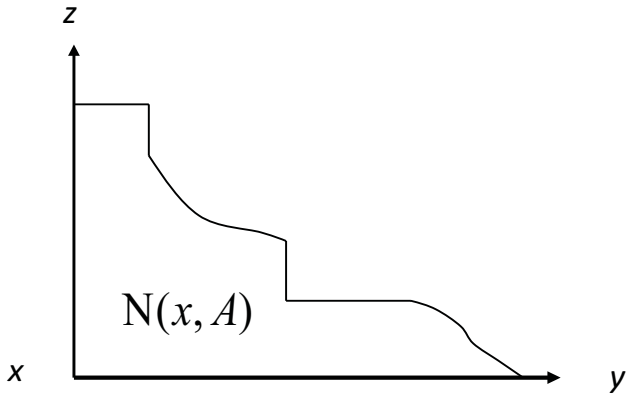


Figure 1: An illustration of $\mathbf{N}(x, A)$ as the count function for a generalized bag.

Two cuts which relate the generalized bag to R-bags and fuzzy sets are defined:

α-cut: Let $\alpha \in (0, 1]$. For the generalized bag A, $[A]_\alpha$ is an R-bag given by

$$C_{[A]_\alpha}(x) = \max\{z : \nu(\alpha, z; x, A) = 1\}.$$

ℓ-cut: Let $\ell \in (0, +\infty)$. For the generalized bag A, $\langle A \rangle_\ell$ is a fuzzy set given by

$$\mu_{\langle A \rangle_\ell}(x) = \max\{y : \nu(y, \ell; x, A) = 1\}.$$

It can be proved that the generalized bags are smallest generalization of fuzzy bags and R-bags [12]. Moreover, s-norms and t-norms can be used for the generalized bags using the α-cuts, while the complementation operation cannot be used for the generalized bags. Bag relations and their compositions can also be used for the generalized bags. The detailed consideration is too lengthy to give here and will be discussed elsewhere.

5 Conclusion

We have shown the operations of complementation and s-norm and t-norm for real-valued bags. Moreover R-bag relations and the max-s (or max-t) compositions are studied. Further generalization has moreover been proposed of which the details will be discussed in near future.

As seen herein, the theory of bags should be studied more deeply, since it has more ample mathematical structure than expected until now. More theoretical studies are thus needed. Relations between rough sets [14] and bags have been discussed [12], but there is much room for further studies. Application of bags such as those in information retrieval [8] and information mining [11, 13] will also be important.

Acknowledgment

This study has partly been supported by the Grant-in-Aid for Scientific Research, Japan Society for the Promotion of Science, No.19650052.

Appendix: Notes on Proofs

Since detailed proofs are too lengthy, we give short notes on how to prove the propositions given here. The proofs are mostly straightforward.

Notes on proofs of Propositions 1 and 2

These propositions are immediately proved by straightforward calculation using $\mathcal{N}$. The details are hence omitted.

Notes on proofs of Propositions 3 and 4

It is enough to check the conditions (I)–(IV) are satisfied.

Notes on proof of Proposition 5

$$\begin{aligned} s(a, b) &= \mathcal{N}(t_0(\mathcal{N}(a), \mathcal{N}(b))) \\ &= \mathcal{N}(\mathcal{N}(s_0(\mathcal{N}(\mathcal{N}(a)), \mathcal{N}(\mathcal{N}(b))))) \\ &= s_0(a, b). \end{aligned}$$

Notes on proof of Proposition 6

By a simple calculation, it is seen that $\mathcal{N}$ by (17) satisfies the conditions for the complementation.

Notes on proofs of Propositions 7 and 8

The conditions for s-norms and t-norms for fuzzy sets should be noted. Mostly it is easy to check that the conditions are satisfied, since the functions h and h^{-1} are monotone. The boundary conditions are satisfied from $h(0) = 0$ and $h(+\infty) = 1$.

Notes on proof of Lemma 1

Let $b \geq c$, then the left hand side is $s(a, b)$. The right hand side is

$$\max\{s(a, b), s(a, c)\} = s(a, b).$$

The case of $b < c$ is calculated likewise and we have the desired equality.

When we use a max-t algebra, we can prove the lemma in a similar way, since the monotone property is common to s-norm and t-norm. We omit the detail.

Notes on proof of Proposition 9

The detailed proof is lengthy, but the way to prove this proposition is just the same as the proof of the matrix properties for ordinary algebra, since those properties for ordinary algebra uses the commutative, associative, and distributive properties which are valid for the present algebra, since (23)–(28) (or (30)) and (31) are valid.

Notes on proofs of Propositions 10 and 11

Omitted as they are straightforward.

ISBN: 978-989-95079-6-8

References

[1] J. Casasnovas, G. Mayor, Discrete t-norms and operations on extended multisets, *Fuzzy Ses and Systems*, Vol.159, pp. 1165–1177, 2008.

[2] B. Heidergott, G.J. Olsder, J. van der Woude, *Max Plus at Work*, Princeton University Press, Princeton, NJ, 2006.

[3] B. Li, W. Peizhang, and L. Xihui, Fuzzy bags with set-valued statistics, *Comput. Math. Applic.*, Vol. 15, pp. 811–818, 1988.

[4] B. Li, Fuzzy bags and applications, *Fuzzy Sets and Systems*, Vol. 34, pp. 61–71, 1990.

[5] D. E. Knuth, *The Art of Computer Programming, Vol. 2 / Seminumerical Algorithms*, Addison-Wesley, Reading, Massachusetts, 1969.

[6] Z. Manna and R. Waldinger, *The Logical Basis for Computer Programming, Vol. 1: Deductive Reasoning*, Addison-Wesley, Reading, Massachusetts, 1985.

[7] G. Mayor, J. Torrens, Triangular norms on discrete settings, in: E.P. Klement, R. Mesiar, eds., *Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms*, Elsevier, Amsterdam, 2005, pp. 189–230.

[8] S. Miyamoto, *Fuzzy Sets in Information Retrieval and Cluster Analysis*, Kluwer, Dordrecht, 1990.

[9] S. Miyamoto, Basic operations of fuzzy multisets, *J. of Japan Society for Fuzzy Theory and Systems*, Vol. 8, No. 4, pp. 639–645, 1996 (in Japanese).

[10] S. Miyamoto, Fuzzy multisets with infinite collections of memberships, *Proc. of the 7th International Fuzzy Systems Association World Congress (IFSA'97)*, June 25-30, 1997, Prague, Czech, Vol.1, pp.61-66, 1997.

[11] S. Miyamoto, Information clustering based on fuzzy multisets, *Information Processing and Management*, Vol.39, No.2, pp. 195–213, 2003.

[12] S. Miyamoto, Generalizations of multisets and rough approximations, *International Journal of Intelligent Systems*, Vol.19, No.7, pp. 639–652, 2004.

[13] S. Miyamoto, Y. Kawasaki, Text Analysis Based on Bag Neighborhood Model and Application to Incident Reports, Da Ruan et al., eds., Computational Intelligence in Decision and Control, Proc. of the 8th Intern. FLINS Conf., Madrid, Spain, 21-24 Sept. 2008, pp.1015-1020.

[14] Z. Pawlak, *Rough Sets*, Kluwer, Dordrecht, 1991.

[15] A. Ramer and C.-C. Wang, Fuzzy multisets, *Proc. of 1996 Asian Fuzzy Systems Symposium*, Dec. 11-14, 1996, Kenting, Taiwan, pp. 429–434.

[16] A. Rebai, Canonical fuzzy bags and bag fuzzy measures as a basis for MADM with mixed non cardinal data, *European J. of Operational Res.*, Vol. 78, pp. 34–48, 1994.

[17] A. Rebai, J.-M. Martel, A fuzzy bag approach to choosing the "best" multiattributed potential actions in a multiple judgment and non cardinal data context, *Fuzzy Sets and Systems*, Vol. 87, pp. 159–166, 1997.

[18] D. Rocacher, On fuzzy bags and their application to flexible querying, *Fuzzy Sets and Systems*, Vol. 140(1), pp.93-110, 2003.

[19] R. R. Yager, On the theory of bags, *Int. J. General Systems*, Vol. 13, pp. 23–37, 1986.

[20] R. R. Yager, Cardinality of fuzzy sets via bags, *Mathl. Modelling*, Vol. 9, No. 6, pp. 441–446, 1987.

ISBN: 978-989-95079-6-8

An Explicit Mapping for Kernel Data Analysis and Application to Text Analysis

Sadaaki Miyamoto[1] Yuichi Kawasaki[2] Keisuke Sawazaki[2]

1.Department of Risk Engineering, University of Tsukuba
Tsukuba, Japan
2.Graduate School of Systems and Information Engineering, University of Tsukuba
Tsukuba, Japan
Email: miyamoto@risk.tuskuba.ac.jp, {kawasaki,sawazaki}@soft.risk.tsukuba.ac.jp

Abstract— *Kernel data analysis is now becoming standard in every application of data analysis and mining. Kernels are used to represent a mapping into a high-dimensional feature space, where an explicit form of the mapping is unknown. Contrary to this common understanding, we introduce an explicit mapping which we consider standard. The reason why we use this mapping is as follows. (1) the use of this mapping does not lose any fundamental information in kernel data analysis and we have the same formulas in every kernel methods. (2) Usually the derivation becomes simpler by using this mapping. (3) New applications of the kernel methods become possible using this mapping. As an application we consider an example of text mining where we use fuzzy c-means clustering and cluster centers in the high-dimensional space and visualize the centers using kernel principal component analysis.*

Keywords— Kernel data analysis, fuzzy clustering, explicit mapping, text mining

1 Introduction

Kernel functions [9] which are not very new but was noted in support vector machines[10] and now is a well-known technique and becoming standard in many applications of data analysis and data mining. An important point in the use of kernel functions is that although we consider a mapping from a data space into a high-dimensional feature space, we need not to know its explicit form but we should know the inner product of the feature space. Generally, the feature space is not uniquely determined. Accordingly every formula in data analysis using kernel functions should be described in terms of the inner product.

Although kernel functions are really useful, but the derivation is sometimes complicated when original formulas should be rewritten by the inner product forms. A typical example is kernel fuzzy c-means clustering [3] and kernel SOM [4] in which the cluster centers in a feature space should be eliminated using inner product forms.

Here is a question: can we have a useful and explicit mapping and explicit representation of a high-dimensional feature space? The answer is YES, and we will show the mapping that is simple enough and useful in the sense that it leads to the same formulas when transformed into the inner product forms. To summarize, this explicit mapping and associated feature space have all information that is used in kernel functions for data analysis.

Another question arises: what is useful in this explicit mapping? We will show a real application of an example of text analysis. A result of kernel fuzzy c-means clustering and cluster centers will be shown using kernel principal component analysis.

Proofs of propositions are mostly omitted to save the space but brief notes are given as appendix.

2 Kernel Functions and an Explicit Mapping

2.1 Preliminary consideration

Although we consider kernel fuzzy c-means (FCM) clustering [3] and kernel principal component analysis (KPCA)[9] later, we describe FCM and omit KPCA for simplicity.

Assume that a set of data $X = \{x_1, \ldots, x_n\} \subset \boldsymbol{R}^p$ is given. Each data unit is also called an object or an individual and it is a point in the p-dimensional real space $x_k = (x_k^1, \ldots, x_k^p)^T \in \boldsymbol{R}^p$. We consider a mapping into a high-dimensional feature space $\Phi\colon \boldsymbol{R}^p \to H$ and associated kernel function

$$\boldsymbol{K}(x, y) = \langle \Phi(x), \Phi(y) \rangle \tag{1}$$

where $\langle \cdot, \cdot \rangle$ is an inner product of H. We also assume $\| \cdot \|_H$ is a norm of H. Thus H is an inner product space. In this section suppose we do not know function $\Phi(\cdot)$ explicitly but we know $\boldsymbol{K}(x, y)$. Specifically, the Gaussian kernel is used frequently:

$$\boldsymbol{K}(x, y) = \exp(-\lambda \|x - y\|^2) \tag{2}$$

where $\lambda > 0$ and $\|x\|$ is the norm of $\boldsymbol{R}^p$.

An objective function of fuzzy c-means using the feature space H is the following.

$$J_H(U, W) = \sum_{i=1}^{c} \sum_{k=1}^{n} (u_{ki})^m \|\Phi(x_k) - W_i\|_H^2, \quad (m > 1) \tag{3}$$

where $U = (u_{ki})$ is $n \times c$ matrix representing membership of x_k to cluster i. U has the next constraint when it is optimized.

$$M = \{U = (u_{ki}) : \sum_{i=1}^{c} u_{ki} = 1,\ u_{ki} \geq 0, \forall k, i\}. \tag{4}$$

Moreover $W = (W_1, \ldots, W_c)$ shows cluster centers.

The iterative algorithm of fuzzy c-means clustering is basically an alternative minimization of $J_H(U, W)$ with respect to

ISBN: 978-989-95079-6-8

U and W until convergence. We have the next solutions.

$$u_{ki} = \left[\sum_{j=1}^{c} \left(\frac{D(x_k, W_i)}{D(x_k, W_j)} \right)^{\frac{1}{m-1}} \right]^{-1}, \tag{5}$$

$$W_i = \frac{\sum_{k=1}^{n} (u_{ki})^m \Phi(x_k)}{\sum_{k=1}^{n} (u_{ki})^m} \tag{6}$$

where we put

$$D(x_k, W_i) = \|\Phi(x_k) - W_i\|_H^2. \tag{7}$$

Equations (5) and (6) are repeated until convergence, but since $\Phi(x_k)$ is unknown, we should use another formula for kernel fuzzy c-means clustering.

The formula is derived by eliminating (6) from iteration, i.e., we substitute (6) into (7) to have an updating formula for $D(x_k, W_i)$[3]:

$$D(x_k, W_i) = \boldsymbol{K}(x_k, x_k) - \frac{2}{\sum_{k=1}^{n} (u_{ki})^m} \sum_{j=1}^{n} (u_{ji})^m \boldsymbol{K}(x_j, x_k) + \frac{1}{\{\sum_{k=1}^{n} (u_{ki})^m\}^2} \sum_{j=1}^{n} \sum_{\ell=1}^{n} (u_{ji} u_{\ell i})^m \boldsymbol{K}(x_j, x_\ell). \tag{8}$$

We hence repeat (5) and (8) until convergence when kernel fuzzy c-means clustering should be used.

2.2 An explicit mapping and its properties

We sometimes have applications in which we wish to have explicit cluster centers. Even if we cannot show them in a low-dimensional data space, there are ways to have approximate representations or visualizations, one of which is to show kernel principal components and another is to use SOM mapping.

To this end we use the following explicit mapping that is simple enough but seems unnoticed.

$$\Phi(x_k) = e_k \quad (k = 1, 2, \ldots, n) \tag{9}$$

where $H = \boldsymbol{R}^n$ and e_k is the k-th unit vector that has unity as ith component and all other components are zero. Note that $\Phi\colon X \to \boldsymbol{R}^n$, i.e., $\Phi(\cdot)$ is not defined on $\boldsymbol{R}^p$ but is limited to the finite set X. Moreover we assume that the inner product of $\boldsymbol{R}^n$ is

$$\langle e_k, e_\ell \rangle = \boldsymbol{K}(x_k, x_\ell) \tag{10}$$

instead of the usual $\langle e_k, e_\ell \rangle = \delta_{k\ell}$.

We have the next proposition.

Proposition 1. *If kernel $\boldsymbol{K}(x, y)$ is positive definite, $\langle e_k, e_\ell \rangle$ defined by (10) satisfy the axioms of the inner product of $\boldsymbol{R}^n$, that is, $\boldsymbol{R}^n$ with (10) is an inner product space.*

We now apply this mapping to fuzzy c-means clustering. It is sufficient to show the optimal solution of W_i. We then have the next proposition of which the proof is easy and omitted.

Proposition 2. *For all positive definite kernel $\boldsymbol{K}(x, y)$ and mapping Φ by (9), the cluster centers are the same and are given by*

$$W_i = \left(\frac{(u_{1i})^m}{\sum_{k=1}^{n} (u_{ki})^m}, \ldots, \frac{(u_{ni})^m}{\sum_{k=1}^{n} (u_{ki})^m} \right)^T, \quad i = 1, \ldots, c \tag{11}$$

The next proposition of which the proof is almost trivial shows an important result of the equivalence between the usual technique and the explicit mapping method.

Proposition 3. *Using the explicit mapping (11) to (7), we have (8).*

That is, (11) derived from the single mapping (9) has all necessary and sufficient information for kernel fuzzy c-means clustering.

As noted above, formulas in kernel principal component analysis are derived likewise. Since the derivation repeats what is described in textbooks [9], we omit the details. The mapping (9) is moreover useful for application to LVQ and SOM [2], where vectors for updating quantization vectors should be based on learning. The explicit mapping enables vector representations in $\boldsymbol{R}^n$, we can use every formulas in LVQ and SOM, while usual kernel methods should eliminate quantization vectors [6].

3 Analysis of Terms from a Set of Texts

We hereafter show an application in which mapping (9) has a natural interpretation. A typical example is analysis of terms from s set of texts. A well-known model for term analysis uses a vector-space model [8] which is also called a bag model. Mizutani and Miyamoto[6] generalize the model into fuzzy multiset model and use kernel based learning and clustering. Note that a term may occur many times in a text which we call here term occurrences, or simply occurrences. Term occurrences are handled as term frequencies or count in bags in the vector space model, or it is treated as fuzzy multisets in [6] when memberships are attached to each term occurrence. When c-means clustering is applied, cluster centers have implications as they are representatives of clusters.

In these models, however, a structure inside a text such as distances or topology between occurrences is not considered. Miyamoto and Kawasaki[5] propose a kernel-based model that handles fuzzy neighborhood. As a result, kernel-based data analysis can be applied but a drawback is that we cannot have cluster centers. We consider this model of fuzzy neighborhood here and uses the above explicit mapping and overcome the last drawback.

3.1 Fuzzy neighborhood model

Let us consider a model that consists of a quintuple:

$$< T, O, d, R, N >$$

where each element is as follows.

1. $T = \{t_1, \ldots, t_n\}$: a finite set of *terms* or *keywords*;
2. O and d: a metric space; O is a finite set of *term occurrences* where a metric d is defined;
3. R : a fuzzy relation of $T \times O$;
4. N : a fuzzy relation of $O \times O$ called *fuzzy neighborhood*.

ISBN: 978-989-95079-6-8

The relation $R(t,o)$ shows correspondence between a term and its occurrence. For simplicity we assume, for every $t \in T$, $R(t,\cdot) = \{o \in O : R(t,o) > 0\}$ is not empty. The most simple relation is

$$R(t,o) = \begin{cases} 1 & (o \text{ is an occurrence of } t), \\ 0 & (\text{otherwise}), \end{cases}$$

and this relation is assumed hereafter in this paper.

Moreover we impose the following conditions to N:

(i) $N(o,o) = 1$ for all $o \in O$;

(ii) $d(o,o') > d(o,o'') \Rightarrow N(o,o') \leq N(o,o'')$;

(iii) $N(o,o') \to 0$ as $d(o,o') \to \infty$;

(iv) $N(o,o') = N(o',o)$.

Our purpose is to define a natural inner product space on T using the structure of occurrence space. We define a proximity relation $p(t,t')$, $t,t' \in T$:

$$p(t,t') = \sum_{o\in O}\sum_{o'\in O} R(t,o)N(o,o')R(t',o'). \tag{12}$$

If the matrix $P = (p(t,t'))$ is positive-definite, we can define an inner product space. If P is positive definite, we can use a normalized relation

$$s(t,t') = \frac{p(t,t')}{\sqrt{p(t,t)p(t',t')}}. \tag{13}$$

which is also positive-definite.

Generally, the above equation (12) does not guarantee the positive definiteness. We hence consider conditions for the positive-definiteness. To obtain a general condition is not practically useful. We therefore describe a specific, but broad enough, application example.

3.2 A model of a text set

A set of texts where each term may occur many times is represented as a sequence. For example, let a term set is $T = \{a,b,c\}$. Then an example of O can be represented by

$$O = aabcbbcccabbccabab$$

or more precisely, we put suffixes:

$$O = a_1\, a_2\, b_1\, c_1\, b_2\, b_3\, c_2\, c_3\, c_4\, a_3\, b_4\, b_5\, c_5\, c_6\, a_4\, b_6\, a_5\, b_7 \tag{14}$$

in order to distinguish occurrences. Obviously, $R(a,\cdot) = \{a_1,\ldots,a_5\}$. What if two or more texts should be handled? We can connect those texts into a sequence using dummy terms. Hence a sequence is sufficient to represent a set of texts. A natural distance d on O is given by

$$d(o,o') = \{\text{the number of occurrences between } o \text{ and } o'\}+1 \tag{15}$$

Thus, $d(a_1,c_1) = 3$, $d(c_6,c_2) = 7$, *etc* in the above example.

Next, two specific examples of fuzzy neighborhood are considered.

Fuzzy neighborhood using a monotone function

Let us show and example of a fuzzy neighborhood. For this purpose we introduce a monotone-decreasing function $f\colon \boldsymbol{R} \to [0,1]$ such that

(I) $f(0) = 1$,

(II) $f(-x) = f(x)$,

(III) $\lim_{x\to\infty} f(x) = 0$.

We define

$$N(o,o') = f(d(o,o')). \tag{16}$$

We have

Proposition 4. *Fuzzy relation $N(o,o')$ defined by (16) satisfies the conditions (i)–(iv) for a fuzzy neighborhood.*

The next proposition is crucial to this model.

Proposition 5. *If function $f(x)$ is convex on $[0,\infty)$, then $P = (p(t,t'))$ is positive-definite. Namely, we can define an inner product space using P.*

Let us consider a simple example where a convex function

$$f(x) = \begin{cases} 1 - \frac{x}{3} & (0 \leq x \leq 3) \\ 0 & (x > 3) \end{cases}$$

is used and $f(-x) = f(x)$. For the example of (14), $N(a_1,b_1) = 1/3$, $N(a_2,b_1) = 2/3$, *etc*. we hence have

$$p(a,b) = \sum_{i=1}^{5}\sum_{j=1}^{7} N(a_i,b_j) = 4.$$

3.3 Fuzzy neighborhood using a hierarchical structure

Another class for fuzzy neighborhood uses a hierarchical structure of a text, that is, a text frequently has chapters, sections, subsections, and paragraphs. Hence the next distance can be induced using $0 < \alpha < \beta < \gamma$.

- If o and o' occurs in a same paragraph, $d(o,o') = \alpha$;
- else if o and o' occurs in a same subsection, $d(o,o') = \beta$;
- else if o and o' occurs in a same section, $d(o,o') = \gamma$;
- else $d(o,o') = \infty$.

Such a distance induced from a hierarchical classification satisfies

$$d(o,o'') \leq \max\{d(o,o'), d(o',o'')\}. \tag{17}$$

which is called ultra-metric property. It is also known that this property is equivalent to fuzzy equivalence that is also called fuzzy similarity. That is,

$$N(o,o') = f(d(o,o')) \tag{18}$$

is a fuzzy equivalence.

We have the next proposition.

Proposition 6. *If $d(o,o')$ satisfies (17) and $N(o,o')$ is defined by (18), then $P = (p(t,t'))$ is positive-definite. Namely, we can define an inner product space using this P.*

ISBN: 978-989-95079-6-8

Let us moreover consider a particular case of d when O consists of documents $D_1, \ldots, D_m$:

$$d(o, o') = \begin{cases} 0 & (o \text{ and } o' \text{ are in a same document}), \\ \infty & (o \text{ and } o' \text{ are in different documents}). \end{cases} \tag{19}$$

Suppose also that the frequency of occurrences of t in D_i is $F_i(t)$.

We have the next proposition.

Proposition 7. *If $d(o, o')$ is given by (19) and $N(o, o')$ is defined by (18). then*

$$p(t, t') = \sum_{i=1}^{m} F_i(t) F_i(t'), \tag{20}$$

which is used in the vector space model.

We thus note that the model in this section generalizes the vector space model.

Finally, we note that the above model can directly be applied to a document set $D = \{d_1, \ldots, d_m\}$ using

$$R_D(d, t) = \begin{cases} 1 & (\ t \text{ occurs in } d), \\ 0 & (\ t \text{ doe no not occur in } d). \end{cases}$$

Then we have a positive definite measure between two documents:

$$p_D(d, d') = \sum_{t, t' \in T} R_D(d, t) p(t, t') R_D(d', t'), \tag{21}$$

and a normalized measure

$$s_D(d, d') = \frac{p_D(d, d')}{\sqrt{p_D(d, d) p_D(d', d')}}, \tag{22}$$

which can be used as inner products of the document space.

4 Numerical Examples

We briefly show the results of kernel fuzzy c-means clustering (KFCM) with the kernel principal components (KPCA) applied to a set of documents extracted from a Japanese newspaper articles. Sixty articles from ASAHI Shinbun (ASAHI News) have been used. They are categorized by the publisher into three categories of 'Economics', 'Politics', and 'Social Affairs'; each category has 20 articles. Each of these categories are subdivided into subcategories of 'economic statistics', 'finance', and so on. Terms occurred more than twice were used and the number of terms is 556. The above measure (22) has been used throughout. The method of fuzzy c-means with $m = 1.2$ and with $c = 3$ has been applied. The following two neighborhood functions have been used.

- First type uses a convex function $f(x) = 1.2^{-d(o,o')}$.
- Second type uses a fuzzy equivalence: If two term occurrences are the same term, $N(o, o') = 1$; if they are in a same paragraph, $N(o, o') = 0.9$; if they are in a same document, $N(o, o') = 0.7$; if they are in a same subcategory: $N(o, o') = 0.5$; if they are in a same category: $N(o, o') = 0.3$.

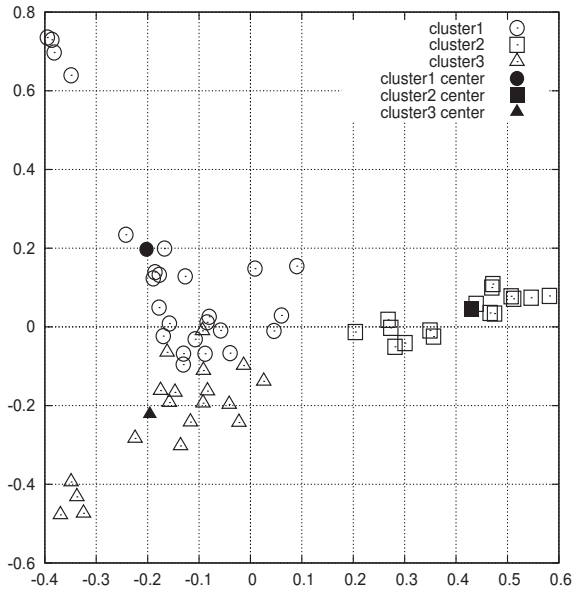


Figure 1: Two-dimensional display from KPCA with three clusters and cluster centers, where the first type of neighborhood function $f(x) = 1.2^{-d(o,o')}$ has been used.

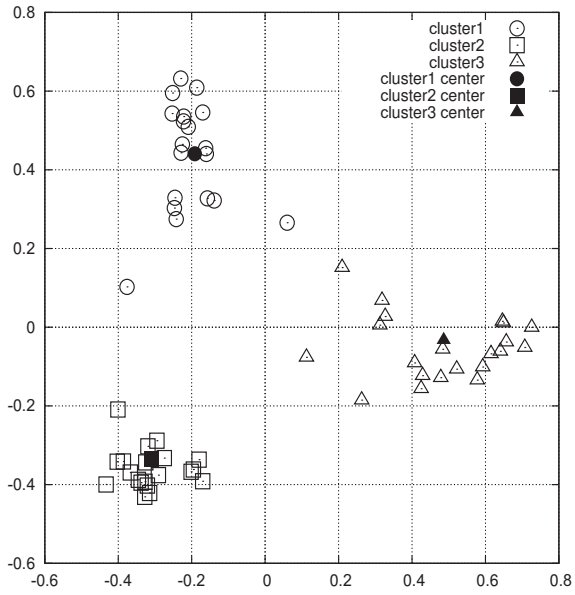


Figure 2: Two-dimensional display from KPCA with three clusters and cluster centers, where the second type of fuzzy equivalence has been used.

Figures 1 and 2 respectively show the two-dimensional figures of two-major axes from KPCA using the first and the second types of neighborhoods. The three symbols of white squares, triangles, and circles show the three clusters obtained from KFCM: fuzzy clusters have been made crisp by the maximum membership rule

$$x_k \to \textit{cluster } i \iff u_{ki} = \max_{1 \le j \le c} u_{kj}.$$

The black square, triangle, and circle are the cluster centers of the corresponding clusters. It seems that the second neigh-

ISBN: 978-989-95079-6-8

borhood of fuzzy equivalence divides more clearly the three clusters.

We have checked the correspondence of the obtained clusters with the actual classification provided from the publisher. The six correspondences have all been checked and the maximum correspondence percentages are given next. In the calculation 1000 trials with random initial values were used and the average numbers are shown below. That is, when we used the first type of neighborhood, the correct classification ratio was 73.5%, while the second type of fuzzy equivalence produced 97.9%.

We also tested a subclass of the above articles: 20 'Economics' and 20 'Politics' articles with the same conditions except that the number of clusters $c = 2$. The number of terms which occurred more than twice is 390. The results are shown as Figures 3 and 4. The separation was better than the previous case of three clusters. The correct classification ratio was 93.2% for the first type of neighborhood and 99.8 for the second neighborhood.

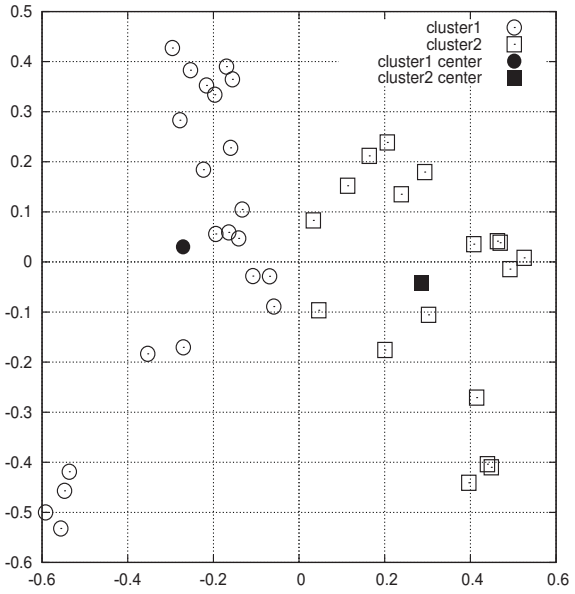


Figure 3: Two-dimensional display from KPCA with two clusters and cluster centers, where the first type of neighborhood function $f(x) = 1.2^{-d(o,o')}$ has been used.

5 Conclusion

We have proposed the use of an explicit mapping for kernel based data analysis. To summarize, we note the following advantages of the present method.

1. Using this mapping, we do not lose any fundamental information in kernel data analysis.
2. Generally the derivation becomes simpler using this mapping.
3. New applications of the kernel methods become easier using this mapping.

The last statement should be put into practice. In relation to fuzzy clustering, the method of fuzzy c-varieties should be

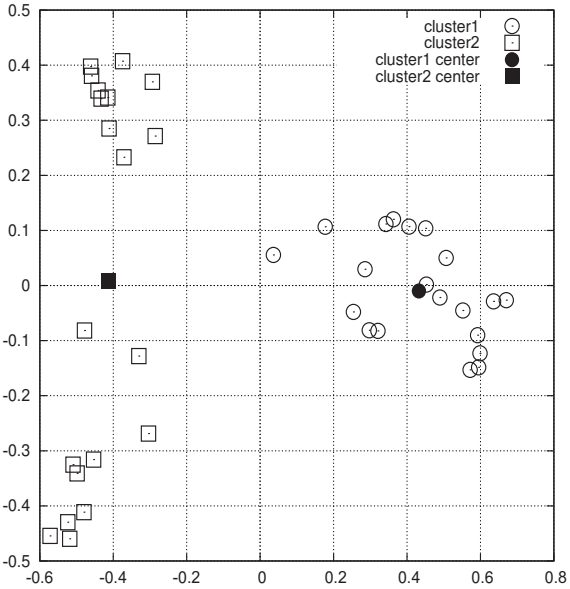


Figure 4: Two-dimensional display from KPCA with two clusters and cluster centers, where the second type of fuzzy equivalence has been used.

studied. We moreover have many research possibilities related to SOM.

This mapping invokes several problems to be solved. For example, when the number of objects are large, we have a problem of many dimensions which should be overcome using some handling large matrix techniques. We also should consider when and where such an explicit mapping become useless. Clarification of such a boundary between usefulness and uselessness is the ultimate objective of the present study.

Acknowledgment

This research has partly been supported by the Grant-in-Aid for Scientific Research, Japan Society for the Promotion of Science, Japan, No.19300074.

Appendix: Notes on Proofs

Since detailed proofs are too lengthy, we give notes on how to prove the propositions given here. The proofs are mostly straightforward.

Note on proof of Proposition 1

The detailed proof is given in standard textbooks [9]. As a rough sketch of the proof, note that the Mercer condition[10]

$$\int \int \boldsymbol{K}(x,y)\eta(x)\eta(y)dxdy \geq 0$$

for all $\eta(x)$ guarantees

$$\sum_{i,j} \boldsymbol{K}(x_i,x_j)\zeta_i\zeta_j \geq 0$$

$\forall \zeta_i \in \boldsymbol{R}$, by putting $\eta(x) = \sum_i \zeta_i \delta(x - x_i)$. Thus the matrix $K = (\boldsymbol{K}(x_i,x_j))$ is positive semi-definite. The kernel function generally does not distinguish positive semi-definiteness

ISBN: 978-989-95079-6-8

and positive definiteness, while positive-definiteness is required for the definition of an inner product. For this purpose a simple trick is to use a regularization which means that we take

$$K \to K + \epsilon I \quad (\epsilon > 0)$$

where ϵ is sufficiently small. The matrix $K + \epsilon I$ is positive definite and approximates K when K is positive semi-definite. In this paper we thus assume positive-definiteness throughout, by using such a regularization when needed.

Note on proof of Proposition 2

The proof of this proposition is not difficult by observing closely (6). Note that W_i given by (6) is the solution of

$$\min_{W_i} \sum_k (u_{ki})^m \|x_k - W_i\|^2$$

where the space can be an arbitrary inner product space, since the derivation of (6) uses a general variational principle valid for any Hilbert space. Note moreover that we substitute (11) into (5) to have the optimal solution of U. It should be noted that although optimal W_i is the same for all positive definite kernel, optimal U differs because (10) give different values for different kernels.

Note on proof of Proposition 3

A straightforward calculation shows this proposition is valid.

Note on proof of Proposition 4

The proof is immediate and omitted here.

Note on proof of Proposition 5

Take an arbitrary $c \in O$. The conclusion is immediately obtained from the Pólya's theorem [7] which states that

$$\sum_{a,b} z_a z_b f(|x(a) - x(b)|) \geq 0$$

when f is convex on $[0, +\infty)$ and $f(x) \to 0$ as $x \to +\infty$. Note that $x(a)$ is the real value defined by $x(a) = D(a, c)$ when a is the left hand of c; $x(a) = -D(a, c)$ when a is right hand of c.

Note on proof of Proposition 6

To prove this proposition, we consider a partition matrix $U = (u_{ij})$. Namely, an $n \times n$ real matrix U is called a partition matrix iff there exists a partition $K_1, \ldots, K_c$ of $\mathbf{n} = \{1, 2, \ldots, n\}$ ($\bigcup_j K_j = \mathbf{n}$ and $K_i \cap K_j = \emptyset$, for $i \neq j$) such that

$$u_{ij} = 1 \quad \forall i, j \in K_h$$

for some h and

$$u_{ij} = 0 \quad \forall i \in K_\ell, j \in K_h$$

for $h \neq \ell$.

We have

Lemma: *A partition matrix U is positive semi-definite. It is positive definite if and only if U is identity matrix ($U = I$).*

The proof of the lemma is immediate by observing

$$x^t U x = \sum_{i=1}^{c} \left(\sum_{x_j \in K_i} x_j \right)^2 .$$

The proof of this proposition is now straightforward. We assume that O is a finite set for simplicity. Then an equivalence relation is represented by a partition matrix. Moreover a fuzzy equivalence relation F is represented by a finite collection $U_1, \ldots, U_k$ of partition matrix and positive $\beta_1, \ldots, \beta_k$:

$$F = \sum_{j=1}^{k} \beta_j U_j.$$

Consequently we have

$$x^t F x = \sum_{j=1}^{k} \beta_j x^t U_j x \geq 0.$$

Hence $N(a, b)$ is positive definite. From Proposition 6, $p(t, t')$ is also positive definite. The proposition is thus proved.

Note on proof of Proposition 7

By straightforward calculation, we see this proposition holds.

References

[1] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Functions, Plenum Press, New York, 1981.

[2] T. Kohonen, Self-Organizing Maps, 2nd Ed., Springer, Berlin, 1995.

[3] S. Miyamoto, D. Suizu, Fuzzy c-means clustering using transformations into high dimensional spaces, Proc. of FSKD'02: 1st International Conference on Fuzzy Systems and Knowledge Discovery, Nov. 18-22, 2002, Singapore, Vol.2, pp. 656–660.

[4] S. Miyamoto, H. Ichihashi, K. Honda, Algorithms for Fuzzy Clustering, Springer, Berlin, 2008.

[5] S. Miyamoto, Y. Kawasaki, Kernel Space for Text Analysis Based on Fuzzy Neighborhoods, *WCCI 2008 Proceedings, 2008 IEEE World Congress on Computational Intelligence*, June 1-6, 2008, Hong Kong, pp.738-743, 2008.

[6] K. Mizutani, S. Miyamoto, Fuzzy multiset space and c-means clustering using kernels with application to information retrieval. in T.Bilgiç et al. Eds.: IFSA 2003, LNAI2715, Springer, Berlin, pp.387-395, 2003.

[7] G. Pólya, Remarks on characteristic functions, Proc. First Berkeley Conf. on Math. Stat. and Prob., pp.115–123, Univ. of Calf Press, Berkeley, 1949.

[8] G. Salton, M.J. McGill, Introduction to Modern Information Retrieval, McGraw-Hill, New York, 1985.

[9] B. Schölkopf, A.J. Smola, Learning with Kernels, the MIT Press, Cambridge, 2002.

[10] V.N. Vapnik, Statistical Learning Theory, Wiley, 1998.

ISBN: 978-989-95079-6-8

Distorted probabilities and m-separable fuzzy measures

Yasuo Narukawa[1] Vicenç Torra[2]

1.Toho Gakuen, 3-1-10, Naka, Kunitachi, Tokyo, 186-0004, Japan
1.The Institute for Research in Business Administration, Waseda University,
1-6-1, Nishi-Waseda, Shinjuku-ku, Tokyo,169-8050, Japan
2.IIIA - CSIC,
Campus UAB s/n, 08193 Bellaterra, Catalonia, Spain
Email: narukawa@d4.dion.ne.jp, vtorra@iiia.csic.es

***Abstract*— *Distorted probabilities are an important family of fuzzy measures. In a recent paper we introduced m-dimensional distorted probabilities, which generalize the former and permit us to have a smooth transition from distorted probabilities to unconstrained ones.*

In this paper we introduce the union condition and the strict union condition, and we show that when these conditions hold for a fuzzy measure, the fuzzy measure is a distorted probability. In addition, we present a few results that establish some relationships between other fuzzy measures.

Keywords: Fuzzy measures, distorted probabilities, m-dimensional distorted probabilities, m-symmetric fuzzy measures, union condition, strict union condition.

1 Introduction

Aggregation operators [28] are used in a large number of applications to combine information from different sources. Although the weighted mean is probably the most well known aggregation operator, other operators exist. E.g., the OWA, the WOWA and some fuzzy integrals. Choquet and Sugeno integrals are examples of such fuzzy integrals.

Fuzzy integrals permits the user to combine information when the sources supplying the information are not independent. To express this a priori knowledge about the sources, fuzzy integrals combine the input data with the information about the sources. Formally, the integrals integrate a function, which represents the data being aggregated, with respect to a fuzzy measure, which represents the a priori information about the sources.

A major difficulty for applying fuzzy integrals in real applications is that they are set functions, and thus, for any aggregation problem with n inputs, 2^n values should be defined. In fact, properly speaking, boundary conditions on the measure reduce this number to $2^n - 2$.

Real applications try to reduce the number of $2^n - 2$ required values using constrained measures. That is, measures that require less than $2^n - 2$ parameters. Sugeno λ-measures are probably the most used ones [22, 23, 20, 21]. Such measures solely require n values as well as an additional parameter λ, which can be deduced from the n values as [10] shows. k-order additive fuzzy measures are another family of measures with reduced complexity. This family, that has been extensively studied, is of special interest because the parameter k permits us to find a trade-off between expressiveness and complexity. In short, when $k = 1$ the measure has the lowest complexity (only $n - 1$ values are required) but it corresponds to a probability distribution. Instead, when $k = n$, any unconstrained fuzzy measure can be represented but at the cost of the highest complexity ($2^n - 2$ values should be defined).

In this paper we study m-dimensional distorted probabilities, another family of measures, introduced in [16], that generalizes distorted probabilities. Informally, distorted probabilities are measures that can be expressed in terms of a probability distribution and a function that distorts this distribution. Such measures, which were originated in psychology [19, 4, 5], have been extensively used. See, for example, the book by Aumann and Shapley (1974) [1] and their recent use in game theory. Recent research on such measures is on their determination either from examples (as in [27]) or from interviews [9]. Nevertheless, the modeling capabilities of distorted probabilities are limited. In [16] it was shown that the number of such measures with respect to the number of unconstrained ones is very small. Moreover, the larger is n, the smaller is the proportion of distorted probabilities. So, in most cases, fuzzy measures cannot be represented using distorted probabilities. To overcome this problem, we introduced in [16] m-dimensional distorted probabilities.

In this work we present some new results with respect to this family of measures. We show some conditions that, when fulfilled, imply distorted probabilities.

Other families of fuzzy measures have been studied in the literature. Two of them, that are relevant for the present paper, are the m-symmetric fuzzy measures [11, 12] and the hierarchically decomposable ones [26]. Some results establishing some relationships between these measures and distorted probabilities will be given.

The structure of the paper is as follows. In Sections 2, we review some concepts that are needed later on. In Section 3, we present the results establishing the connections among different kinds of fuzzy measures. The paper finishes with some conclusions.

2 Preliminaries

This section reviews some previous results in the literature that are needed in the rest of the paper. We start by defining fuzzy measures, and some of their families. Among them, we review m-dimensional distorted probabilities and a few results concerning these measures. The section finishes with a review of a few aggregation operators that are relevant for the purpose of this paper.

ISBN: 978-989-95079-6-8

2.1 Fuzzy measures

In this paper we will consider fuzzy measures on a finite universal set $X = \{x_1, \ldots, x_n\}$. For the sake of simplicity, when possible, we will consider $X := \{1, \ldots, n\}$. Now, we review the definition of fuzzy measure.

Definition 1 *A set function $\mu : 2^X \to [0, 1]$ is a fuzzy measure if it satisfies the following axioms:*

(i) $\mu(\emptyset) = 0$, $\mu(X) = 1$ *(boundary conditions)*

(ii) $A \subseteq B$ *implies* $\mu(A) \leq \mu(B)$ *(monotonicity)*

In order to distinguish measures satisfying (i) and (ii) with others that also satisfy some additional constraints (*e.g.* additivity $\mu(A \cup B) = \mu(A) + \mu(B)$ when $A \cap B = \emptyset$), we use the terms *unconstrained* fuzzy measures for the former ones and *constrained* fuzzy measures for the others.

2.2 m-symmetric fuzzy measures

The definition of these measures is based on the concept of *set of indifference*. Roughly speaking, a set of indifference is defined by elements that do not affect the value of the measure. That is, the elements of a set are indistinguishable with respect to the fuzzy measure.

Definition 2 *[11, 12] Given a subset A of X, we say that A is a set of indifference if and only if:*

$$\forall B_1, B_2 \subseteq A, |B_1| = |B_2|,$$

$$\forall C \subseteq X \setminus A \quad \mu(B_1 \cup C) = \mu(B_2 \cup C)$$

In this definition $|\cdot|$ corresponds to the cardinality of a set. We now consider m-symmetric fuzzy measures for the particular case of $m = 2$ and, then, we give the general definition.

Definition 3 *[11, 12] Given a fuzzy measure μ, we say that μ is an at most 2-symmetric fuzzy measure if and only if there exists a partition of the universal set $\{X_1, X_2\}$, with $X_1, X_2 \neq \emptyset$ such that both X_1 and X_2 are sets of indifference. An at most 2-symmetric fuzzy measure is 2-symmetric if X is not a set of indifference.*

Definition 4 *[11, 12] Given a fuzzy measure μ, we say that μ is an at most m-symmetric fuzzy measure if and only if there exists a partition of the universal set $\{X_1, \ldots, X_m\}$, with $X_1, \ldots, X_m \neq \emptyset$ such that $X_1, \ldots X_m$ are sets of indifference.*

The next proposition follows from this definition.

Proposition 1 *Every fuzzy measure μ is an at most n-symmetric fuzzy measure for $n = |X|$.*

So, all fuzzy measures can be considered as m-symmetric for a value of m large enough.

Definition 5 *[11, 12] Given two partitions $\{X_1, \ldots, X_p\}$ and $\{Y_1, \ldots, Y_r\}$ on the finite universal set X, we say that $\{X_1, \ldots, X_p\}$ is coarser than $\{Y_1, \ldots, Y_r\}$ if the following holds:*

$$\forall X_i \exists Y_j \text{ such that } Y_j \subseteq X_i$$

Definition 6 *Given a fuzzy measure μ, we say that μ is m-symmetric if and only if the coarsest partition of the universal set in sets of indifference contains m non empty sets. That is, the coarsest partition is of the form: $\{X_1, \ldots, X_m\}$, with $X_i \neq \emptyset$ for all $i \in \{1, \ldots, m\}$.*

Proposition 2 *[11, 12] Let μ be an m-symmetric measure with respect to the partition $\{X_1, \ldots, X_m\}$. Then, the number of values that are needed in order to determine μ is:*

$$\big[(|X_1| + 1) \cdots (|X_m| + 1)\big] - 2$$

An m-symmetric fuzzy measure can be represented in a $(|X_1| + 1) \cdots (|X_m| + 1)$ matrix M.

2.3 Hierarchically S-Decomposable Fuzzy Measures

[26] introduced Hierarchically S-Decomposable Fuzzy Measures. These measures (HDFM for short) can be seen as a generalization of S-decomposable measures. An important characteristic of S-decomposable fuzzy measures, is that the measure for any subset of X can be built from the measures on the singletons and a t-conorm S. When interactions among information sources are considered, such construction means that the interactions among pairs (or subsets) of sources can be expressed in a single and unique way. In particular, all interactions are modeled using the t-conorm S.

The so-called hierarchically S-decomposable fuzzy measures define a more general family of fuzzy measures as they permit us to express different kind of interactions between different subsets. This is achieved permitting us the use of different t-conorms for combining the measures of different singletons (and of different subsets).

This is obtained as follows: (i) the elements in X are structured in a hierarchy that gathers together elements that are *similar* (from the interactions point of view); (ii) each node of the hierarchy has associated a t-conorm to be used to *combine* the interactions. In this way, a richer variety of interactions can be expressed.

For example, if we have a fuzzy measure $\mu : 2^X \to [0, 1]$ with $X = \{x_1, x_2, x_3, x_4, \ldots, x_n\}$ such that $\mu(\{x_1\}) = 0.2$, $\mu(\{x_2\}) = 0.4$, $\mu(\{x_3\}) = 0.3$ and $\mu(\{x_4\}) = 0.3$. Then, we have a negative interaction between x_1 and x_2 defining $\mu(\{x_1, x_2\}) = \max(\mu(\{x_1\}), \mu(\{x_2\}))$. Instead, for a positive interaction between x_3 and x_4 we define $\mu(\{x_3, x_4\}) = \min(1, \mu(\{x_3\}) + \mu(\{x_4\}))$. Both situations can be modeled with the t-conorms $S_1(x, y) = \max(x, y)$ and $S_2(x, y) = \min(1, x + y)$.

We give below the definition for the particular case of 2-level HDFM. That is, a measure where the hierarchy has only two levels.

Definition 7 *[26] Given a fuzzy measure μ, we say that μ is a 2-level Hierarchically Decomposable Fuzzy Measure (2-level HDFM) if there is a partition $\{X_1, \ldots, X_m\}$ on X (we denote the elements in X_i by $X_i = \{x_{i,1}, \ldots, x_{i,m_i}\}$) and t-conorms S, S_1, …, S_m such that:*

$$\mu(A) = S(r_1(A), \ldots, r_m(A))$$

where

$$r_i(A) = S_i(\mu(\{x_{i,1}\} \cap A), \ldots, \mu(\{x_{1,m_i}\} \cap A))$$

ISBN: 978-989-95079-6-8

In the general case of HDFM, not presented here, a complete hierarchy is permitted and, then, the measure is defined recursively for each node using the t-conorm attached to the node, and the partition associated to the node.

2.4 Distorted probabilities

As briefly described in the introduction, distorted probabilities correspond to fuzzy measures that can be represented by a probability distribution and a distortion function. We formalize these measures as well as the required concepts below:

Definition 8 *Let $P : 2^X \to [0,1]$ be a probability measure. Then, we say that a function f is strictly increasing with respect to P if and only if*

$$P(A) > P(B) \text{ implies } f(P(A)) > f(P(B))$$

Remark: Since we suppose that X is a finite set, when there is no restriction on the function f, a strictly increasing function f with respect to P can be regarded as a strictly increasing function on $[0,1]$. Note that with respect to increasingness only the points in $\{P(A)|A \in 2^X\}$ are essential, the others are not considered by $f(P(A))$.

Definition 9 *[1, 2] Let μ be a fuzzy measure. We say that μ is a distorted probability if there exists a probability distribution P and a strictly increasing function f with respect to P such that $\mu = f \circ P$.*

The next theorem gives the necessary and sufficient condition for a fuzzy measure μ to be a distorted probability. The theorem is based on *Scott's condition*:

Definition 10 *[16] Let μ be a fuzzy measure, μ satisfies* Scott's condition *when for all $A_i, B_i \in 2^X$ such that $\sum_{i=1}^n 1_{A_i} = \sum_{i=1}^n 1_{B_i}$ the condition below holds:*
$\mu(A_i) \leq \mu(B_i)$ for $i = 2, 3, \ldots, n$ implies $\mu(A_1) \geq \mu(B_1)$.

Here 1_A represents the characteristic function of the set A. That is $1_A(x) = 1$ if and only if $x \in A$.

Using this condition, we can characterize distorted probabilities as follows:

Theorem 1 *[16] Let μ be a fuzzy measure; then, μ is a distorted probability if and only if Scott's condition holds.*

2.5 m-dimensional distorted probabilities

m-dimensional distorted probabilities were presented in [16] to overcome the limited expressiveness of distorted probabilities. They are defined as follows:

Definition 11 *[16] Let $\{X_1, X_2, \cdots, X_m\}$ be a partition of X; then, we say that μ is an at most m dimensional distorted probability if there exists a function f on $\mathbb{R}^m$ and probabilities P_i on $(X_i, 2^{X_i})$ such that:*

$$\mu(A) = f(P_1(A \cap X_1), P_2(A \cap X_2), \cdots, P_m(A \cap X_m)) \quad (1)$$

where f on $\mathbb{R}^m$ is strictly increasing with respect to each variable.

We say that an at most m dimensional distorted probability μ is an m dimensional distorted probability if μ is not an at most $m-1$ dimensional.

The next proposition follows from the definition above.

Proposition 3 *Every fuzzy measure is an at most n-dimensional distorted probability with $n = |X|$.*

Note that for $n = |X|$, we are considering the following partition of X: $\{X_1 = \{x_1\}, \ldots, X_n = \{x_n\}\}$. So, $f(a_1, \ldots, a_n) = \mu(A)$ when $a_i = 1$ if and only if $x_i \in A$.

Also, we can prove that m-dimensional distorted probabilities define a family of measures with increasing complexity with respect to m. This means that increasing the value of m, the number of measures representable increases. The following proposition establishes this property.

Proposition 4 *Let $\mathcal{M}_k$ be the set of all fuzzy measures that are k-dimensional distorted probabilities and let $\mathcal{M}_0$ be the empty set. Then $\mathcal{M}_{k-1} \subset \mathcal{M}_k$ for all $k = 1, 2, \ldots, |X|$.*

Corollary 1 *Given a fuzzy measure μ, there exists a $k \in \{1, 2, \ldots, |X|\}$ such that $\mu \in \mathcal{M}_k$ and $\mu \notin \mathcal{M}_{k-1}$.*

Therefore, the proposed family of fuzzy measures permits us to cover the whole set of fuzzy measures.

2.6 Aggregation operators

Now we define the OWA and the WOWA operators. They will be of relevance in this work. As explained in detail in [25], the OWA operator permits to give importance to the data (with respect to their position) while the WOWA permits to give importance to the data (as the OWA operator) and also to the information sources (as the weighted mean does).

Definition 12 *[29, 30] Let $\mathbf{w}$ be a weighting vector of dimension n (*i.e.*, $w_i \geq 0$ and $\sum_{i=1}^n w_i = 1$), then a mapping OWA: $\mathbb{R}^n \to \mathbb{R}$ is an* Ordered Weighted Averaging (OWA) operator *of dimension n if*

$$OWA_{\mathbf{w}}(a_1, ..., a_n) = \sum_{i=1}^{n} w_i a_{\sigma(i)}$$

where $\{\sigma(1), ..., \sigma(n)\}$ is a permutation of $\{1, ..., n\}$ such that $a_{\sigma(i-1)} \geq a_{\sigma(i)}$ for all $i = \{2, ..., n\}$ (i.e. $a_{\sigma(i)}$ is the i-th largest element in the collection $a_1, ..., a_n$).

Definition 13 *[25] Let $\mathbf{p}$ and $\mathbf{w}$ be two weighting vectors of dimension n, then a mapping WOWA: $\mathbb{R}^n \to \mathbb{R}$ is a* Weighted Ordered Weighted Averaging (WOWA) operator *of dimension n if*

$$WOWA_{\mathbf{p},\mathbf{w}}(a_1, ..., a_n) = \sum_{i=1}^{n} \omega_i a_{\sigma(i)}$$

where σ is defined as in the case of the OWA, and the weight ω_i is defined as:

$$\omega_i = w^*(\sum_{j \leq i} p_{\sigma(j)}) - w^*(\sum_{j<i} p_{\sigma(j)})$$

with w^* being a non-decreasing function that interpolates the points $\{(i/n, \sum_{j \leq i} w_j)\}_{i=1,\ldots,n}$ together with the point $(0,0)$. The function w^* is required to be a straight line when the points can be interpolated in this way.

ISBN: 978-989-95079-6-8

Alternatively, it is possible to define the WOWA operator directly using the function w^*. This will be denoted by $WOWA_{\mathbf{p},w^*}(a_1,\ldots,a_n)$ or $WOWA_{\mathbf{p},w^*}(f)$ when f is a function $f : X \to \mathbb{R}$ such that $f(x_i) = a_i$. In this latter case, we will read $WOWA_{\mathbf{p},w^*}(f)$ as the $WOWA$ of f with respect to $\mathbf{p}$ and w^*.

We finish the section with the definition of the Choquet integral.

Definition 14 *[3] Let μ be a fuzzy measure, then the* Choquet integral *of a function $f : X \to \mathbb{R}^+$ with respect to the fuzzy measure μ is defined by:*

$$(C)\int f d\mu (= C_\mu(f)) = \sum_{i=1}^{n} [f(x_{s(i)}) - f(x_{s(i-1)})]\mu(A_{s(i)})$$

where $x_i \in X$ and where $f(x_{s(i)})$ indicates that the indices have been permuted so that
$0 \le f(x_{s(1)}) \le \cdots \le f(x_{s(n)}) \le 1$,
$A_{s(i)} = \{x_{s(i)},\ldots,x_{s(n)}\}$ *and* $f(x_{s(0)}) = 0$.

3 Distorted probabilities and symmetric fuzzy measures

3.1 A sufficient condition for distorted probablity

We have discussed above Scott's condition: Definition 10 for distorted probability. Nevertheless, this condition is not easy to check. We present below another condition under which a fuzzy measure is a distorted probability. The new condition is easier to check.

Definition 15 *We say that a fuzzy measure μ satisfies a union condition (for short UC), if $A \cap C = \emptyset, B \cap D = \emptyset$,*

$$\mu(A) \ge \mu(B), \mu(C) \ge \mu(D) \Rightarrow \mu(A \cup C) \ge \mu(B \cup D).$$

The next proposition is obvious from the definition.

Proposition 5 *Suppose that a fuzzy measure μ satisfies UC. We have $A \cap C = \emptyset, B \cap D = \emptyset$,*

$$\mu(A) = \mu(B), \mu(C) = \mu(D) \Rightarrow \mu(A \cup C) = \mu(B \cup D).$$

Definition 16 *We say that a fuzzy measure μ satisfies a strict union condition (for short SUC), if $A \cap C = \emptyset, B \cap D = \emptyset$,*

$$\mu(A) > \mu(B), \mu(C) \ge \mu(D) \Rightarrow \mu(A \cup C) > \mu(B \cup D).$$

Proposition 6 *Suppose that a fuzzy measure μ satisfies SUC. We have $A \cap C = \emptyset, B \cap C = \emptyset$,*

$$\mu(A) > \mu(B) \Rightarrow \mu(A \cup C) > \mu(B \cup C).$$

Applying Proposition 5,6,we have the next proposition.

Proposition 7 *Suppose that a fuzzy measure μ satisfies UC. There exists a function F on*

$$\{(x,y) | x = \mu(A), y = \mu(B), A \cap B = \emptyset, A, B \in 2^X\}$$

such that

$$\mu(A \cup B) = F(\mu(A), \mu(B))$$

for $A, B \in 2^X, A \cap B = \emptyset$, and

$$F(x,0) = x, F(F(x,y),z) = F(x,F(y,z)).$$

Moreover if μ satisfies both UC and SUC, F is strictly monotone with respect to each variable.

Suppose that a fuzzy measure μ satisfies both UC and SUC. Since F is strictly monotone on

$$\{(x,y) | x = \mu(A), y = \mu(B), A \cap B = \emptyset, A, B \in 2^X\},$$

the domain of F can be extended to $[0,1] \times [0,1]$ and F is monotone with respect to each variable and F is continuous. Then F can be represented by strictly monotone function φ on [0,1] as

$$F(x,y) = \varphi^{-1}(\varphi(x) + \varphi(y)).$$

That is,

$$F(\mu(A), \mu(B)) = \varphi^{-1}(\varphi(\mu(A)) + \varphi(\mu(B)))$$

for $A \cap B = \emptyset$. Therefore we have

$$\varphi(\mu(A \cup B)) = \varphi(\mu(A)) + \varphi(\mu(B)).$$

Let $P(A) := \varphi(\mu(A))$. Then, P is a probability and we have $\mu(A) = \varphi^{-1}(P(A))$. Therefore we have the next theorem.

Theorem 2 *A fuzzy measure μ is distorted probability if μ satisfies both UC and SUC.*

3.2 Symmetric fuzzy measure

We start showing that a 1-Symmetric fuzzy measure is a special case of distorted probabilities.

Proposition 8 *Let $\mu = f \circ P$ be a distorted probability. Then, μ is a 1-symmetric fuzzy measure if and only if $P(A) = |A|/|X|$.*

Now we show that all m-symmetric fuzzy measures are m-dimensional distorted probabilities. This implies that 1-symmetric fuzzy measures are distorted probabilities.

Proposition 9 *Let μ be an m-symmetric fuzzy measure with respect to the partition $\{X_1,\ldots,X_m\}$. Then, μ is an m-dimensional distorted probability.*

Although the reversal of this proposition is not true, the next proposition characterizes one case in which m-dimensional distorted probabilities are m-symmetric fuzzy measures.

Proposition 10 *Let μ be an m-dimensional distorted probability. If, $p_i(x_j) = p_i(x_k)$ for all $x_j, x_k \in X_i$ and for all $i = 1,\ldots,m$, then μ is an m-symmetric fuzzy measure.*

It is known that OWA operators are equivalent to Choquet integrals with respect to symmetric fuzzy measures. Therefore, m-symmetric fuzzy measures permit us to define a generalization of OWA operators. The m-dimensional OWA is defined below:

Definition 17 *The m-dimensional OWA is defined as the Choquet integral with respect to an m-symmetric fuzzy measure.*

As proven in [24], a Weighted OWA (WOWA) operator is equivalent to a Choquet integral with respect to a distorted probability. Therefore, a Choquet integral with an m-dimensional probability can be seen as a generalization of the WOWA operator. We define an m-dimensional WOWA as follows:

ISBN: 978-989-95079-6-8

Definition 18 *The m-dimensional WOWA is defined as the Choquet integral with respect to an m-dimensional distorted probability.*

Then, considering Definitions 17 and 18 above, we have the following corollary from Proposition 9:

Corollary 2 *An m-dimensional OWA is a particular case of an m-dimensional WOWA. In other words, a Choquet integral with respect to an m-symmetric fuzzy measure is a particular case of a Choquet integral with respect to an m-dimensional distorted probability.*

3.3 m-separable fuzzy measure

As shown in Proposition 9, m-symmetric fuzzy measures are one special class of the m-dimensional distorted probabilities. We present another special class of fuzzy measure

Definition 19 *Let μ be a fuzzy measure. Then, we say that μ is a m-separable fuzzy measure if there exists a function g and a partition $\{X_1, \ldots, X_m\}$ of X such that*

$$\mu(A) = g(\mu(A \cap X_1), \ldots, \mu(A \cap X_m)) \tag{2}$$

where g is a m-dimensional function on $\mathbb{R}^m$. We say that g is a generating function for μ. We say that a generating function g is induced by h on $[0,1] \times [0,1]$ if $g(x_1, \ldots, x_m) = h(h(\ldots h(x_1, x_2), \ldots, x_{m-1}), x_m)$, $g(x_1, x_2, 0, \ldots, 0) = h(x_1, x_2)$.

Example 1 *Let $\{X_1, \ldots, X_m\}$ be a partition of X.*

1. *Suppose $g(x_1, \ldots, x_m) = x_1 + \cdots + x_m$, so g is induced by $h(x, y) = x + y$. Then we have*

$$\mu(A) = \mu(A \cap X_1) + \cdots + \mu(A \cap X_m).$$

This is an interadditivity defined in [14].

2. *Suppose $g(x_1, \ldots, x_m) = x_1 \vee \cdots \vee x_m$, so g is induced by $h(x, y) = x \vee y$. Then we have*

$$\mu(A) = \mu(A \cap X_1) \vee \cdots \vee \mu(A \cap X_m).$$

3. *Suppose $g(x_1, \ldots, x_m) = (x_1^2 + \cdots + x_m^2)^{1/2}$, so g is induced by $h(x, y) = (x^2 + y^2)^{1/2}$. Then we have*

$$\mu(A) = (\mu(A \cap X_1)^2 + \cdots + \mu(A \cap X_m)^2)^{1/2}.$$

Suppose that μ is a m-separable fuzzy measure generated by g and that g is induced by h. We say that h is associative if $h(h(x, y), z) = h(x, h(y, z))$. Suppose that h is strictly monotone and associative. Since g is symmetric, then h is symmetric, that is $h(x, y) = h(y, x)$. Then there exists a strictly monotone function φ such that $h(x, y) = \varphi^{-1}(\varphi(x) + \varphi(y))$. Define φ- Möbius inverse m^φ by

$$m^\varphi(A) := \sum_{B \subset A} (-1)^{|A \setminus B|} \varphi(\mu(B)).$$

Let $\mathcal{P}$ be a partition of X for a m-separable fuzzy measure. Applying theorem in [14], $m^\varphi(A) = 0 \Rightarrow A \not\subset C, C \in \mathcal{P}$. Let $M := \{A | m^\varphi(A) \neq 0\}$ We have $\mathcal{P} \subset M$. Define $A_l := \{A | A \in M, x_l \in A\}$ for $X = \{x_1, \ldots, x_l, \ldots, x_n\}$ and $M_l := \max\{|A| | A \in A_l\}$. Since $A \subset C, C \in \mathcal{P}$ for $A \in A_l$, we have the next proposition.

Proposition 11 *Let μ be a m - separable fuzzy measure generated by g, and g be induced by a strict monotone and associative h on $[0,1]$. Then we have*

$$m \times \min_{l \in |1,\ldots,n|} M_l \leq n.$$

We say that a fuzzy measure μ is a φ $k-$order additive if $max\{|A| | A \in M\} = k$. If $\varphi(x) = x$, a φ $k-$order additive fuzzy measure is a k-order additive fuzzy measure [6, 7].

Proposition 12 *Let μ be a m-separable fuzzy measure generated by g, and g be induced by strict monotone and associative h on $[0,1]$. If μ is k- additive, then we have $m \times k \geq n$.*

Now, we consider the relationship between the m-separable fuzzy measures and some other families of measures.

Proposition 13 *2-level HDFMs with S_i Archimedean t-conorms are a m-separable fuzzy measures.*

Theorem 3 *Let $\{X_1, \ldots, X_m\}$ be a partition of X and μ_i $i = 1, \ldots m$ be distorted probabilities represented by f_i and P_i (i.e., $\mu_i = f_i \circ P_i$). Then, there exists a m-separable fuzzy measure μ such that*

$$\sum_{i=1}^{m} ((C) \int f df_i \circ P_i) = (C) \int f d\mu \tag{3}$$

for all measurable function f.

As a corollary of this theorem, we have that the Choquet integral with respect to a m-separable fuzzy measure μ with $g(x_1, \ldots, x_m) = x_1 + \cdots + x_m$ can be represented as a two step Choquet integral.

Corollary 3 *Let μ be a m-separable fuzzy measure μ with $g(x_1, \ldots, x_m) = x_1 + \cdots + x_m$. Then the Choquet integral with respect to μ is represented as a two step Choquet integral of a 1st step integral with respect to a probability on $(1, \ldots, m)$. That is,*

$$(C) \int f d\mu = \int ((C) \int f df_i \circ P_i) dP(i). \tag{4}$$

4 Conclusions

In this paper we have introduced two new conditions, the Union Condition (UC) and the Strict Union Condition (SUC), and we have studied distorted probability under these conditions. We have shown that a fuzzy measure is a distorted probability when both UC and SUC conditions are satisfied.

Acknowledgements

Partial support by the Spanish MEC (projects ARES – CONSOLIDER INGENIO 2010 CSD2007-00004 – and eAEGIS – TSI2007-65406-C03-02) is acknowledged.

References

[1] Aumann, R. J., Shapley, L. S., (1974), Values of Non-atomic Games, Princeton Univ. Press.

[2] Chateauneuf, A., Decomposable Measures, (1996), Distorted Probabilities and Concave Capacities, Mathematical Social Sciences 31, 19-37.

[3] Choquet, G., (1953/54), Theory of capacities, Ann. Inst. Fourier, 5 131-295.

[4] Edwards, W., (1953), Probability-Preferences in Gambling, American Journal of Psychology 66 349-364.

[5] Edwards, W., (1962), Subjective probabilities inferred from decisions, Psychological review 69 109-135.

[6] Grabisch, M., (1996), k-order additive fuzzy measures, Proc. 6th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU), 1345-1350, Granada, Spain.

[7] Grabisch,M. k-order additive discrete fuzzy measures and their representation. Fuzzy Sets and Systems 92 (1997), 167-189.

[8] Grabisch, M., Murofushi, T., Sugeno, M., (2000), Fuzzy Measures and Integrals: Theory and Applications, Physica-Verlag.

[9] Honda, A., Nakano, T., Okazaki, Y., (2002), Subjective evaluation based on distorted probability, Proc. of the SCIS/ISIS conference.

[10] Leszczyski, K., Penczek, P., Grochulski, W., (1985), Sugeno's fuzzy measure and fuzzy clustering, Fuzzy Sets and Systems, 15 147-158.

[11] Miranda, P., Grabisch, M., (2002), p-symmetric fuzzy measures, Proc. 8th IPMU, 545-552, Annecy, France.

[12] Miranda, P., Grabisch, M., P. Gil, (2002),p-symmetric fuzzy measures, Int. J. of Unc., Fuzz. and Knowledge-Based Systems, 10 (Supplement), 105-123.

[13] Murofushi, T., Sugeno, M., Machida, M., (1994), Non-monotonic fuzzy measures and the Choquet integral, Fuzzy Sets and Systems, 64:1 73-86.

[14] Murofushi, T., Sugeno, M., Fujimoto, K., (1997), Separated hierarchical decomposition of the Choquet integral, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 5:5 563-585.

[15] Narukawa, Y., Torra, V., (2004), On m-dimensional distorted probabilities and p-symmetric fuzzy measures, IPMU 2004 (ISBN 88-87242-54-2), vol. 2, 1279-1284, Perugia, Italy, July 4-9.

[16] Narukawa, Y., Torra, V., (2005), Fuzzy measure and probability distributions: distorted probabilities, IEEE Trans. on Fuzzy Systems, 13:5 617- 629.

[17] Narukawa, Y., Torra, V., (2005), On multidimensional distorted probability, IFSA 2005, Beijing, China.

[18] Ovchinnikov, S., (1998), An Analytic Characterization of Some Aggregation Operators, Int. J. of Intel. Systems, 13 59-68.

[19] Preston, M. G., Baratta, P., (1948), An Experimental Study of the Auction-Value of an Uncertain Outcome, American Journal of Psychology, 61 183-93.

[20] Soria-Frisch, A., (2003), Hybrid SOM and fuzzy integral frameworks for fuzzy classification, in Proc. IEEE Int. Conf. on Fuzzy Systems 840-845.

[21] Soria-Frisch, A., (2006), Unsupervised Construction of Fuzzy Measures through Self-Organizing Feature Maps and its Application in Color Image Segmentation, Int. J. of Approx. Reasoning, in press.

[22] Sugeno, M., (1974), Theory of fuzzy integrals and its applications, Ph. D. Dissertation, Tokyo Institute of Technology, Tokyo, Japan.

[23] Tahani, H., Keller, J. M., (1990), Information fusion in computer vision using the fuzzy integral, IEEE Trans. on Systems, Man and Cybernetics, 20:3 733-741.

[24] Torra, V., (1998), On some relationships between the WOWA operator and the Choquet integral, Proc. 7th IPMU 818-824, Paris, France.

[25] Torra, V., (1997), The Weighted OWA operator, Int. J. of Intel. Systems, 12, 153-166.

[26] Torra, V., (1999), On hierarchically S-decomposable fuzzy measures, Int. J. of Intel. Systems, 14:9, 923-934.

[27] Torra, V., (2000), Learning Weights for Weighted OWA Operators, Proc. IEEE Int. Conf. Industrial Elect., Control and Instrumentation (IECON 2000), Nagoya, Japan. (CD-ROM).

[28] Torra, V., Narukawa, Y., (2007), Modeling decisions: aggregation operators and information fusion, Springer, in press. ISBN: 978-3-540-68789-4.

[29] Yager, R. R., (1988), On ordered weighted averaging aggregation operators in multi-criteria decision making, IEEE Trans. Syst., Man, Cybern., 18 183-190.

[30] Yager, R. R., (1996), Quantifier Guided Aggregation Using OWA operators, Int. J. of Intel. Systems, 11, 49-73.

[31] Yager, R. R., (1999), Nonmonotonic OWA operators, Soft Computing 3 187-196.

ISBN: 978-989-95079-6-8

Image Analysis Applications of Morphological Operators based on Uninorms

Manuel González Hidalgo[1] Arnau Mir Torres[2] Daniel Ruiz Aguilera[2] Joan Torrens Sastre[2]

[1] Computer Vision, Graphics and Artificial Intelligence Group, Maths. and Computer Science Department.
[2] Fuzzy Logic and Information Fusion Group, Maths. and Computer Science Department,
Universitat de les Illes Balears. Palma de Mallorca, Spain.
Email: {manuel.gonzalez,arnau.mir,dmijts0,daniel.ruiz}@uib.es

***Abstract*—** This paper presents a continuation of the study on a mathematical morphology based on left-continuous conjunctive uninorms given in [1]. Experimental results are displayed using the morphological Top-Hat transformation, used to highlight certain components of the image, and on the reduction and elimination of noise using alternate filters that are generated with the operators of opening and closing associated with these morphological operators

***Keywords*—** Mathematical morphology, Top-Hat, alternate filters, uninorms, representable uninorms, idempotent uninorm.

1 Introduction

The fuzzy mathematical morphology is a generalization of binary morphology [2] using techniques of fuzzy sets [3, 4, 5, 6]. The basic tools of mathematical morphology are the so-called morphological operations that are applied to an image A, which is modified through a structural element B, whose size and shape are chosen in order to study the structure of the image A. The basic morphological operations are: erosion, dilation, opening and closing. Other fuzzy extensions have been introduced by others authors, see fo example [7, 6].

The fuzzy operators used to build a fuzzy morphology are conjunctors and implications. Among them, the most commonly used are the t-norms and their residual implications. Recently conjunctive uninorms, as a particular case of conjunctors, have also been used in this area [8, 9, 1]. In these two last recent works the authors use two kinds of conjunctive uninorms in order to obtain a mathematical morphology with "good" properties, including duality between the morphological operators.

This work can be seen as a continuation of [1] where the authors made a comparative study of the results obtained using the morphology based on uninorms with those obtained using t-norms and using the classical umbra approach. In that work, it is checked out that uninorms detect the edges of the images better than other approaches. Also, in [1], properties of idempotence and generalized idempotence for opening and closing operators based on uninorms are shown. Thus, the aim of this paper is to extend this study to " Top-Hat" operators (residual between openings and closings, a morphological gradient) based on uninorms, as well as to the so-called alternate filters (alternate compositions of openings and closings). The Top-Hat is used to highlight certain components of the image, while the alternate filters are used in elimination and reduction of noise.

The residuals and morphological gradients are still under study and have applications in image analysis. Thus, in [10], it is proposed a directional transition detection algorithm based on morphological residues, considering linear structurant elements. In 2002, T. Chen et al. in [11], design on the basis of mathematical morphology, a new detector of soft edges in dark regions called "Pseudo Top-Hat" which provides, in the case of the classical image of the video cameraman (see Figure 1), a better performance than those achieved by other edge detection methods used to compare the results. In [12], morphological gradients based on openings and closings were used to detect edges in CT medical images altered by noise. The results were compared with other contours detectors, improving the obtained edge images. More recently, in [13, 14] gradient operators are applied again to edge detection and image segmentation, respectively. Thus, in [13], an edge detector is developed for obtaining thin edges in regions or images with low-contrast. Tested with multiple images, even some of them disturbed by noise type salt and pepper and Gaussian noise of mean zero and variance 0.02, the method performs very well in comparison with the other four methods with which they work. In [14] operators were designed based on multiscale morphological gradients, addressing the segmentation problem of images to Computed Axial Tomography (CAT). All these works show that the morphological gradients remain relevant and useful in the analysis and image processing.

The removal, reduction and smoothing of noise is one of the key tasks of analysis and image processing as a preliminary step to artificial vision. Mainly because many techniques of interpretation, measurement, segmentation, detection of structures, among others, fail or diminish its effectiveness in presence of noise. In [12, 13] the designed gradient operators are robust in the presence of noise. Many efforts have been devoted in the literature, and also are currently devoted, to the smoothing, reduction and elimination of noise, from both points of view: crisp and fuzzy. Some examples can be seen in [15, 16, 17]. More specifically, we can find a good job on fuzzy filters in [18] and more recently, in [19], where noise reduction is handled. The design of basic and fast filters remains a subject of interest. In this work it will be studied the feasibility of alternate filters, from opening and closing of the fuzzy morphology based on uninorms, in order to use them in the elimination and reduction of noise.

This work is organized as follows. In the next section we will introduce the fuzzy morphological operators based on left-continuous conjunctive uninorms and their basic properties. Section 3 analyzes the Top-Hat and filtering based on this morphology and shows some experimental results together with their interpretations. Finally, Section 4 is devoted to conclusions and future work.

ISBN: 978-989-95079-6-8

2 Fuzzy morphological operators and its properties

We assume known the basics facts on uninorms used in this work which, in any case, can be found in [20, 21, 22]. We will use the following notation: $\mathcal{I}$ is an implication, $\mathcal{C}$ a conjunction, $\mathcal{N}$ a strong negation, U a conjunctive uninorm with a neutral element e, $\mathcal{I}_U$ its residual implication, and finally A a gray-scale image and B a gray-scale structural element.

We recall the definitions of fuzzy morphological operators following the ideas of De Baets in [3]. The method consists in fuzzify the logical operations, i.e. the Boolean conjunction and the Boolean implication, to obtain fuzzy operators. An n-dimensional gray-scale image is modeled as an $\mathbb{R}^n \longrightarrow [0,1]$ function. The values of an image must be in $[0,1]$ in order to consider it as a fuzzy object. Then, we will proceed to explain this method from the following definitions and propositions.

Definition. The *fuzzy dilation* $D_{\mathcal{C}}(A,B)$ and *fuzzy erosion* $E_{\mathcal{I}}(A,B)$ of A by B are the gray-scale images defined by

$$D_{\mathcal{C}}(A,B)(y) = \sup_x \mathcal{C}(B(x-y), A(x)) \quad (1)$$

$$E_{\mathcal{I}}(A,B)(y) = \inf_x \mathcal{I}(B(x-y), A(x)). \quad (2)$$

Note that the reflection $-B$ of a fuzzy image B is defined by $-B(y) = B(-y)$, for all $y \in \mathbb{R}^n$. Given two images B_1, B_2, we will say that $B_1 \subseteq B_2$ when $B_1(y) \leq B_2(y)$ for all $y \in \mathbb{R}^n$.

Definition. The *fuzzy closing* $C_{\mathcal{C},\mathcal{I}}(A,B)$ and *fuzzy opening* $O_{\mathcal{C},\mathcal{I}}(A,B)$ of A by B are the gray-scale images defined by

$$C_{\mathcal{C},\mathcal{I}}(A,B)(y) = E_{\mathcal{I}}(D_{\mathcal{C}}(A,B), -B)(y) \quad (3)$$

$$O_{\mathcal{C},\mathcal{I}}(A,B)(y) = D_{\mathcal{C}}(E_{\mathcal{I}}(A,B), -B)(y). \quad (4)$$

In this paper, we use as conjunction two types of left-continuous conjunctive uninorms and as implication, their residual implications. Specifically these two types of uninorms are the following.

- The representable uninorms: Let $e \in]0,1[$ and let $h : [0,1] \longrightarrow [-\infty, \infty]$ be a strictly increasing, continuous function with $h(0) = -\infty$, $h(e) = 0$ y $h(1) = \infty$. Then $U_h(x,y) =$

$$\begin{cases} h^{-1}(h(x)+h(y)), & \text{if } (x,y) \notin \{(1,0),(0,1)\}, \\ 0, & \text{in other case,} \end{cases}$$

  is a conjunctive representable uninorm with neutral element e, see [20], and its residual implication $\mathcal{I}_{U_h}$ is given by $\mathcal{I}_{U_h}(x,y) =$

$$\begin{cases} h^{-1}(h(y)-h(x)), & \text{if } (x,y) \notin \{(0,0),(1,1)\}, \\ 1, & \text{in other case.} \end{cases}$$

  Moreover, U_h satisfies self duality (except at the points (0,1) and (1,0)) with respect to the strong negation $\mathcal{N}(x) = h^{-1}(-h(x))$, see [21].

- A specific type of idempotent uninorms: Let $\mathcal{N}$ be a strong negation. The function given by

$$U^{\mathcal{N}}(x,y) = \begin{cases} \min(x,y), & \text{if } y \leq \mathcal{N}(x), \\ \max(x,y), & \text{in other case,} \end{cases}$$

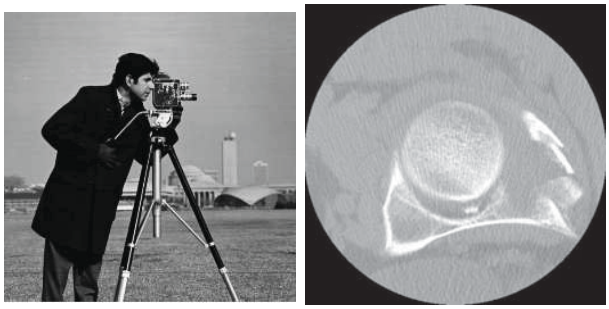


Figure 1: Original images used in experiments.

is a conjunctive idempotent uninorm. Its residual implication is given by (see [22])

$$\mathcal{I}_{U^{\mathcal{N}}}(x,y) = \begin{cases} \min(\mathcal{N}(x), y), & \text{if } y < x, \\ \max(\mathcal{N}(x), y), & \text{if } y \geq x. \end{cases}$$

These two types of conjunctive uninorms guarantees most of the good algebraic and morphological properties associated with the morphological operators obtained from them ([1, 9, 21, 22]). Among them, we highlight those described below. In all properties, U is a left-continuous conjunctive uninorm with neutral element $e \in]0,1[$, $\mathcal{I}_U$ its residual implication, A is a gray-level image and B a gray-scale structural element.

- The fuzzy dilation D_U is increasing in both arguments, the fuzzy erosion $E_{\mathcal{I}}$ is increasing in their first argument and decreasing in their second one, the fuzzy closing $C_{U,\mathcal{I}_U}$ and the fuzzy opening $O_{U,\mathcal{I}_U}$ are both increasing in their first argument.
- If $B(0) = e$ the fuzzy dilation is extensive and the fuzzy erosion is anti-extensive $E_{\mathcal{I}_U}(A,B) \subseteq A \subseteq D_U(A,B)$.
- Moreover, the fuzzy closing is extensive and the fuzzy opening is anti-extensive: $O_{U,\mathcal{I}_U} \subseteq A \subseteq C_{U,\mathcal{I}_U}(A,B)$.
- The fuzzy closing and the fuzzy opening are idempotent, i.e.: $C_{U,\mathcal{I}_U}(C_{U,\mathcal{I}_U}(A,B),B) = C_{U,\mathcal{I}_U}(A,B)$, and $O_{U,\mathcal{I}_U}(O_{U,\mathcal{I}_U}(A,B),B) = O_{U,\mathcal{I}_U}(A,B)$.
- If $B(0) = e$, then $E_{\mathcal{I}_U}(A,B) \subseteq O_{U,\mathcal{I}_U}(A,B) \subseteq A \subseteq C_{U,\mathcal{I}_U}(A,B) \subseteq D_U(A,B)$.
- For the two previous conjunctives uninorms of type U_h and $U^{\mathcal{N}}$, the duality between fuzzy morphological operators is guaranteed.

3 Residuals and basic filters

In the following experiments, the idempotent uninorm $U^{\mathcal{N}}$ with $\mathcal{N}(x) = 1 - x$ and the representable uninorm U_h with $h(x) = \ln\left(\frac{x}{1-x}\right)$ have been used. The obtained results are compared with the Łuckasiewicz t-norm and the classic development based on *umbra approach* (see [7]). In particular, the structural element used by the morphological operators is given by the following matrix:

$$B = e \cdot \begin{pmatrix} 0.86 & 0.86 & 0.86 \\ 0.86 & 1.00 & 0.86 \\ 0.86 & 0.86 & 0.86 \end{pmatrix}$$

ISBN: 978-989-95079-6-8



Figure 2: Left: closing. Right: opening. Top, an idempotent uninorm is used. Down, a representable uninorm is used.

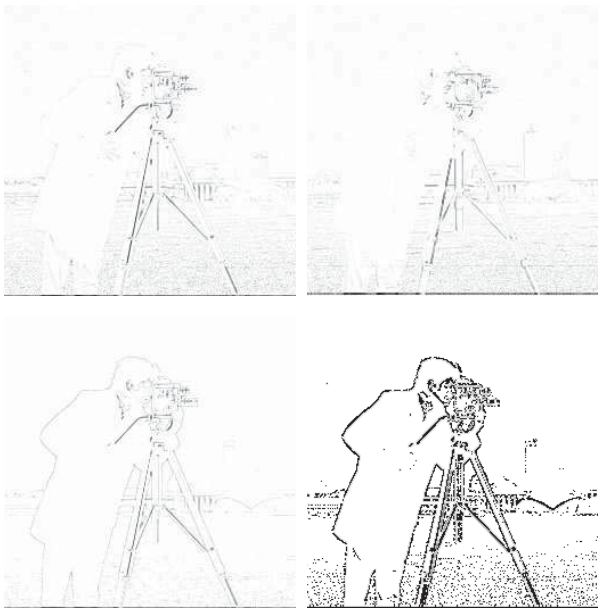


Figure 3: Left: Top-Hat. Right: Dual Top-Hat. Top: *umbra approach*. Down: Łuckasiewicz t-norm.

where e is the neutral element of the considered uninorm. The original images used in the experiments are shown in Figures 1 and 6 (they are represented by A in the fuzzy morphological operators). Two gray level original images are shown in Figure 1; one of them with size 256×256 is the classical image of a cameraman, and the other one of size 512×512 is the medical image of a hip.

The residual of two morphological operations or transformations is their difference. The first residual that can be defined in mathematical morphology is the morphological gradient, which is used as an edge detector and as a first approximation to a morphological segmentation. The morphological gradient, known in morphology as the Beucher gradient, is the difference between a dilation and an erosion, a dilation and the original image or the difference between the original image and its erosion. When we use the dilation and the erosion of a fuzzy set A by the structural element B, where the uninorm U is considered as a conjunctor and $\mathcal{I}_U$ as the residual implicator, the gradients are given by:

$$\nabla^{-}_{U,\mathcal{I}_U}(A,B) = A \setminus E_{\mathcal{I}_U}(A,B), \quad (5)$$

$$\nabla^{+}_{U,\mathcal{I}_U}(A,B) = D_U(A,B) \setminus A, \quad (6)$$

$$\nabla_{U,\mathcal{I}_U}(A,B) = D_U(A,B) \setminus E_{\mathcal{I}_U}(A,B), \quad (7)$$

where expression (5) represents the gradient by erosion, (6) the gradient by dilation and (7) the symmetrical gradient. From the point of view of the fuzzy inclusion, $\nabla^{-}_{U,\mathcal{I}_U}(A,B) \subseteq \nabla_{U,\mathcal{I}_U}(A,B)$ and $\nabla_{U,\mathcal{I}_U}(A,B) \supseteq \nabla^{+}_{U,\mathcal{I}_U}(A,B)$ are satisfied.

The goal of the gradients is the detection of the edges and/or the perimeter of the objects of the image. The correct choice of the structural element and the gradient will depend on the geometry of the objects. The application of the symmetrical gradient based on uninorms to the edge detection can be seen in [1], where the obtained results are better than those obtained by the gradients using t-norms or the classic approach.

The opening and closing defined in the previous section are the most elementary morphological filters, which are called *basic filters*. As we have seen in the previous section, the opening is an antiextensive morphological filter and the closing is an extensive morphological filter. So, in a first step, they can be used to remove non desired objects in an image. We can observe that the opening of a gray level image by a structural element removes the light zones of less size than the element and makes the light objects darker. The morphological closing helps to remove dark structures of less size than the structural element, toning down the dark objects. The size and shape of the used structurant element in the opening must agree with the image structures that we want to remove. Sometimes, structurant element of great size will remove the non desired shapes of an image, but also they can remove others shapes and the rest of the structures can be affected. Reduced sizes will be optimal when the images have small details. The effect of the closing and the opening on the cameraman's image can be observed in Figure 2, using the idempotent uninorm (top) and the representable uninorm (down).

The opening and the closing are morphological transformations and they are sensitive to compute the associated residuals, which are called *Top-Hat* transformations. The Top-Hat transformation, initiated by F. Meyer in 1977 (see [2]), finds structures of the images that have been removed by the opening or closing filter. If we choose the appropriated size and shape of the structural element, it is possible to filter and to remove some elements in the original image. The difference operation between the original map and the filtered map increases considerably the contrast of the removed zones.

The Top-Hat transformations are defined as a residual between the identity and the opening, which is called opening Top-Hat or white Top-Hat, or between the closing and the identity, which is called dual Top-Hat or black Top-Hat. They

ISBN: 978-989-95079-6-8

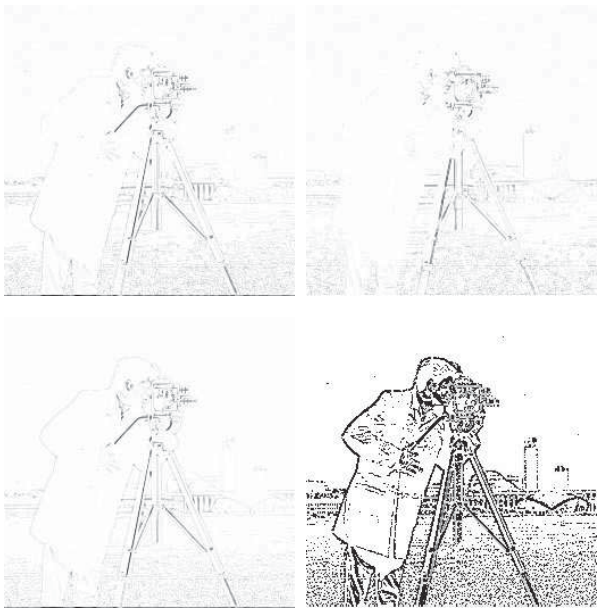


Figure 4: Left: Top-Hat. Right: Dual Top-Hat. Top, idempotent uninorm. Down, representable uninorm.

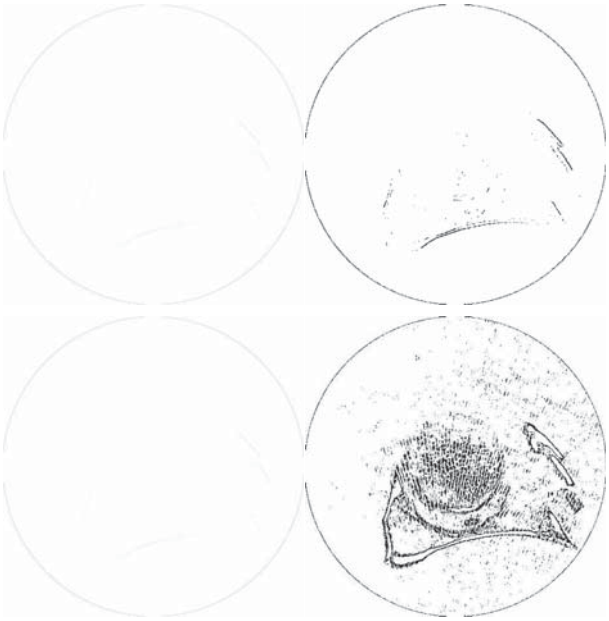


Figure 5: Left: Top-Hat. Right: Dual Top-Hat. Top, Łuckasiewicz t-norm. Down, representable uninorm.

are defined by:

$$\rho_{U,\mathcal{I}_U}(A,B) = A \setminus O_{U,\mathcal{I}_U}(A,B) \text{ (Top-Hat)}$$
$$\rho^d_{U,\mathcal{I}_U}(A,B) = C_{U,\mathcal{I}_U}(A,B) \setminus A \text{ (dual Top-Hat)}$$

The Top-Hat enhances the light objects that have been removed by the opening and it is used to extract contrasted components from the background, while the dual Top-Hat extracts the dark components that have been removed by the closing. Usually, the Top-Hat removes the soft trends, making thus an enhanced contrast. The detection of light objects is improved with an opening Top-Hat, in the same way with the dual Top-Hat does with the dark objects. The size of the structurant element is a deciding factor in order to apply the Top-Hat transformations.

The Top-Hat is idempotent (but it is not increasing) and it is antiextensive. The dual Top-Hat is neither increasing nor idempotent.

The obtained edges using the Top-Hat and the dual Top-Hat can be seen in Figures 3 and 4 for the cameraman image. The collected information using the idempotent uninorm is similar to the classical approach, and in the case of the representable uninorm, we can obtain (with the dual approach) more information than in the rest of the cases. Both top-hats are shown in Figure 5 for the hip image and it can be observed that the representable uninorm shows again more information.

It has been indicated that opening and closing are the basic filters of the fuzzy mathematical morphology. In order to build new mophological filters ([2]), we start with the basic filters and using composition or combination with other operations. New filters as sequential filters, center operators, contrast filters, etc. can be implemented. The basic applications of the morphological filters are the noise reduction and the selected extraction of image objects. Both applications are important in artificial vision since interpretation or measure techniques will fail in presence of noise, whereas the structure and the object selection are fundamental in segmentation processes.

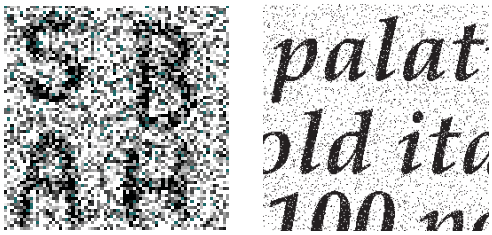


Figure 6: Original images with noise.

The first built filters from opening and closing are the so called *alternate filters*. Let $\xi(A,B)$ and $\psi(A,B)$ be the opening and closing, respectively, of a fuzzy object A by the structural element B using uninorm U as a conjuctor and $\mathcal{I}_U$ as a residual implicator. Using these two filters, four idempotent and growing filters can be generated: $\xi\,\psi$, $\psi\,\xi$, $\psi\,\xi\,\psi$ and $\xi\,\psi\,\xi$. These filters have similar properties to the classical morphological filters. Due to the idempotent property, the composition of more of three operators doesn't provide a new filter, except that we change the structural element.

In Figure 6, two binary images with added noise are shown. In the left image of size 70×74 the 60% of the points have been substituted by random graylevel noise[1]. In the right image of size 256×256, a gaussian noise has been added. Figure 7 shows the results obtained when we apply, from left to right, the alternate filters ψ, $\xi\,\psi$ and $\psi\,\xi\,\psi$, respectively. From top to bottom, the obtained results are shown using the different approaches. We can see that the uninorms remove the noise of the image in such a way that the *original* image is discernible. The same structure is shown in Figure 8.

The filtered image using the filter $\psi\,\xi$ and using the idempotent uninorm (very similar to the obtained image using the representable uninorm), Łuckasiewicz t-norm and *umbra approach* are shown in Figure 9. The filtered image with the

[1]This image has been downloaded of the web page of J. A. Sethian, http://math.berkeley.edu/~sethian/.

ISBN: 978-989-95079-6-8

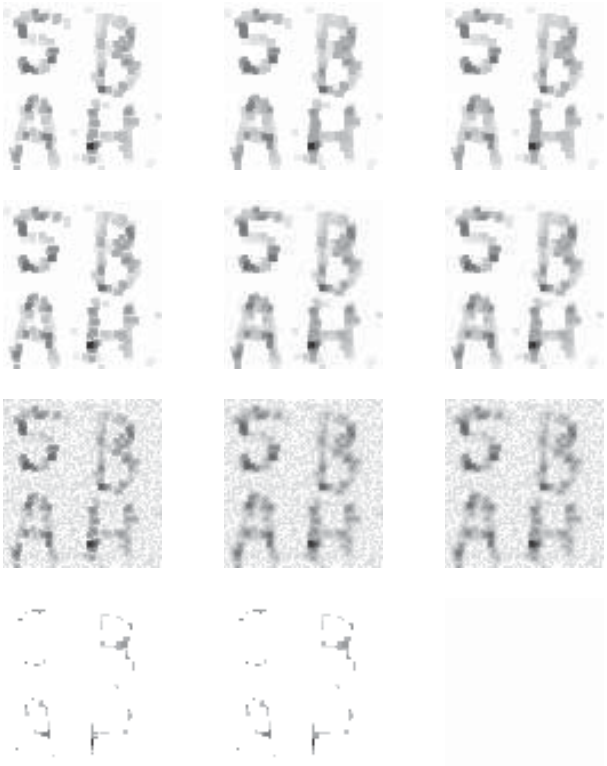


Figure 7: From left to right: filtered image using ξ, $\xi\,\psi$ and $\psi\,\xi\,\psi$, respectively. Top to bottom: idempotent uninorm, representable uninorm, Łuckasiewicz t-norm and *umbra approach*.



Figure 8: From left to right: filtered image using ξ, $\xi\,\psi$ and $\psi\,\xi\,\psi$, respectively. From top to bottom: idempotent uninorm, representable uninorm, Łuckasiewicz t-norm and *umbra approach*.

structure $\psi\,\xi\,\psi$ is very similar, except in the classical approach where the image gets smudged and in the case of the t-norm, the noise is softer.

A chest CT image of size 425×334 with "salt and pepper" added noise is shown in Figure 10 using 0.02 as a parameter. The results obtained using filter $\psi\,\xi\,\psi$ are shown in Figure 11. In the top left image, the Łuckasiewicz t-norm has been used and most of the noise remains. In the right image, the filtered image using the *umbra approach* is shown, a lot of "salt" noise is magnified. In the bottom images, the results obtained using uninorms are shown. The results using an idempotent uninorm are shown in the left image and the results using a representable uninorm are shown in the right image. The function generator of the representable uninorm is $h(x) = \frac{x-0.5}{x(1-x)}$. In the cases using uninorms the results are better than the ones using t-norms and umbra approach, because the noise is practically removed.

These type of filters have been used in the industrial control of pieces with partial occlusions or even in edge detection with noise elimination.

4 Conclusions and future work

In this work we have shown how the alternate filters based on opening and closing morphological operators using uninorms can remove the noise of an image preserving their structure. The obtained results improve the obtained results using t-norms and *umbra approach*. Also, we have obtained images very similar to the original ones without noise. So, the alternate filters based on uninorms are suitable to remove different

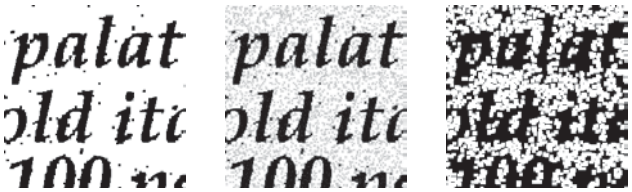


Figure 9: From left to right: filtered image using $\psi\,\xi$, an idempotent uninorm, Łuckasiewicz t-norm and *umbra approach*.

types of noise. The future work will consist on the construction of new filters. Also, we have shown the usefullness of the opening and closing filters in the edge detection. The future work includes the study of the effect of the choice of the structural element too. We have used isotropic structural elements in the experiments. That is, they have the same effect in all directions where they can be applied. We think that in some images, the use of anisotropic structural elements would improve the results.

We also want to apply the previous techniques to a more extended set of gray level images, even color images.

Acknowledgment

This work has been supported by the Government Spanish Grant MTM2006-05540, with FEDER support.

References

[1] M. González-Hidalgo, A. Mir-Torres, D. Ruiz-Aguilera, and J. Torrens. Edge-images using a uninorm-based fuzzy mathematical morphology: Opening and closing. In João Tavares

ISBN: 978-989-95079-6-8

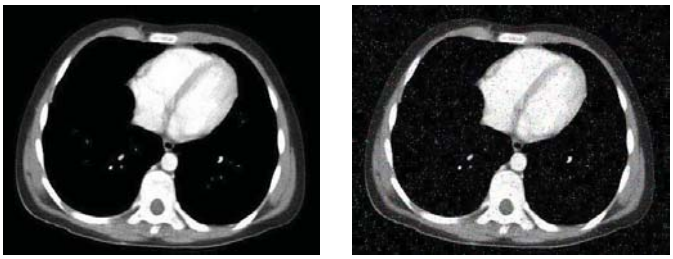


Figure 10: Left: Original image. Right: image with "salt and pepper" added noise.

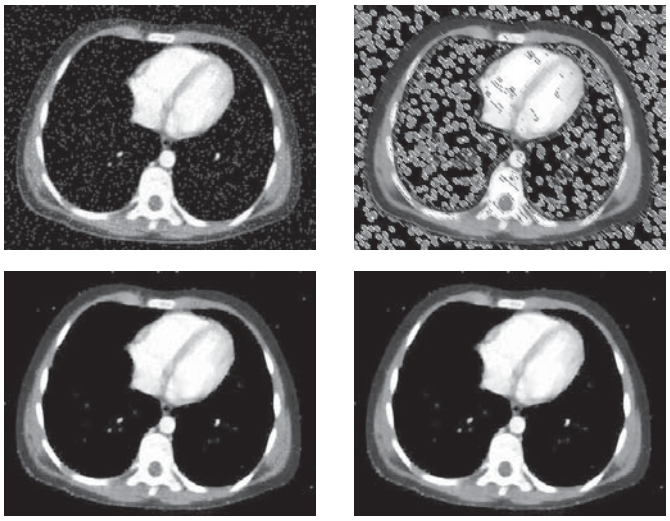


Figure 11: From left to right and from top to bottom, filtered image using $\psi\,\xi\,\psi$, and Łuckasiewicz t-norm, *umbra approach*, idempotent uninorm and representable uninorm.

and Natal Jorge, editors, *Advances in Computational Vision and Medical Image Processing*, number 13 in Computational Methods in Applied Sciences, chapter 8, pages 137–157. Springer, Netherlands, 2009.

[2] J. Serra. *Image analysis and mathematical morphology, vols. 1, 2*. Academic Press, London, 1982,1988.

[3] B. De Baets. Fuzzy morphology: A logical approach. In B. M. Ayyub and M. M. Gupta, editors, *Uncertainty Analysis in Engineering and Science: Fuzzy Logic, Statistics, and Neural Network Approach*, pages 53–68. Kluwer Academic Publishers, Norwell, 1997.

[4] B. De Baets, E.E. Kerre, and M. Gupta. The fundamentals of fuzzy mathematical morfologies part i: basics concepts. *International Journal of General Systems*, 23:155–171, 1995.

[5] B. De Baets, E.E. Kerre, and M. Gupta. The fundamentals of fuzzy mathematical morfologies part ii: idempotence, convexity and decomposition. *International Journal of General Systems*, 23:307–322, 1995.

[6] I. Bloch and H. Maître. Fuzzy mathematical morphologies: a comparative study. *Pattern Recognition*, 28:1341–1387, 1995.

[7] M. Nachtegael and E.E. Kerre. Classical and fuzzy approaches towards mathematical morphology. In Etienne E. Kerre and Mike Nachtegael, editors, *Fuzzy techniques in image processing*, number 52 in Studies in Fuzziness and Soft Computing, chapter 1, pages 3–57. Physica-Verlag, New York, 2000.

[8] B. De Baets, N. Kwasnikowska, and E. Kerre. Fuzzy morphology based on uninorms. In *Proceedings of the seventh IFSA World Congress*, pages 215–220, Prague, 1997.

[9] M. González, D. Ruiz-Aguilera, and J. Torrens. Algebraic properties of fuzzy morphological operators based on uninorms. In *Artificial Intelligence Research and Development*, volume 100 of *Frontiers in Artificial Intelligence and Applications*, pages 27–38, Amsterdam, 2003. IOS Press.

[10] S. Jamil Ferzoli Guimarães, N. Jerônimo Leite, M. Couprie, and A. de Albuquerque Araújo. A directional and parametrized transition detection algorithm based on morphological residues. In *Proceedings of the 15th Brazilian Symposium on Computer Graphics and Image Processing, SIBGRAPI'02*, pages 261–268, Washington, DC, USA, 2002. IEEE Computer Society.

[11] T. Chen, Q. H. Wu, R. Rahmani-Torkaman, and J. Hughes. A pseudo top-hat mathematical morphological approach to edge detection in dark regions. *Pattern Recognition*, 35(1):199–210, 2002.

[12] Zhao Yu-qian, Gui Wei-hua, Chen Zhen-cheng, Tang Jing-tian, and Li Ling-yun. Medical images edge detection based on mathematical morphology. In *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pages 6492–6495, Shanghai, China, 2005.

[13] J.-A. Jiang, C.-L. Chuang, Y.-L. Lu, and C.-S-Fahn. Mathematical-morphology-based edge detectors for detection of thin edges in low-contrast regions. *IET Image Processing*, 1(3):269–277, 2007.

[14] J. I. Pastore, E. Moler, and V. Ballarin. Multiscale morphological operators and geodesic distance applied to computer axial tomography segmentation. *IEEE Latin America Transactions*, 5(1):28–31, 2007.

[15] Hamid R. Tizhoosh. Fuzzy image enhancement: An overview. In Etienne E. Kerre and Mike Nachtegael, editors, *Fuzzy techniques in image processing*, number 52 in Studies in Fuzziness and Soft Computing, chapter 5, pages 137–171. Physica-Verlag, New York, 2000.

[16] Chang-Shing Lee and Yau-Hwang Kuo. Adaptative fuzzy filter and its application to image enhancement. In Etienne E. Kerre and Mike Nachtegael, editors, *Fuzzy techniques in image processing*, number 52 in Studies in Fuzziness and Soft Computing, chapter 6, pages 172–193. Physica-Verlag, New York, 2000.

[17] F. Farbiz and M. B. Menhaj. A fuzzy logic control based approach for image filtering. In Etienne E. Kerre and Mike Nachtegael, editors, *Fuzzy techniques in image processing*, number 52 in Studies in Fuzziness and Soft Computing, chapter 7, pages 194–221. Physica-Verlag, New York, 2000.

[18] M. Nachtegael, D. Van der Weken, D. Van De Ville, and E.E. Kerre, editors. *Fuzzy filters for image processing*, volume 122 of *Studies in Fuzziness and Soft Computing*. Springer-Verlag, 2003.

[19] D. Van De Ville, M. Nachtegael, D. Van der Weken, E.E. Kerre, and W. Philips. Noise reduction by fuzzy images filtering. *IEEE Transactions on Fuzzy Systems*, 11(4):429–436, 2003.

[20] J.C. Fodor, R.R. Yager, and A. Rybalov. Structure of uninorms. *Int. J. Uncertainty, Fuzziness, Knowledge-Based Systems*, 5:411–427, 1997.

[21] B. De Baets and J.C. Fodor. Residual operators of uninorms. *Soft Computing*, 3:89–100, 1999.

[22] D. Ruiz-Aguilera and J. Torrens. Residual implications and coimplications from idempotent uninorms. *Kybernetika*, 40:21–38, 2004.

ISBN: 978-989-95079-6-8

A value via posets induced by graph-restricted communication situations

Katsushige Fujimoto[1] Aoi Honda[2]
1.College of Symbiotic Systems Science, Fukushima University
1 Kanayagawa Fukushima 960-1296, Japan
2.Department of Systems Design and Informatics, Kyushu Institute of Technology
680-4 Kawazu Iizuka Fukuoka 820-8502, Japan
Email: fujimoto@sss.fukushima-u.ac.jp, aoi@ces.kyutech.ac.jp

Abstract— *This paper provides a new value (solution concept or allocation rule) of cooperative games via posets induced by graphs. Several values in a graph-restricted communication situation have been proposed or introduced by Myerson, Borm, and Hamiache... However, these values have been subjected to some criticisms in certain types of games. The value proposed in this paper withstands these criticisms. Moreover, these existing values have been defined only in situations represented by undirected graphs, while the notion of the value proposed in this paper can be extended to situations represented by directed graphs.*

Keywords— graph-restricted situations, communication situations, values, posets, cooperative games.

1 Introduction and Preliminaries

Throughout the paper, N denotes the universal set of n elements. For convenience, we often number the elements such that the universal set is $N = \{1, 2, \ldots, n\}$. A real-valued function $v : 2^N \to \mathbb{R}$ with $v(\emptyset) = 0$ is called a *game*. A monotone game (i.e., $v(A) \leq v(B)$ whenever $A \subseteq B \subseteq N$) is called a *capacity* or *a fuzzy measure*. We often call the pair (N, v), rather than v, a game or a capacity. The set of all games on N is denoted by $\mathcal{G}^N$. A real vector-valued function $\Phi : \mathcal{G}^N \to \mathbb{R}^{|N|}$ is called a *value*. In cooperative game theory, N is considered to be the set of all players. For every subset S of N, often called a *coalition*, $v(S)$ represents the (transferable) utility/profits that players in S can obtain if they decide to cooperate. For every game (N, v), the value $\Phi(N, v)$ represents an allocation rule, which provides an assessment of the benefits for each player from participating in a game v. For the sake of simplicity, we mainly discuss games in terms of various set functions (e.g., games, capacities, fuzzy measures, and so forth.) on N.

To avoid cumbersome notations, we often omit braces for singletons, e.g., by writing $v(i)$, $U \setminus i$ instead of $v(\{i\})$, $U \setminus \{i\}$. Similarly, for pairs, we write ij instead of $\{i, j\}$. Furthermore, cardinalities of subsets $S, T, \ldots$, are often denoted by the corresponding lower case letters $s, t, \ldots$, otherwise by the standard notation $|S|, |T|$,...

1.1 Games and capacities with graph restricted situations

In ordinary cooperative game theory it is implicitly assumed that all coalitions of N can be formed; however, this is generally not the case. For players to coordinate their actions, they must be able to communicate. The bilateral communication channels between players in N are described by a *communication network*. Such a network can be represented by an *undirected graph* (N, L), which has the set of players as its *nodes* $S \subseteq N$ and in which the players are connected by the set of *links* $L \subseteq \{ij \mid i, j \in N, i \neq j\}$; i.e., players i and j can communicate (directly) with each other if $ij \in L$. This paper will deals with only situations induced by communication networks described by undirected graphs. Many other approaches to the situations can be seen via the literatures [1, 2].

Definition 1.1 (communication situation)
The triple (N, v, L), which reflects a situation consisting of a game v on N and a communication network (N, L), is called a *communication situation*. We denote the set consisting of all communication situations on N by CS^N. For a coalition $T \subseteq N$, the restriction of (N, L) to T is denoted by $(T, L(T))$ and defined by $L(T) := \{ij \in L \mid ij \subseteq T\}$.

Definition 1.2 (feasible coalition) We say that players j and k are *connected in* $S \subseteq N$ if $j = k$ or there exists a subset $\{i_1, \cdots, i_m\} \subseteq S$ such that $j = i_1$, $k = i_m$, and $\{i_t, i_{t+1}\} \in L$ for all $t \in \{1, \cdots, m-1\}$. Then we denote $j \sim_S k$. Clearly, this relation $\sim_S$ is an equivalence relation. Hence, the notion of connectedness in S induces a partition $S/L := S/\sim_S$ of S. A coalition $S \subseteq N$ is said to be *feasible* in the communication network (N, L) if any two players, $j \in S$ and $k \in S$, are connected in S (i.e., $S/L = \{S\}$).

Example 1.1
Consider the communication situation (N_1, v, L_1) with $N_1 = \{1, 2, 3, 4, 5, 6, 7\}$ and $L_1 = \{12, 15, 26, 37, 47, 56\}$ (Fig.1). Then, all the players in $\{1, 2, 6\}$ can communicate with other;

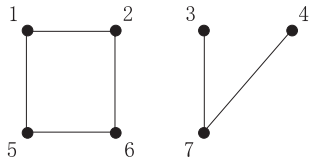


Figure 1: Communication network(N, L_1).

i.e., the coalition $\{1, 2, 6\}$ is feasible. Hence, they can fully coordinate their actions and obtain the value $v(\{1, 2, 6\})$. On the other hand, in the coalition $\{1, 2, 3, 4\}$, players 1 and 2 can communicate with each other, but players 3 and 4 cannot communicate with any other players in $\{1, 2, 3, 4\}$. Thus, feasible subcoalitions of $\{1, 2, 3, 4\}$ are $\{1, 2\}$, $\{3\}$, and $\{4\}$ (i.e., forming the coalition $\{1, 2, 3, 4\}$ is unfeasible). Hence, the value attainable by the players in $\{1, 2, 3, 4\}$ should be $v(\{1, 2\}) + v(\{3\}) + v(\{4\})$. In general, the value attainable by the players in $S \in N$ under a communication situation (N, v, L) is represented by

$$\sum_{T \in S/L} v(T). \tag{1}$$

ISBN: 978-989-95079-6-8

Definition 1.3 (network-restricted game [3]) The *network-restricted game* (N, v^L) associated with (N, v, L) is defined as

$$v^L(S) := \sum_{T \in S/L} v(T) \quad \text{for each } S \subseteq N. \tag{2}$$

Note that if (N, L) is the complete graph (i.e., $L = \{ij \mid i, j \in N, i \neq j\}$), the network-restricted game v^L is equal to the original game v.

The network-restricted game evaluates the possible gains from cooperation in a communication situation from the viewpoint of the players. The next example focuses on the importance of communication channels and links in a communication situation.

Example 1.2 In the communication situation L_1 depicted in Fig.1, the value obtainable by the players in the grand coalition N is

$$v^{L_1}(N) = v(\{1, 2, 5, 6\}) + v(\{3, 4, 7\}), \tag{3}$$

since $N/L_1 = \{\{1, 2, 5, 6\}, \{3, 4, 7\}\}$. If for some reason the communication link between players 4 and 7 is lost, the communication network L_1 becomes the new communication network $L_2 = \{12, 15, 26, 37, 56\}$. Then, $N/L_2 = \{\{1, 2, 5, 6\}, \{4\}, \{3, 7\}\}$ and the value obtainable by the players in the grand coalition N becomes

$$v^{L_2}(N) = v(\{1, 2, 5, 6\}) + v(\{4\}) + v(\{3, 7\}). \tag{4}$$

Then,

$$v^{L_1}(N) - v^{L_2}(N) \tag{5}$$

can be interpreted as a type of marginal contribution of the link $\{4, 7\} \in L_1$ to the communication network L_1.

Definition 1.4 (link game [4]) The *link game* (L, γ^v) associated with (N, v, L) consisting of a zero-normalized game v is a game on L defined by

$$\gamma^v(M) := v^M(N) = \sum_{T \in N/M} v(T) \quad \text{for each } M \subseteq L. \tag{6}$$

Note that, for an ordinary game v, γ^v is not a game on L since $\gamma^v(\emptyset) = \sum_{T \in N/\emptyset} v(T) = \sum_{i \in N} v(\{i\}) \neq 0$.

The link game $\gamma^v(M)$ represents the worth of the communication network $M \subseteq L$ as the worth of the grand coalition in the communication situation (N, v, M) through the network-restricted game v^M.

Definition 1.5 (Möbius Transform [5]**)** The *Möbius transform* of a game $v : 2^N \to \mathbb{R}$ (resp. $\gamma : 2^L \to \mathbb{R}$) is a game on N (resp. L) denoted by $\Delta^v : 2^N \to \mathbb{R}$ (resp. $\Delta^\gamma : 2^L \to \mathbb{R}$) and is defined by

$$\Delta^v(S) := \sum_{T \subseteq S} (-1)^{|S \setminus T|} v(T) \quad \text{for each } S \in 2^N. \tag{7}$$

$$(resp.\ \Delta^\gamma(M) := \sum_{K \subseteq M} (-1)^{|M \setminus K|} \gamma(K) \quad \text{for each } M \in 2^L). \tag{8}$$

Equivalently, we have that

$$v(S) = \sum_{T \subseteq S} \Delta^v(T) \quad \forall S \in 2^N. \tag{9}$$

$$(resp.\ \gamma(M) = \sum_{K \subseteq M} \Delta^\gamma(K) \quad \forall M \in 2^L). \tag{10}$$

Thus, the worth $v(S)$ (resp. $\gamma(M)$) of a coalition S (resp. communication network M) is equal to the sum of the Möbius transform of all its subcoalitions (subnetworks). This gives a recursive definition of the Möbius transform. The Möbius transform of every singleton is equal to its worth, while recursively, the Möbius transform of every coalition (resp. communication network) of at least two players (resp. links) is equal to its worth minus the sum of the Möbius transform of all its proper subcoalitions (resp. subnetworks). In this sense, the Möbius transform of a coalition S (resp. communication network M) can be interpreted as the extra contribution of the cooperation/synergy among the players in S (resp. links in M) that they did not already achieve by smaller coalitions (resp. networks). In fact, in the context of *interaction indices* (e.g.,[6, 7]), the Möbius transform $\Delta^v(S)$ is called the *internal interaction index* of S, which represents the magnitude of a type of interaction among the elements in S. The Möbius transform is also occasionally called the *Harsanyi dividends*[8].

Definition 1.6 (unanimity game) The *unanimity game* for a non-empty coalition $T \subseteq N$ is denoted by u_T and defined by

$$u_T(S) = \begin{cases} 1 & \text{if } S \supseteq T, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

For any game $v : 2^N \to \mathbb{R}$, v can be represented as

$$v(S) = \sum_{T(\neq \emptyset) \in 2^N} \Delta^v(T) \cdot u_T(S) \quad \forall S(\neq \emptyset) \in 2^N. \tag{12}$$

2 Values for communication situations

In this section, we briefly introduce the Shapley value for ordinary cooperative games and three existing values for communication situations that appear in the literatures [3, 4, 9], *the Myerson value, the position value, and the Hamiache value*.

Definition 2.1 (the Shapley value [10]) The *Shapley value* $\Phi : \mathcal{G}^N \to \mathbb{R}^{|N|}$ for a game $(N, v) \in \mathcal{G}^N$ is defined by

$$\Phi_i(N, v) := \sum_{T \ni i} \frac{1}{|T|} \Delta^v(T) \quad \text{for each } i \in N. \tag{13}$$

Definition 2.2 (the Myerson value [3]) The *Myerson value* $\Psi : CS^N \to \mathbb{R}^{|N|}$ for a communication situation $(N, v, L) \in CS^N$ is defined by

$$\Psi(N, v, L) := \Phi(N, v^L). \tag{14}$$

The Myerson value is the allocation rule that assigns to every communication situation (N, v, L) the Shapley value of the network-restricted game (N, v^L). Note that $\Psi(N, v, L) = \Phi(N, v)$ if (N, L) is the complete graph.

Definition 2.3 (position value [4]) The *position value* $\pi : CS^N \to \mathbb{R}^{|N|}$ for a communication situation $(N, v, L) \in CS^N$ is defined by

$$\pi_i(N, v, L) := \frac{1}{2} \sum_{\substack{l \in L \\ l \ni i}} \Phi_l(L, \gamma^v) \quad \text{for each } i \in N. \tag{15}$$

ISBN: 978-989-95079-6-8

The Shapley value $\Phi_l(L,\gamma^v)$ of a link $l \in L$, which is induced via (13) for the link game (L,γ^v), can be interpreted as a type of *expected marginal contribution* of the link l to all communication networks containing l. Then, the value is divided equally between the two players at the ends of the considered link $l \in L$. The position value of a given player $i \in N$ is obtained as the sum of all these shares.

We focus to a third value for communication situations, introduced by Hamiache [9]. Given a communication situation (N,v,L) and $S \subseteq N$, we denote by S^* the set of all nodes of the communication network (N,L) that are adjacent to at least one of the nodes of S,

$$S^* := \{i \in N \mid \exists j \in S \text{ such that } ij \in L\}. \tag{16}$$

Definition 2.4 (associated game [9]) For a value ϕ on CS^N (i.e., $\phi : CS^N \to \mathbb{R}^{|N|}$), the *associated game* v^*_ϕ of v with respect to ϕ is defined for $S \subseteq N$, by

$$v^*_\phi(S) := \begin{cases} v(S) + \displaystyle\sum_{j \in S^* \setminus S} \Big(\phi_i(S^{+j}, v|_{S^{+j}}, L(S^{+j})) - v(j)\Big) & \text{if } |S/L| = 1, \\ \displaystyle\sum_{T \in S/L} v^*_\phi(T) & \text{otherwise,} \end{cases} \tag{17}$$

where $S^{+j} := S \cup \{j\}$ and $v|_{S^{+j}}$ is the restriction of v to S^{+j}.

Hamiache [9] claims that there is a unique value ϕ, the so-called *Hamiache value*, for communication situations satisfying the following five properties, *component-efficiency*, *linearity w.r.t. games*, *independence of irrelevant players*, *positivity*, and *associated consistency*:

Component-efficiency :
For any (N,v,L) and any $S \in N/L$,

$$\sum_{i \in S} \phi_i(N,v,L) = v(S). \tag{18}$$

Linearity w.r.t. games :
For any $\alpha,\beta \in \mathbb{R}$ and $(N,v,L),(N,w,L) \in CS^N$,

$$\phi(N,\alpha v + \beta w, L) = \alpha\phi(N,v,L) + \beta\phi(N,w,L). \tag{19}$$

Independence of irrelevant players :
For any (N,L) and for any two feasible coalitions $R \subseteq T$,

$$\phi_i(N,u_R,L) = \phi_i(T,u_R,L(T)) \ \ \forall i \in T. \tag{20}$$

Positivity :
For any feasible coalition $T \subseteq N$,

$$\phi_i(T,u_T,L(T)) \geq 0 \ \ \forall i \in T. \tag{21}$$

Associated consistency:
For any $(N,v,L) \in CS^N$,

$$\phi(N,v,L) = \phi(N,v^*_\phi,L). \tag{22}$$

Note that $\phi(N,v,L) = \Phi(N,v)$ if (N,L) is the complete graph.

3 Posets induced by communication networks

3.1 Communication networks and posets

In this subsection, we consider and introduce a subposet of $B(n) := (2^N, \subseteq)$ induced by a communication network (N,L).

For a communication network (N,L), the set of all feasible coalitions in (N,L) is denoted by $P(N,L)$. i.e.,

$$P(N,L) := \{S \subseteq N \mid |S/L| = 1\}. \tag{23}$$

The set $P(N,L)$, together with set inclusion $\subseteq$ as an order on $P(N,L)$, is called the *poset induced by the communication network* (N,L).

Example 3.1 Let $N = \{1,2,3\}$, $L_a = \{12,13,23\}$, $L_b = \{13,23\}$, and $L_c = \{12\}$. Then the posets induced by communication networks (N,L_a), (N,L_b), and (N,L_c), as shown in (a) – (c) in Fig. 2, are represented as shown in (a) – (c) in Fig. 3, respectively.

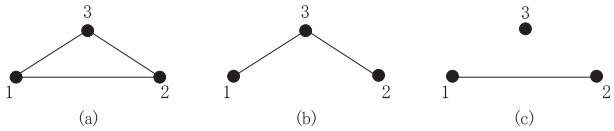


Figure 2: Communication networks on $N = \{1,2,3\}$.

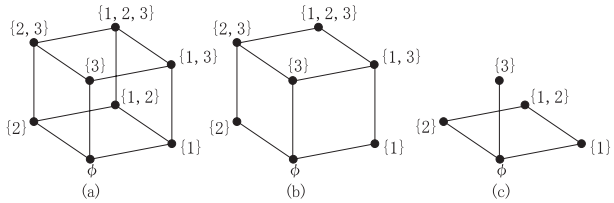


Figure 3: Posets corresponding to networks in Fig. 2.

Definition 3.1 (Möbius transform on posets)
Let $P := (N,\leq)$ be a poset. For a function $v : P \to \mathbb{R}$, the *Möbius transform* Δ^v of v is a function on P satisfying the following equation:

$$v(x) = \sum_{y \leq x} \Delta^v(y) \quad \forall x \in P. \tag{24}$$

Definition 3.2 (representation functions)
The *representation function of a communication situation* (N,v,L) is a function v^P on the poset $P(N,L)$ defined by

$$v^P(S) = v(S) \quad \text{for each } S \in P(N,L). \tag{25}$$

Then, the Möbius transform Δ^{v^P} of v^P is represented as

$$\Delta^{v^P}(S) := \sum_{\substack{T \in P(N,L) \\ T \subseteq S}} (-1)^{|S \setminus T|} v^P(T) \quad \forall S \in P(N,L). \tag{26}$$

Conversely,

$$v^P(S) := \sum_{\substack{T \in P(N,L) \\ T \subseteq S}} \Delta^{v^P}(T) \quad \forall S \in P(N,L). \tag{27}$$

Definition 3.3 (poset representation) The *poset representation of a communication situation* (N,v,L) is the pair $(P(N,L),\Delta^{v^P})$ of the poset induced by (N,v,L) and the Möbius transform Δ^{v^P} of representation function v^P of (N,v,L).

ISBN: 978-989-95079-6-8

4 A new value in communication situations

In this section, we introduce a new value for communication situations.

4.1 An interpretation of the Shapley value

Now, we consider the case $N = \{1,2\}$; the Shapley value $\Phi_1(N,v)$ of player 1 in a game v is obtained, from (13), as

$$\Phi_1(N,v) = \frac{1}{1}\Delta^v(\{1\}) + \frac{1}{2}\Delta^v(\{1,2\}). \tag{28}$$

This can be interpreted as an allocation rule of *Harsanyi dividends* (i.e., the Möbius transform) described as follows:

Allocation rule of Harsanyi dividends : We consider a process to form the coalition $\{1,2\}$. Then, there are two shortest paths from $\emptyset$ to $\{1,2\}$ in Fig. 2. One is the path $\emptyset \to \{1\} \to \{1,2\}$; another is the path $\emptyset \to \{2\} \to \{1,2\}$. The path $\emptyset \to \{1\} \to \{1,2\}$ can be interpreted as follows: *Player 1 makes an offer to player 2 for forming the coalition* $\{1,2\}$. *Player 2 accepts the offer and adds to the coalition* $\{1\}$ *to form the new coalition* $\{1,2\}$. Among these two paths, the only path that passes through $\{1\}$ is $\emptyset \to \{1\} \to \{1,2\}$. That is, the number of paths from $\emptyset$ to $\{1,2\}$ is 2, while of the number of paths via $\{1\}$ is 1. Then player 1 obtains $\frac{1 \text{ path}}{2 \text{ paths}}$ of the amount of the Harsanyi dividend $\Delta^v(\{1,2\})$ (i.e., $\frac{1}{2}\Delta^v(\{1,2\})$). In the same way, player 1 obtains $\frac{1}{1}\Delta^v(\{1\})$ and $\frac{0}{1}\Delta^v(\{2\})$. The Shapley value of player 1 is obtained as the sum of all these shares.

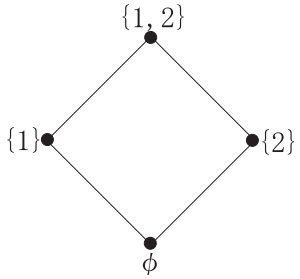


Figure 4: The Boolean lattice $B(2)$ on $N = \{1,2\}$.

This allocation rule can be extended to the case $N = \{1,2,3\}$ (Fig. 5).

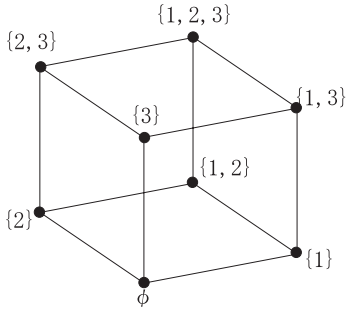


Figure 5: The Boolean lattice on $N = \{1,2,3\}$.

Indeed,

$$\begin{aligned}\Phi_1(N,v) &= \frac{1}{1}\Delta^v(\{1\}) + \frac{1}{2}\Delta^v(\{1,2\}) + \frac{1}{2}\Delta^v(\{1,3\}) \\ &+ \frac{0}{2}\Delta^v(\{2,3\}) + \frac{2}{6}\Delta^v(\{1,2,3\}).\end{aligned} \tag{29}$$

For instance, there are six shortest paths from $\emptyset$ to $\{1,2,3\}$. Among them, two paths pass through $\{1\}$, as shown in Fig. 6.

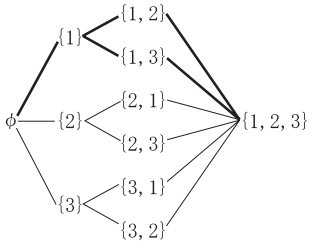


Figure 6: Shortest paths from $\emptyset$ to $\{1,2,3\}$.

4.2 An interpretation of the Myerson value

The Myerson value of (N,v,L) is the Shapley value of the network-restricted game (v^L,N). That is, the Myerson value is obtained by applying the above allocation rule to Harsanyi dividends $\{\Delta^{v^L}\}$ of v^L. Then Δ^{v^L} is given as follows:

Proposition 4.1 Let $(N,v,L) \in CS^N$ be a communication situation and $(B(n),\Delta^{v^L})$ the poset representation of the network-restricted game (N,v^L) associated with (N,v,L). Then,

$$\Delta^{v^L}(S) = \begin{cases} \Delta^v(S) & \text{if } S \in P(N,L), \\ 0 & \text{otherwise.} \end{cases} \tag{30}$$

4.3 Criticisms of existing values

Each of the existing values for communication situations, the Myerson value, the position value, and the Hamiache value, has been subject to criticisms, as follows.

The Myerson value :

$$\Psi_i(N,u_S,L) = \Psi_i(N,u_S,M) = \frac{1}{|S|} \quad \forall i \in N \tag{31}$$

whenever S is a feasible coalition in both (N,L) and (N,M). For example, consider the communication situation with $L = \{ij \subseteq N \mid j \in N \setminus i\}$ (i.e., L is a star with a central player i); then every player receives the same value (see Example $\Psi(N,v,L_e)$ in Example 5.3).

The position value :

Irrelevant null players often have positive values (see Example 5.2), where a null player $i \in N$ of the game (N,v) is a player satisfying $v(S \cup i) = v(S)$ for any $S \subseteq N$.

The Hamiache value :

It is very complex to compute the Hamiache value. Not only that, associated consistency is rather technical.

ISBN: 978-989-95079-6-8

4.4 A new value in communication situations

In this section, we propose a new value for communication situations that withstands all these criticisms.

Definition 4.1 (chain, saturated chain) A *chain* (or a *totally ordered set* or *linear ordered set*) is a poset in which any two elements are comparative. That is, a subset C of $P(N,L)$ is called a chain if $S \subseteq T$ or $T \subseteq S$ for any $S, T \in C$. The chain C of $P(N,L)$ is *saturated* (or *unrefinable*) if there does not exist $W \in P(N,L) \setminus C$ such that $S \subsetneq W \subsetneq T$ for some $S, T \in C$ and that $C \cup W$ is a chain.

Definition 4.2 (shortest path) For two feasible coalitions S, $T \in P(N,L)$, a saturated chain $\mathcal{P}$ of $P(N,L)$ is called a *shortest path* from S to T if $S, T \in \mathcal{P}$ and $S \subseteq W \subseteq T$ for any $W \in \mathcal{P}$. Then, we denote the set of all shortest paths from S to T by $\{S \to T\}$.

In the following, we propose a new value in communication situations, based on the interpretation of the Shapley value mentioned in Subsection 4.1.

Definition 4.3 We now propose a new value $\sigma(N,v,L)$ of a communication situation (N,v,L), as follows.

$$\sigma_i(N,v,L) := \sum_{S \in P(N,L)} \frac{|\{i \to S\}|}{|\{\emptyset \to S\}|} \Delta^{v^P}(S) \quad \text{for each } i \in N. \tag{32}$$

The number $|\{\emptyset \to S\}|$ of all shortest paths from $\emptyset$ to S indicates the number of all processes in which the feasible coalition S is formed. Also, $|\{i \to S\}|$ indicates the number of all processes in which the feasible coalition S is formed by the initiator $i \in N$. Then, the player $i \in N$ obtains $\frac{|\{i \to S\}|}{|\{\emptyset \to S\}|}$ of the amount of $\Delta^{v^P}(S)$ if $\Delta^{v^P}(S)$ is allocated in proportion to the frequency with which the player i initiates the formation of the feasible coalition S. The value proposed here of a given player $i \in N$ is obtained as the sum of all these shares.

Now we show an example that supports the naturalness of the definition of this value.

Example 4.1 We consider the communication situation (N,v,L) with $N = \{1,2,3\}$, $L = \{13,23\}$, and $\Delta^{v^P}(S) \geq 0$ for any $S \in P(N,L)$. The value $\sigma_i(N,v,L)$ proposed here of player $i \in N$ is represented as the values of the ammeters A_i in the electric circuit with current sources $I_S = \Delta^{v^P}(S)$, as shown in Fig.7.

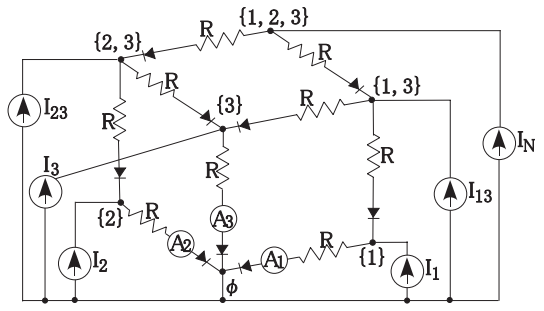


Figure 7: Electric circuit representing (N,v,L).

Property 1 The value σ proposed here satisfies *component-efficiency*, *linearity w.r.t. games*, *independence of irrelevant players*, and *positivity*.

Property 2 Let (N,u_N,L_c^*) be a communication situation with $L_c^* = \{cj \mid j \in N \setminus c\}$, $c \in N$. Then,

$$\sigma_i(N,u_N,L_c^*) = \begin{cases} \dfrac{1}{2} & \text{if } i = c, \\ \dfrac{1}{2(n-1)} & \text{otherwise.} \end{cases} \tag{33}$$

That is, if the communication network (N,L) is a star-graph with central player $c \in N$, in the unanimity game u_N, the central player obtains a half of the total amount of $u_N(N) = 1$ and the rest of the amount are shared out equally among the other players (see L_b, L_e in example 5.3).

However, we have not found any axiomatic characterization of the value proposed in this paper yet.

5 Comparison of existing values

In this section, we compare the existing four values (the Shapley, Myerson, position, and Hamiache values) and the value proposed in this paper. Examples 5.1 and 5.2 not only compare them but also illustrate the criticisms against the Shapley, Myerson, and position values, respectively.

Example 5.1 Consider the communication situation (N,v,L) with $N = \{1,2,3\}$, $L = \{13,23\}$ ((b) in Fig. 2), and

$$v(S) = \begin{cases} 0 & \text{if } |S| \leq 1 \\ 30 & \text{if } |S| = 2 \\ 36 & \text{if } S = N. \end{cases} \tag{34}$$

Then,

$$\Phi(N,v) = (12,12,12), \quad \Psi(N,v,L) = (7,7,22),$$
$$\pi(N,v,L) = (9,9,18), \quad \phi(N,v,L) = (9,9,18),$$
$$\sigma(N,v,L) = (9,9,18).$$

Example 5.2 Consider the communication situation (N,v,L) with $N = \{1,2,3\}$, $L = \{12,13,23\}$ (L_d in Fig. 8), and

$$v(S) = \begin{cases} 12 & \text{if } S \supseteq \{1,2\} \\ 0 & \text{otherwise.} \end{cases} \tag{35}$$

Then,

$$\Phi(N,v) = (6,6,0), \quad \Psi(N,v,L) = (6,6,0),$$
$$\pi(N,v,L) = (5,5,2), \quad \phi(N,v,L) = (6,6,0),$$
$$\sigma(N,v,L) = (6,6,0).$$

Example 5.3 Consider communication situations (N,u_N,L) with $2 \leq |N| \leq 4$, $|N/L| = 1$ (i.e., (N,L) is connected). Fig.8 shows all connected graphs (up to isomorphism) with $2 \leq n \leq 4$ nodes. Then, for any such communication situations (N,u_N,L),

$$\Phi_i(N,u_N,L) = \Psi_i(N,u_N,L) = \frac{1}{|N|} \quad \forall i \in N. \tag{36}$$

Table 1 shows comparisons of the remaining values (i.e., the position value π, the Hamiache value ϕ, and the value σ proposed in this paper), and illustrates that the value σ does not always coincide with the Hamiache value ϕ.

ISBN: 978-989-95079-6-8

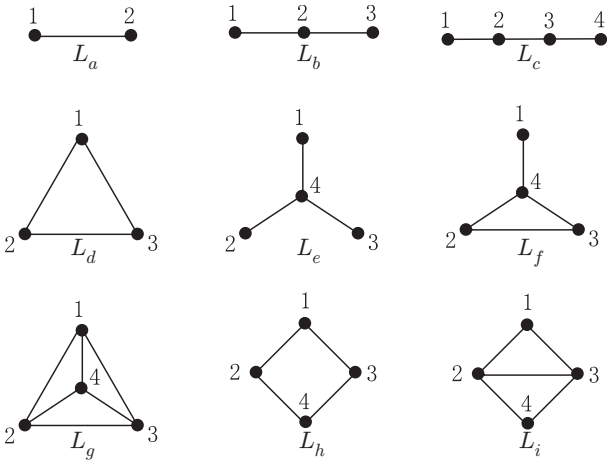


Figure 8: Graphs with at most four nodes.

Table 1: Comparison of existing values.

	π	ϕ	σ
L_a	$(\frac{1}{2}, \frac{1}{2})$	$(\frac{1}{2}, \frac{1}{2})$	$(\frac{1}{2}, \frac{1}{2})$
L_b	$(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$	$(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$	$(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$
L_c	$(\frac{1}{6}, \frac{2}{6}, \frac{2}{6}, \frac{1}{6})$	$(\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8})$	$(\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8})$
L_d	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$
L_e	$(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{3}{6})$	$(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{3}{6})$	$(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{3}{6})$
L_f	$(\frac{3}{12}, \frac{2}{12}, \frac{2}{12}, \frac{5}{12})$	$(0.172, 0.190, 0.190, 0.448)$	$(\frac{2}{14}, \frac{3}{14}, \frac{3}{14}, \frac{6}{14})$
L_g	$(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$	$(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$	$(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$
L_h	$(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$	$(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$	$(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$
L_i	$(\frac{13}{60}, \frac{17}{60}, \frac{17}{60}, \frac{13}{60})$	$(\frac{3}{14}, \frac{4}{14}, \frac{4}{14}, \frac{3}{14})$	$(\frac{2}{10}, \frac{3}{10}, \frac{3}{10}, \frac{2}{10})$

Acknowledgment

This research was partially supported by the Ministry of Education, Culture, Sports, Science and Technology-Japan, Grant-in-Aid for Scientific Research (C), 19510136, 2007-2009.

References

[1] M. Grabisch and F. Lange. Games on latices, multichoice games and the shapley value: a new approach. *mathematical Methods of Operations Research*, 65(1):153–167, 2007.

[2] J. M. .Bilbao. *Cooperative games on combinatorial syructures*. Kluwer academic publishers, Boston, 2000.

[3] R. Myerson. Graphs and cooperation in games. *Mathematics of Operations Research*, 2:225–229, 1977.

[4] P. Borm, G. Owen, and S. Tijs. On the position value for communication situuations. *SIAM Journal on Discrete Mathematics*, 3:305–320, 1992.

[5] G.-C. Rota. On the foundations of combinatorial theory I. theory of Möbius functions. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 2:340–368, 1964.

[6] K. Fujimoto. Interaction indices with respect to fuzzy measures (in japanese). *Journal of Japan Society of Fuzzy Theory and Intelligent Informatics*, 16(4):303–310, 2004.

[7] K. Fujimoto, I. Kojadinovic, and J.-L. Marichal. Axiomatic characterizations of probabilistic and cardinal-probabilistic interaction indices. *Games and Economic Behavior*, 55:72–99, 2006.

[8] J. Harsanyi. A simplified bargaining model for the n-person cooperative game. *International Economic Review*, 4:194–220, 1963.

[9] G. Hamiache. A value with incomplete information. *Games and Economic Behaviour*, 26:59–78, 1999.

[10] L. Shapley. A value for *n*-person games. *Contribution Theory of Games, II. Annals of Mathematics Studies*, 28:307–317, 1953.

ISBN: 978-989-95079-6-8

Ranking of fuzzy numbers, some recent and new formulas

Saeid Abbasbandy

Department of Mathematics, Science and Research Branch
Islamic Azad University, Tehran, 14778, Iran
Email: abbasbandy@yahoo.com

Abstract— *Ranking of fuzzy numbers plays a very important role in linguistic decision making and some other fuzzy application systems. Several strategies have been proposed for ranking of fuzzy numbers. Each of these techniques has been shown to produce non-intuitive results in certain cases. In this paper, some new approaches for ranking of trapezoidal fuzzy numbers are introduced.*

Keywords— Magnitude of fuzzy number, Parametric form of fuzzy number, Ranking of fuzzy numbers, Trapezoidal fuzzy number.

1 Introduction

Ranking of fuzzy numbers is an important component of the decision process in many applications. More than 30 fuzzy ranking indices have been proposed since 1976. In 1976 and 1977, Jain [1, 2] proposed a method using the concept of maximizing set to order the fuzzy numbers. Jain's method is that the decision maker considers only the right side membership function. A canonical way to extend the natural ordering of real numbers to fuzzy numbers was suggested by Bass and Kwakernaak [3] as early as 1977. Dubios and Prade 1978 [4], used maximizing sets to order fuzzy numbers. In 1979, Baldwin and Guild [5] indicated that these two methods have some disturbing disadvantages. Also, in 1980, Adamo [6] used the concept of α-level set in order to introduce α-preference rule. In 1981 Chang [7] introduced the concept of the preference function of an alternative. Yager in 1981 [8, 9] proposed four indices which may be employed for the purpose of ordering fuzzy quantities in $[0, 1]$. Bortolan and Degani have been compared and reviewed some of these ranking methods [10]. Chen and Hwang [11] thoroughly reviewed the existing approaches, and pointed out some illogical conditions that arise among them. Chen [12], Choobineh [13], Cheng [14] have presented some methods , and also more recently numerous ranking techniques have been proposed and investigated by Chu, Tsao [15] and Ma, Kandel and Friedman [16]. Nowadays many researchers have developed methods to compare and to rank fuzzy numbers. Some of those methods are counter-intuitive and non discriminating [18, 19, 20, 21, 22] and recently some methods based on different distance functions have been introduced for ranking of fuzzy numbers [23, 24, 25, 26, 27].

2 Preliminaries

Though there are a number of ways of defining fuzzy numbers, for the purposes of this paper we adopt the following definition, we will identify the name of the number with that of its membership function for simplicity. Throughout this paper, $\mathbb{R}$ stands for the set of all real numbers, E stands the set of fuzzy numbers, $u(x)$ for the membership function of every $u \in E$ and $x \in \mathbb{R}$.

Definition 2.1 *[28] A fuzzy number is a fuzzy set like $u : \mathbb{R} \to I = [0, 1]$ which satisfies:*

1. *u is upper semi-continuous,*
2. *$u(x) = 0$ outside some interval $[a, d]$,*
3. *There are real numbers $a,\ b$ such that $a \le b \le c \le d$ and*
   a. *$u(x)$ is monotonic increasing on $[a, b]$,*
   b. *$u(x)$ is monotonic decreasing on $[c, d]$,*
   c. *$u(x) = 1,\ b \le x \le c$.*

The membership function u can be expressed as

$$u(x) = \begin{cases} u_L(x), & a \le x \le b, \\ 1, & b \le x \le c, \\ u_R(x), & c \le x \le d, \\ 0, & otherwise, \end{cases}$$

where $u_L : [a, b] \longrightarrow [0, 1]$ and $u_R : [c, d] \longrightarrow [0, 1]$ are left and right membership functions of fuzzy number u. An equivalent parametric form is also given in [29] as follows:

Definition 2.2 *[29] A fuzzy number u in parametric form is a pair $(\underline{u}, \overline{u})$ of functions $\underline{u}(r), \overline{u}(r),\ 0 \le r \le 1$, which satisfy the following requirements:*

1. *$\underline{u}(r)$ is a bounded monotonic increasing left continuous function,*
2. *$\overline{u}(r)$ is a bounded monotonic decreasing left continuous function,*
3. *$\underline{u}(r) \le \overline{u}(r),\ 0 \le r \le 1$.*

The trapezoidal fuzzy number $u = (x_0, y_0, \sigma, \beta)$, with two defuzzifier x_0, y_0, and left fuzziness $\sigma > 0$ and right fuzziness $\beta > 0$ is a fuzzy set where the membership function is as

$$u(x) = \begin{cases} \dfrac{1}{\sigma}(x - x_0 + \sigma), & x_0 - \sigma \le x \le x_0, \\ 1 & x \in [x_0, y_0], \\ \dfrac{1}{\beta}(y_0 - x + \beta), & y_0 \le x \le y_0 + \beta, \\ 0, & otherwise, \end{cases}$$

and its parametric form is

$$\underline{u}(r) = x_0 - \sigma + \sigma r, \quad \overline{u}(r) = y_0 + \beta - \beta r.$$

ISBN: 978-989-95079-6-8

Let E_{TR} be the set of all trapezoidal fuzzy numbers on the real line. Provided that, $x_0 = y_0$ then u is a triangular fuzzy number, and we write $u = (x_0, \sigma, \beta)$. The support of fuzzy number u is defined as follows:

$$supp(u) = \overline{\{x \mid u(x) > 0\}},$$

where $\overline{\{x \mid u(x) > 0\}}$ is closure of set $\{x \mid u(x) > 0\}$.

The addition and scalar multiplication of fuzzy numbers are defined by the extension principle and can be equivalently represented in [30, 31, 32] as follows. For arbitrary $u = (\underline{u}, \overline{u})$, $v = (\underline{v}, \overline{v})$ we define addition $(u + v)$ and multiplication by scalar $k > 0$ as

$$(\underline{u+v})(r) = \underline{u}(r) + \underline{v}(r), \ \ (\overline{u+v})(r) = \overline{u}(r) + \overline{v}(r), \quad (1)$$

$$(\underline{ku})(r) = k\underline{u}(r), \ \ (\overline{ku})(r) = k\overline{u}(r). \quad (2)$$

To emphasis the collection of all fuzzy numbers with addition and multiplication as defined by (1) and (2) is denoted by E, which is a convex cone. The image (opposite) of $u = (x_0, y_0, \sigma, \beta)$, can be defined by $-u = (-y_0, -x_0, \beta, \sigma)$ (see [32, 33]).

Definition 2.3 *For arbitrary fuzzy numbers $u = (\underline{u}, \overline{u})$ and $v = (\underline{v}, \overline{v})$ the quantity*

$$D(u,v) = \left[\int_0^1 (\underline{u}(r) - \underline{v}(r))^2 \,\mathrm{dr} + \int_0^1 (\overline{u}(r) - \overline{v}(r))^2 \,\mathrm{dr}\right]^{1/2},$$

is the distance between u and v, [16, 17, 34]. The function $D(u, v)$ is a metric in E and (E, D) is a complete metric space.

The ordering indices are organized into three categories by Wang and Kerre [35] as follows:

1- **Defuzzification method**: Each index is associated with a mapping from the set of fuzzy quantities to the real line. In this case fuzzy quantities are compared according to the corresponding real numbers.

2- **Reference set method**: In this case a fuzzy set as a reference set is set up and all the fuzzy quantities to be ranked are compared with the reference set.

3- **Fuzzy relation method**: In this case a fuzzy relation is constructed to make pairwise comparisons between the fuzzy quantities involved.

Let M be an ordering method on E. The statement "two elements u and v in E satisfy that u has a higher ranking than v when M is applied" will be written as "$u \succ v$ by M". "$u \sim v$" and "$u \succeq v$" are similarly interpreted. The following reasonable properties for the ordering approaches are introduced by Wang and Kerre [35].

Reasonable properties (axioms)

A_1 : For an arbitrary finite subset Γ of E and $u \in \Gamma, u \succeq u$.

A_2 : For an arbitrary finite subset Γ of E and $(u, v) \in \Gamma^2, u \succeq v$ and $v \succeq u$, we should have $u \sim v$.

A_3 : For an arbitrary finite subset Γ of E and $(u, v, w) \in \Gamma^3, u \succeq v$ and $v \succeq w$, we should have $u \succeq w$.

A_4 : For an arbitrary finite subset Γ of E and $(u, v) \in \Gamma^2$, $\inf supp(u) \geq \sup supp(v)$, we should have $u \succeq v$.

A'_4 : For an arbitrary finite subset Γ of E and $(u, v) \in \Gamma^2$, $\inf supp(u) > \sup supp(v)$, we should have $u \succ v$.

A_5 : Let Γ and Γ' be two arbitrary finite subsets of E also u and v are in $\Gamma \cap \Gamma'$. We obtain the ranking order $u \succ v$ by M on Γ' if and only if $u \succ v$ by M on Γ.

A_6 : Let $u,\ v$, $u + w$ and $v + w$ be elements of E. If $u \succeq v$, then $u + w \succeq v + w$.

A'_6 : Let $u,\ v$, $u + w$ and $v + w$ be elements of E. If $u \succ v$, then $u + w \succ v + w$, when $w \neq 0$.

A_7 : Let $u,\ v,\ uw$, and vw be elements of E and $w \geq 0$. If $u \succeq v$ then $uw \succeq vw$.

3 Some new and recent methods

3.1 Method of D-distance

Let all of fuzzy numbers are positive or negative. Without less of generality assume that all of them are positive. The membership function of $a \in R$ is $u_a(x) = 1$, if $x = a$; and $u_a(x) = 0$, if $x \neq a$. Hence if $a = 0$ we have the following

$$u_0(x) = \begin{cases} 1 & x = 0, \\ 0 & x \neq 0. \end{cases}$$

Since $u_0(x) \in E$, left fuzziness σ and right fuzziness β are 0, so for each $u \in E$

$$D(u, u_0) = \left[\int_0^1 (\underline{u}(r)^2 + \overline{u}(r)^2)\mathrm{dr}\right]^{1/2}.$$

Thus we have the following definition.

Definition 3.1 *For u and $v \in E$, define the ranking of u and v by saying*

$$u > v \ \ iff \ \ d(u, u_0) > d(v, u_0),$$
$$u < v \ \ iff \ \ d(u, u_0) > d(v, u_0),$$
$$u \approx v \ \ iff \ \ d(u, u_0) = d(v, u_0).$$

Property 3.1. Suppose u and $v \in E$ are arbitrary there

(I) If $u = v$ then $u \approx v$.

(II) If $v \subseteq u$ and $\underline{u}(r)^2 + \overline{u}(r)^2 > \underline{v}(r)^2 + \overline{v}(r)^2$ for all $r \in [0, 1]$ then $v < u$.

Remark 3.1. (I) The distance triangular fuzzy number $u = (x_0, \sigma, \beta)$ of u_0 is defined as following

$$d(u, u_0) = \left[2x_0^2 + \sigma^2/3 + \beta^2/3 + x_0(\beta - \sigma)\right]^{1/2}.$$

(II) The distance trapezoidal fuzzy number $u = (x_0, y_0, \sigma, \beta)$ of u_0 is defined as following

$$d(u, u_0) = \left[2x_0^2 + \sigma^2/3 + \beta^2/3 - x_0\sigma + y_0\beta\right]^{1/2}.$$

(III) If $u \approx v$, it is not necessary that $u = v$. Since if $u \neq v$ and $(\underline{u}(r)^2 + \overline{u}(r)^2)^{1/2}$=$(\underline{v}(r)^2 + \overline{v}(r)^2)^{1/2}$ then $u \approx v$.

3.2 Method of min distance

In this subsection, we will propose the ranking of fuzzy numbers associated with the metric D in E.

Definition 3.2 *Let $\gamma = \{v_1, v_2, \ldots, v_n\} \subseteq E$ be is set of fuzzy numbers, we define a ranking on γ by min distance as*

$$v_i \succ v_j \ \text{ iff } \ D(v_i, \tilde{M}) > D(v_j, \tilde{M}),$$

ISBN: 978-989-95079-6-8

$$v_i \prec v_j \text{ iff } D(v_i, \tilde{M}) < D(v_j, \tilde{M}),$$
$$v_i \sim v_j \text{ iff } D(v_i, \tilde{M}) = D(v_j, \tilde{M}),$$

where $\tilde{M} = \tilde{min}\{v_1, \ldots, v_n\}$.

Dubois and Prade [30] present rules for computing $\tilde{min}$ and also comment on the properties of $\tilde{min}$.

Remark 3.2. The min distance, has the properties $A_1, A_2, \ldots, A_4'$.

3.3 Method of sign distance

Definition 3.3 *For arbitrary fuzzy numbers* $u = (\underline{u}, \overline{u})$ *and* $v = (\underline{v}, \overline{v})$*,the function*

$$D_p(u,v) = \left[\int_0^1 |\underline{u}(r) - \underline{v}(r)|^p \mathrm{dr} + \int_0^1 |\overline{u}(r) - \overline{v}(r)|^p \mathrm{dr}\right]^{1/p}, \quad (p \geq 1)$$

is the distance between u *and* v.

Definition 3.4 *Let* $\gamma : E \longrightarrow \{-1, 1\}$ *be a function that is defined as follows:*

$$\forall\, u \in E : \gamma(u) = sign\left[\int_0^1 (\underline{u}(r) + \overline{u}(r))\mathrm{dr}\right],$$

where

$$\gamma(u) = \begin{cases} 1 & \text{if } sign\left(\int_0^1 (\underline{u} + \overline{u})(r)\mathrm{dr}\right) \geq 0, \\ -1 & \text{if } sign\left(\int_0^1 (\underline{u} + \overline{u})(r)\mathrm{dr}\right) < 0. \end{cases}$$

Remark 3.3. (I) If $\inf supp(u) \geq 0$ or $\inf \underline{u}(r) \geq 0$ then $\gamma(u) = 1$.

(II) If $\sup supp(u) < 0$ or $\sup \overline{u}(r) < 0$ then $\gamma(u) = -1$.

Definition 3.5 *For* $u \in E$,

$$d_p(u, u_0) = \gamma(u) D_p(u, u_0),$$

is called sign distance.

Definition 3.6 *For* u *and* $v \in E$*, define the ranking of* u *and* v *by* d_p *on* E*, i.e.*

$$u \succ v \text{ iff } d_p(u, u_0) > d_p(v, u_0),$$
$$u \prec v \text{ iff } d_p(u, u_0) < d_p(v, u_0),$$
$$u \sim v \text{ iff } d_p(u, u_0) = d_p(v, u_0).$$

Remark 3.4. (I) The function d_p, sign distance, has the properties $A_1, A_2, \ldots, A_5$.

(II) The function d_p, sign distance, for $p = 1$ has the properties A_6, A_6' if

$$\inf\{supp(u), supp(v), supp(u+w), supp(v+w)\} \geq 0$$

or

$$\sup\{supp(u), supp(v), supp(u+w), supp(v+w)\} \leq 0.$$

(III) Suppose u and $v \in E$ are arbitrary, then

(a) If $u = v$ then $u \sim v$,

(b) If $v \subseteq u$ and $\gamma(u)\left(|\underline{u}(r)|^p + |\overline{u}(r)|^p\right) > \gamma(v)\left(|\underline{v}(r)|^p + |\overline{v}(r)|^p\right)$ for all $r \in [0,1]$ then $v \prec u$.

(IV) If $u \sim v$, it is not necessary that $u = v$. Since if $u \neq v$ and $\gamma(u)(|\underline{u}(r)|^p + |\overline{u}(r)|^p) = \gamma(v)(|\underline{v}(r)|^p + |\overline{v}(r)|^p)$ then $u \sim v$.

(V) If $u \preceq v$ then $-u \succeq -v$.

Therefore we can simply rank the fuzzy numbers by the defuzzification of $d_p(u, u_0)$. By Remark 3.4(V) we can logically infer ranking order of the images of the fuzzy numbers.

3.4 Method of H-distance

Definition 3.7 *A continuous function* $s : [0,1] \longrightarrow [0,1]$ *with the following properties is a* source *function*

1. $s(0) = 0$,
2. $s(1) = 1$,
3. $s(r)$ *is increasing.*
4. $\int_0^1 s(r)\mathrm{dr} = \frac{1}{2}$.

In fact, a reducing has the reflect of weighting the influence of the different r-cuts and diminishes the contribution of the lower r-levels. This is reasonable since these levels arises from values of membership function for which there is a considerable amount of uncertainty. For example, we can use $s(r) = r$.

Definition 3.8 *The Value and Ambiguity of a fuzzy number* $\widetilde{u}$ *are defined as follows, [36],*

$$Val_s(\widetilde{u}) = \int_0^1 s(r)[\overline{u}(r) + \underline{u}(r)]\mathrm{dr},$$
$$Amb_s(\widetilde{u}) = \int_0^1 s(r)[\overline{u}(r) - \underline{u}(r)]\mathrm{dr}.$$

Definition 3.9 *For* $\widetilde{u}, \widetilde{v} \in E$*, we define* ***H-distance*** *of* $\widetilde{u}$ *and* $\widetilde{v}$ *by*

$$D_H^s(\widetilde{u}, \widetilde{v}) = \frac{1}{2}\big\{\, |Val_s(\widetilde{u}) - Val_s(\widetilde{v})| + |Amb_s(\widetilde{u}) - Amb_s(\widetilde{v})| + d_H([\widetilde{u}]^1, [\widetilde{v}]^1)\big\},$$

where d_H *is the Hausdorff metric between intervals, and* $[.]^1$ *is the 1-cut representation of a fuzzy number.*

Property 3.2. The source distance, D_H^s, is a metric on E_{TR} and a pseudo-metric on E.

Remark 3.5. By the metric D_H^s and an arbitrary reference fuzzy set (like subsection 3.1 or 3.2), we can define a new ordering for fuzzy numbers.

3.5 Method of source distance

Definition 3.10 *For* $\widetilde{u}, \widetilde{v} \in E$*, we define* ***source distance*** *of* $\widetilde{u}$ *and* $\widetilde{v}$ *by*

$$D_s(\widetilde{u}, \widetilde{v}) = \frac{1}{2}\big\{\, |Val_s(\widetilde{u}) - Val_s(\widetilde{v})| + |Amb_s(\widetilde{u}) - Amb_s(\widetilde{v})| + \max\big\{\, |t_v - t_u|, |m_v - m_u|\big\}\big\},$$

where $[m_u, t_u]$ *and* $[m_v, t_v]$ *are the cores of fuzzy numbers* $\widetilde{u}$ *and* $\widetilde{v}$ *respectively.*

ISBN: 978-989-95079-6-8

Property 3.3. The source distance, D_s, is a metric on E_{TR} and a pseudo-metric on E.

Remark 3.6. By the metric D_s and an arbitrary reference fuzzy set (like subsection 3.1 or 3.2), we can define a new ordering for fuzzy numbers.

3.6 Method of magnitude

For an arbitrary trapezoidal fuzzy number $u = (x_0, y_0, \sigma, \beta)$, with parametric form $u = (\underline{u}(r), \overline{u}(r))$, we define the *magnitude* of the trapezoidal fuzzy number u as

$$Mag(u) = \frac{1}{2}\Big(\int_0^1 (\underline{u}(r) + \overline{u}(r) + x_0 + y_0) f(r) \mathrm{dr} \Big),$$

where the function $f(r)$ is a non-negative and increasing function on $[0, 1]$ with $f(0) = 0$, $f(1) = 1$ and $\int_0^1 f(r)\mathrm{dr} = \frac{1}{2}$. For example, we can use $f(r) = r$. The resulting scalar value is used to rank the fuzzy numbers. In the other words $Mag(u)$ is used to rank fuzzy numbers. The larger $Mag(u)$, the larger fuzzy number. Therefore for any two trapezoidal fuzzy numbers u and $v \in E$, we define the ranking of u and v by the $Mag(.)$ on E as follows:

1. $Mag(u) > Mag(v)$ if and only if $u \succ v$,
2. $Mag(u) < Mag(v)$ if and only if $u \prec v$,
3. $Mag(u) = Mag(v)$ if and only if $u \sim v$.

Then we formulate the order $\succeq$ and $\preceq$ as $u \succeq v$ if and only if $u \succ v$ or $u \sim v$, $u \preceq v$ if and only if $u \prec v$ or $u \sim v$. In the other words, this method is placed in the first class of Kerre's categories [35].

Remark 3.7. **(I)** If $\inf supp(u) \geq 0$ or $\inf \underline{u}(r) \geq 0$ then $Mag(u) \geq 0$.

(II) If $\sup supp(u) \leq 0$ or $\sup \overline{u}(r) \leq 0$ then $Mag(u) \leq 0$.

(III) For two arbitrary trapezoidal fuzzy numbers u and v, we have

$$Mag(u + v) = Mag(u) + Mag(v).$$

(IV) For all symmetric trapezoidal fuzzy numbers $u = (-x_0, x_0, \sigma, \sigma)$,

$$Mag(u) = 0.$$

(V) For any two symmetric trapezoidal fuzzy numbers $u = (x_0, y_0, \sigma, \sigma)$ and $v = (x_0, y_0, \beta, \beta)$,

$$Mag(u) = Mag(v).$$

Property 3.4. The function $Mag(.)$ has the properties $A_1, A_2, A_3, \ldots, A_6'$.

4 Conclusions

In spite of many ranking methods, no one can rank fuzzy numbers with human intuition consistently in all cases. The proposed methods can effectively rank various fuzzy numbers and their images. These methods have some mathematical properties. Moreover some pseudo metric on the set of fuzzy numbers and metric on trapezoidal fuzzy numbers are introduced. We may conclude that these ordering methods are relatively reasonable for fuzzy numbers based on the introduced axioms.

Acknowledgment

The author is thankful to the Islamic Azad University of Iran, Science and Research Branch, for the financial support.

References

[1] R. Jain. Decision-making in the presence of fuzzy variable. *IEEE Transactions on Systems, Man, and Cybernetics*, 6:698-703,1976.

[2] R. Jain. A procedure for multi-aspect decision making using fuzzy sets. *International Journal of Systems Science*, 8:1-7,1977.

[3] S. Bass and H. Kwakernaak. Rating and ranking of multiple-aspect alternatives using fuzzy sets. *Automatica*, 13:47-58,1977.

[4] D. Dubios and H. Prade. Oprations on fuzzy numbers. *International Journal of Systems Science*, 9:613-626,1978.

[5] J.F. Baldwin and N.C.F. Guild. Comparison of fuzzy numbers on the same decision space. *Fuzzy Sets and Systems*, 2:213-233,1979.

[6] M. Adamo. Fuzzy decision trees. *Fuzzy Sets and Systems*, 4:207-219,1980.

[7] W.K. Chang. Ranking of fuzzy utilities with triangular membership functions. *International Conference on Plicy Analysis and Informations Systems, Tamkang University*, 163-171,1981.

[8] R.R. Yager. On choosing between fuzzy subsets. *Kybernetes*, 9:151-154,1980.

[9] R.R. Yager. A procedure for ordering fuzzy subests of the unit interval. *Information Sciences*, 24:143-161,1981.

[10] G. Bortolan and R. Degani. A review of some methods for ranking fuzzy numbers. *Fuzzy Sets and Systems*, 15:1-19,1985.

[11] S.J. Chen and C.L. Hwang. *Fuzzy multiple Attribute Decision Making*. Spinger-Verlag, Berlin, 1992.

[12] S. Chen. Ranking fuzzy numbers with maximizing set and minimizing set. *Fuzzy Sets and Systems*, 17:113-129,1985.

[13] F. Choobineh and H. Li. An index for ordering fuzzy numbers. *Fuzzy Sets and Systems*, 54:287-294,1993.

[14] C.H. Cheng. A new approach for ranking fuzzy numbers by distance method. *Fuzzy Sets and Systems*, 95:307-317,1998.

[15] T. Chu and C.Tsao. Ranking fuzzy numbers with an area between the centroid point and orginal point. *Computers and Mathematics with Applications*, 43:111-117,2002.

[16] M. Ma and A. Kandel and M. Friedman. A new approach for defuzzification. *Fuzzy Sets and Systems*, 111:351-356,2000.

[17] S. Abbasbandy and B. Asady. A Note on "A new approach for defuzzification". *Fuzzy Sets and Systems*, 128:131-132,2002.

[18] S. Abbasbandy and M. Amirfakhrian. A new approach to universal approximation of fuzzy functions on a discrete set of points. *Applied Mathematical Modelling*, 30:1525-1534,2006.

[19] S. Abbasbandy and M. Amirfakhrian. The nearest trapezoidal form of a generalized left right fuzzy number. *International Journal of Approximate Reasoning*, 43:166-178,2006.

[20] S. Abbasbandy and B. Asady. Ranking of fuzzy numbers by sign distance. *Information Sciences*, 176:2405-2416,2006.

[21] L.H. Chen and H.W. Lu. An approximaite approach for ranking fuzzy numbers based on left and right dominance. *Computers and Mathematics with Applications*, 41:1589-1602,2001.

[22] J. Yao and K. Wu. Ranking fuzzy numbers based on decomposition principle and signed distance. *Fuzzy Sets and Systems*, 116:275-288,2000.

[23] S. Abbasbandy and C. Lucas and B. Asady. A new method for ranking of fuzzy numbers through using distance method. *Quarterly Journal of Science, Teacher Training University*, 3:83-88,2003.

[24] G. Facchinetti and R.G. Ricci. A characterization of a general class of ranking functions on triangular fuzzy numbers. *Fuzzy Sets and Systems*, 146:297-312,2004.

[25] S. Abbasbandy and M. Otadi and M. Mosleh. Ranking of fuzzy numbers by min distance. *Mathematics Scientific Journal, Islamic Azad University of Arak*, 2:1-14,2006-2007.

[26] S. Abbasbandy and M. Amirfakhrian. The nearest approximation of a Fuzzy quantity in parametric form. *Applied Mathematics and Computation*, 172:624-632,2006.

[27] S. Abbasbandy and T. Hajjari. A new approach for ranking of trapezoidal fuzzy numbers. *Computers and Mathematics with Applications*, 57:413-419,2009.

[28] R. Goetschel and W. Voxman. Elementary calculus. *Fuzzy Sets and Systems*, 18:31-43,1986.

[29] M. Ma and M. Friedman and A. Kandel. A new fuzzy arithmetic. *Fuzzy Sets and Systems*, 108:83-90,1999.

[30] D. Dubois and H. Prade. *Fuzzy Sets and System: Theory and Application*. Academic Press, New York, 1980.

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

[32] H.J. Zimmermann. *Fuzzy Sets Theory and its Application*. Kluwer Academic Press, Dordrecht, 1991.

[33] A. Kauffman and M.M. Gupta. *Introduction to fuzzy arithmetic: Theory and Application*. Van Nostrand Reinhold, New York, 1991.

[34] M. Ma and A. Kandel and M. Friedman. Correction to "A New approach for defuzzification". *Fuzzy Sets and Systems*, 128:133-134,2002.

[35] X. Wang and E.E. Kerre. Reasonable properties for the ordering of fuzzy quantities (I,II). *Fuzzy Sets and Systems*, 118:375-405,2001.

[36] W. Voxman. Some remarks on distance between fuzzy numbers. *Fuzzy Sets and Systems*, 100:353-365,1998.

ISBN: 978-989-95079-6-8

Lattice properties of discrete fuzzy numbers under extended min and max

Jaume Casasnovas Juan Vicente Riera
Dept. Mathematics and Computer Science ,University of the Balearic Islands
Palma de Mallorca, Spain
Email: {jaume.casasnovas,jvicente.riera}@uib.es

Abstract— *This paper proposes to study the lattice properties of two closed binary operations in the set of discrete fuzzy numbers. Using these operations to represent the meet and the join, we prove that the set of discrete fuzzy numbers whose support is a set of consecutive natural numbers is a distributive lattice. Finally, we demonstrate that the subsets of discrete fuzzy numbers, which have the same support, are distributive lattices too.*

Keywords— Fuzzy numbers, discrete fuzzy numbers, distributive lattice.

1 Introduction

It is possible to approach the theory of fuzzy numbers in different directions: theoretical[8, 9, 10, 13], geometric [1, 2], applications in engineering [13], social science [12], lattice theory[13, 20], etc.. Voxman [18] introduced the concept of a discrete fuzzy number as a fuzzy subset of real numbers with discrete support and analogous properties to a fuzzy number (convexity, normality). Also, like fuzzy numbers, it is possible to consider discrete fuzzy numbers from different points of view: theoretical [16, 18], applications in engineering [11, 19], social sciences [17], etc.

It is well known that, the arithmetic and lattice operations such as maximum and minimum on fuzzy numbers can be approached either by the direct use of the membership function (by the Zadeh's extension principle) or by the equivalent use of the *r-cuts* representation, for instance, [13, 14, 20].

Nevertheless, in the discrete case using the same methods, this process can yield a fuzzy subset that does not satisfy the conditions to be a discrete fuzzy number [3, 19]. In previous works [3, 4], we have presented a technique that allows us to obtain a closed addition on the set of discrete fuzzy numbers, DFN, and moreover, we focus on the addition of discrete fuzzy numbers whose support is an arithmetic sequence and even a subset of consecutive natural numbers. This type of numbers arise mainly when a fuzzy cardinality of a fuzzy set [5, 6] or a fuzzy multiset [15] is considered.

In [7] we define two closed binary operations in the set of discrete fuzzy numbers to obtain the maximum and the minimum of discrete fuzzy numbers. We prove as well that in the set $\mathcal{A}_1$, of discrete fuzzy numbers whose support is a set of consecutive natural numbers, these operations coincide with the function maximum and minimum obtained using the Zadeh's extension principle.

The aim of this paper is to continue studying the properties of these operations and if it is possible to obtain a structure of lattice using them. We will see that, in general, these binary operations only fulfill the associative, commutative and idempotent laws. We show as well that the triplets ($\mathcal{A}_1$,MAX_w,MIN_w) and (F_A,MIN_w,MAX_w) are distributive lattices, where F_A is the subset of discrete fuzzy numbers whose support is the support of A, with $A \in DFN$.

2 Preliminaries

In this section, we recall some definitions and the main results about discrete fuzzy numbers which will be used later.

Definition 2.1 *[13] A fuzzy subset u of $\mathbb{R}$ with membership mapping $u : \mathbb{R} \to [0,1]$ is called* fuzzy number *if its* support *is an interval $[a,b]$ and there exist real numbers s,t with $a \leq s \leq t \leq b$ and such that:*

1. *$u(x)$=1 with $s \leq x \leq t$*
2. *$u(x) \leq u(y)$ with $a \leq x \leq y \leq s$*
3. *$u(x) \geq u(y)$ with $t \leq x \leq y \leq b$*
4. *$u(x)$ is upper semi-continuous.*

We will denote the set of fuzzy numbers by FN.

Definition 2.2 *[18] A fuzzy subset u of $\mathbb{R}$ with membership mapping $u : \mathbb{R} \to [0,1]$ is called* discrete fuzzy number *if its support is finite, i.e., there are $x_1, ..., x_n \in \mathbb{R}$ with $x_1 < x_2 < ... < x_n$ such that $supp(u) = \{x_1, ..., x_n\}$, and there are natural numbers s,t with $1 \leq s \leq t \leq n$ such that:*

1. *$u(x_i)$=1 for any natural number and i with $s \leq i \leq t$ (*core*)*
2. *$u(x_i) \leq u(x_j)$ for each natural number i,j with $1 \leq i \leq j \leq s$*
3. *$u(x_i) \geq u(x_j)$ for each natural number i,j with $t \leq i \leq j \leq n$*

From now on, we will denote the set of discrete fuzzy numbers by DFN and $DFN(\mathbb{N})$ will stand for the set of discrete fuzzy numbers whose support is a subset of the set of Natural Numbers. Finally, a discrete fuzzy number will be denoted by *dfn*.

In general, the operations on fuzzy numbers f, g can be approached either by the direct use of their membership function, $\mu_f(x), \mu_g(x)$, as fuzzy subsets of $\mathbb{R}$ and the Zadeh's extension principle:

$$O(f,g)(z) = sup\{\mu_f(x) \wedge \mu_g(y) | O(x,y) = z\}$$

or by the equivalent use of the α-*cuts* representation[13]:

ISBN: 978-989-95079-6-8

$$O(f,g)^{\alpha} = O(f^{\alpha}, g^{\alpha}) = \{O(x,y) | x \in f^{\alpha}, y \in g^{\alpha}\}$$

and

$$O(f,g)(z) = sup\{\alpha \in [0,1] | z \in O(f,g)^{\alpha}\}$$

Nevertheless, in the discrete case, this process can yield a fuzzy subset that does not satisfy the conditions to be a discrete fuzzy number [3, 19]. For example, let $u = \{0.3/1, 1/3, 0.5/7\}$ and $v = \{0.4/2, 1/5, 1/6, 0.8/9\}$ be two discrete fuzzy numbers. If we use the Zadeh's extension principle to obtain their addition, it results the fuzzy subset $S = \{0.3/3, 0.4/5, 0.3/6, 0.3/7, 1/8, 1/9, 0.3/10, 0.8/12, 0.5/13, 0.5/16\}$ which doesn't fulfill the conditions to be a discrete fuzzy number, because the third property of the definition 2.2 fails. In a previous work [3, 4] we have presented an approach to a closed extended addition ($\oplus$) of discrete fuzzy numbers after associating suitable non-discrete fuzzy numbers, which can be used like a carrier to obtain the desired addition.
In a recent paper [4] we proved that a suitable carrier can be a discrete fuzzy number whose support is an arithmetic sequence and even a subset of consecutive natural numbers. Thus, we obtained the following results:

Proposition 2.3 *[4]*
Let $\mathcal{A}_r$ be the set $\{f \in DFN\ (\mathbb{N})$, such that supp(f) is the set of terms of an arithmetic sequence with r as common difference$\}$. If $f, g \in \mathcal{A}_r$. The following facts:

1. $f \oplus g \in DFN(\mathbb{N})$
2. $f \oplus g \in \mathcal{A}_r$

hold.

Remark 2.4 *[4] Note that the set $\mathcal{A}_1$ is the set of discrete fuzzy numbers whose support is a set of consecutive natural numbers.*

Finally, we will use a kind of representation in the study of discrete fuzzy numbers:

Theorem 2.5 *[19] Let u be a dfn and let u^r be the r-cut $=\{x \in \mathbb{R} | u(x) \geq r\}$ for any $r \in (0,1]$. Let's u^0 denote the support of u. Then the following statements (1)-(4) hold:*

1. *u^r is a nonempty finite subset of $\mathbb{R}$, for any $r \in [0,1]$*
2. *$u^{r_2} \subset u^{r_1}$ for any $r_1, r_2 \in [0,1]$ with $0 \leq r_1 \leq r_2 \leq 1$*
3. *For any $r_1, r_2 \in [0,1]$ with $0 \leq r_1 \leq r_2 \leq 1$, if $x \in u^{r_1} - u^{r_2}$ we have $x < y$ for all $y \in u^{r_2}$, or x>y for all $y \in u^{r_2}$*
4. *For any $r_0 \in [0,1]$, there exist some real numbers $r_0^{'}$ with $0 < r_0' < r_0$ such that $u^{r_0'} = u^{r_0}$ (i.e. $u^r = u^{r_0}$ for any $r \in [r_0', r_0]$).*

And conversely, if for any $r \in [0,1]$, there exist $A^r \subset \mathbb{R}$ satisfying the following conditions (1)-(4):

1. *A^r is a nonempty finite for any $r \in [0,1]$*
2. *$A^{r_2} \subset A^{r_1}$, for any $r \in [0,1]$ with $0 \leq r_1 \leq r_2 \leq 1$*
3. *For any $r_1, r_2 \in [0,1]$ with $0 \leq r_1 \leq r_2 \leq 1$, if $x \in A^{r_1} - A^{r_2}$ we have $x < y$ for all $y \in A^{r_2}$, or x>y for all $y \in A^{r_2}$*
4. *For any $r_0 \in [0,1]$, there exists a real number $r_0^{'}$ with $0 < r_0^{'} < r_0$ such that $A^{r_0'} = A^{r_0}$ (i.e. $A^r = A^{r_0}$, for any $r \in [r_0', r_0]$)*

then there exists a unique $u \in DFN$ such that $u^r = A^r$ for any $r \in [0,1]$.

3 Maximum and Minimum of discrete fuzzy numbers

It's well known, for example [13], that the set of fuzzy numbers is a distributive lattice using the following operations

$$MIN(u,v)(z) = \sup_{z=\min(x,y)} \min(u(x), v(y)), \forall z \in \mathbb{R} \quad (1)$$

$$MAX(u,v)(z) = \sup_{z=\max(x,y)} \min(u(x), v(y)), \forall z \in \mathbb{R} \quad (2)$$

for each couple $u, v \in FN$. If we use the same operations to obtain a similar structure in the set of discrete fuzzy numbers we see that this result is not possible. For instance, if we consider the discrete fuzzy numbers $u = \{0.3/1, 0.4/3, 1/4\}$ and $v = \{0.5/2, 1/5, 1/6\}$ and we use the previous definition to calculate the $MIN(u,v)$, we obtain the fuzzy subset $M = \{0.3/1, 0.5/2, 0.4/3, 1/4\}$ which doesn't satisfy the conditions of the definition 2.2. In [7], the authors study this drawback and we propose a new method to calculate them. Using this method, we will see later that, it is possible to provide different subsets of the set DFN with a structure of distributive lattice.

Definition 3.1 *[7] Let u, v be two dfn. For each $\alpha \in [0,1]$, let's consider the α-cut sets: $u^{\alpha} = \{x_1^{\alpha}, ..., x_p^{\alpha}\}, v^{\alpha} = \{y_1^{\alpha}, ..., y_k^{\alpha}\}$ for u and v respectively and the set $supp(u) \bigvee supp(v) = \{x \vee y | x \in supp(u), y \in supp(v)\}$. Let's define the set:*

$$A^{\alpha} = \{z \in supp(u) \bigvee supp(v) \text{ such that}$$

$$\min(u^{\alpha} \vee v^{\alpha}) \leq z \leq \max(u^{\alpha} \vee v^{\alpha})\} =$$

$$= \{z \in supp(u) \bigvee supp(v) \text{ such that}$$

$$(\min u^{\alpha} \vee \min v^{\alpha}) \leq z \leq (\max u^{\alpha} \vee \max v^{\alpha})\}$$

i.e.:

$$A^{\alpha} = \{z \in supp(u) \bigvee supp(v) | (x_1^{\alpha} \vee y_1^{\alpha}) \leq z \leq (x_p^{\alpha} \vee y_k^{\alpha})\}$$

Proposition 3.2 *[7] The finite set A^{α}, as defined above, satisfies the properties 1,2,3 and 4 of theorem 2.5 and a discrete fuzzy number, $MAX_w(u,v)$, whose α-cuts are the finite set A^{α} exists and $MAX_w(u,v)(z) = \sup\{\alpha \in [0,1]$ such that $z \in A^{\alpha}\}$.*

Proposition 3.3 *[7] If $u, v \in \mathcal{A}_1$, then $MAX(u,v)$, defined through the extension principle, coincides with $MAX_w(u,v)$. So, if $u, v \in \mathcal{A}_1$, $MAX(u,v)$ is a discrete fuzzy number and $MAX(u,v) \in \mathcal{A}_1$.*

ISBN: 978-989-95079-6-8

Definition 3.4 *[7] Let u, v be two dfn. For each $\alpha \in [0,1]$, let's consider the α-cut sets: $u^\alpha = \{x_1^\alpha, ..., x_p^\alpha\}$,$v^\alpha = \{y_1^\alpha, ..., y_k^\alpha\}$ for u and v respectively and the set $supp(u) \bigwedge supp(v) = \{x \wedge y | x \in supp(u), y \in supp(v)\}$. Let's define the set:*

$$B^\alpha = \{z \in supp(u) \bigwedge supp(v) \text{ such that}$$

$$\min(u^\alpha \wedge v^\alpha) \leq z \leq \max(u^\alpha \wedge v^\alpha)\} =$$

$$= \{z \in supp(u) \bigwedge supp(v) \text{ such that}$$

$$(\min u^\alpha \wedge \min v^\alpha) \leq z \leq (\max u^\alpha \wedge \max v^\alpha)\}$$

i.e.:

$$B^\alpha = \{z \in supp(u) \bigwedge supp(v) | (x_1^\alpha \wedge y_1^\alpha) \leq z \leq (x_p^\alpha \wedge y_k^\alpha)\}$$

Remark 3.5 *In general, if "$*$" is a binary operation the equalities*

- $\min(u^\alpha * v^\alpha) = \min u^\alpha * \min v^\alpha$
- $\max(u^\alpha * v^\alpha) = \max u^\alpha * \max v^\alpha$

do not hold. For example, if we consider the sets $u^\alpha = \{-10, -9, -8, -7, -6, -1\}$, $v^\alpha = \{0, 1, 2, 3, 4\}$ and the usual product of real numbers as binary operation then $\min(u^\alpha \cdot v^\alpha) = -40$ and $\min(u^\alpha) \cdot \min(v^\alpha) = -10 \cdot 0 = 0$, and $\max(u^\alpha \cdot v^\alpha) = 0$ and $\max(u^\alpha) \cdot \max(v^\alpha) = -1 \cdot 4 = -4$

Proposition 3.6 *[7] The finite set B^α, as defined above, satisfies the properties 1,2,3 and 4 of Proposition 2.5 and a discrete fuzzy number, $MIN_w(u, v)$, whose α-cuts are the finite set B^α exists and $MIN_w(u, v)(z) = \sup\{\alpha \in [0, 1]$ such that $z \in B^\alpha\}$.*

Proposition 3.7 *[7] If $u, v \in \mathcal{A}_1$, then $MIN(u, v)$, defined through the extension principle, coincides with $MIN_w(u, v)$. So, if $u, v \in \mathcal{A}_1$, $MIN(u, v)$ is a discrete fuzzy number and $MIN(u, v) \in \mathcal{A}_1$.*

We have seen, in the previous propositions, that the operations $MAX_w(u, v)$ and $MIN_w(u, v)$ are discrete fuzzy numbers.

Example 3.8 *If we use the method, as explained above, to calculate the minimum of $u = \{0.3/1, 0.4/3, 1/4\}$ and $v = \{0.5/2, 1/5, 1/6\}$, we obtain the following discrete fuzzy number $MIN_w(u, v) = \{0.3/1, 0.5/2, 0.5/3, 1/4\}$ where*

$$B^{0.3} = \{z \in \{1, 2, 3, 4\} \text{ such that } 1 \leq z \leq 4\} = \{1, 2, 3, 4\}$$

$$B^{0.4} = \{z \in \{1, 2, 3, 4\} \text{ such that } 2 \leq z \leq 4\} = \{2, 3, 4\}$$

$$B^{0.5} = \{z \in \{1, 2, 3, 4\} \text{ such that } 2 \leq z \leq 4\} = \{2, 3, 4\}$$

$$B^1 = \{z \in \{1, 2, 3, 4\} \text{ such that } 4 \leq z \leq 4\} = \{4\}$$

Proposition 3.9 *Let A, B and C be three dfn such that their supports are $supp(A)$,$supp(B)$ and $supp(C)$ respectively. The following properties hold:*

1. *Commutativity*

$$supp(A) \bigwedge supp(B) = supp(B) \bigwedge supp(A)$$

and

$$supp(A) \bigvee supp(B) = supp(B) \bigvee supp(A)$$

2. *Associativity*

$$(supp(A) \bigwedge supp(B)) \bigwedge supp(C) =$$

$$= supp(A) \bigwedge (supp(B) \bigwedge supp(C))$$

and

$$(supp(A) \bigvee supp(B)) \bigvee supp(C) =$$

$$= supp(A) \bigvee (supp(B) \bigvee supp(C))$$

3. *Idempotence*

$$supp(A) \bigwedge supp(A) = supp(A)$$

and

$$supp(A) \bigvee supp(A) = supp(A)$$

4. *If the supports of A and B are the same or $supp(B) \subseteq supp(A)$ or $A, B \in \mathcal{A}_1$ then the next absorption laws*

$$supp(A) \bigwedge (supp(A) \bigvee supp(B)) = supp(A)$$

$$supp(A) \bigvee (supp(A) \bigwedge supp(B)) = supp(A)$$

hold.

5. *If the supports of A, B and C are the same or $A, B, C \in \mathcal{A}_1$ then the following distributive properties*

$$supp(A) \bigwedge (supp(B) \bigvee supp(C)) =$$

$$= (supp(A) \bigwedge supp(B)) \bigvee (supp(A) \bigwedge supp(C))$$

and

$$supp(A) \bigvee (supp(B) \bigwedge supp(C)) =$$

$$= (supp(A) \bigvee supp(B)) \bigwedge (supp(A) \bigvee supp(C))$$

hold.

6. *For any $A, B \in DFN$ then*

$$supp(A) \bigwedge (supp(B) \bigvee supp(C))) \subseteq$$

$$\subseteq (supp(A) \bigwedge supp(B)) \bigvee (supp(A) \bigwedge supp(C))$$

and

$$supp(A) \bigvee (supp(B) \bigwedge supp(C))) \subseteq$$

$$\subseteq (supp(A) \bigvee supp(B)) \bigwedge (supp(A) \bigvee supp(C))$$

7. *For any $A, B \in DFN$ then*

$$supp(A) \subseteq supp(A) \bigwedge (supp(A) \bigvee supp(B))$$

$$supp(A) \subseteq supp(A) \bigvee (supp(A) \bigwedge supp(B))$$

ISBN: 978-989-95079-6-8

Proof From the definition 2.2, the support of a discrete fuzzy number is a finite linearly ordered subset of real numbers. Moreover, it is well known that the set of real numbers is a distributive lattice with the usual operations maximum(max) and minimum(min). Let X, Y and Z denote the supports of $A, B, C \in DFN$ respectively.

1. Using the commutative property of the real functions maximum and minimum the proof is trivial.

2. Associativity

   If $z \in (X \bigwedge Y) \bigwedge Z$ then $z = \min(x, c)$ where $x = \min(a, b)$, $a \in X$, $b \in Y$ and $c \in Z$. So $z = \min(\min(a, b), c)$. Using the associativity of the function minimum, $z = \min(\min(a, b), c) = \min(a, \min(b, c))$. Then $z \in X \bigwedge (Y \bigwedge Z)$. Therefore

   $$(X \bigwedge Y) \bigwedge Z \subseteq X \bigwedge (Y \bigwedge Z)$$

   If $z \in X \bigwedge (Y \bigwedge Z)$ then $z = \min(a, x)$ where $x = \min(b, c)$, $a \in X$, $b \in Y$ and $c \in Z$. So $z = \min(a, \min(b, c))$. Using the associativity of the function minimum, $z = \min(a, \min(b, c)) = \min(\min(a, b), c)$. Then $z \in (X \bigwedge Y) \bigwedge Z$. Therefore

   $$X \bigwedge (Y \bigwedge Z) \subseteq (X \bigwedge Y) \bigwedge Z$$

   The proof of the other associative law is analogous.

3. Idempotence

   If $z \in X \bigwedge X$ then $z = \min(a, a')$ with $a, a' \in X$. Therefore $z = a \in X$ or $z = a' \in X$. This means that $z \in X$ and so $X \bigwedge X \subseteq X$. On the other hand, it is evident that $X \subseteq X \bigwedge X$ because the function minimum is idempotent and then for each $z \in X$, $z = \min(z, z)$.

   The proof of the other idempotence law is similar.

4. Absorption

   If $z \in X \bigwedge (X \bigvee Y)$ then $z = \min(a, \max(a', b))$ where $a, a' \in X$ and $b \in Y$. Then if $z = a$ or $z = a'$ obviously $z \in X$. But if $z = b$ we have that $a' \leq b \leq a$. Using the hypothesis of the proposition we obtain that $b \in X$. Therefore $X \bigwedge (X \bigvee Y) \subseteq X$.

   If $z \in X$ then $z = \min(z, \max(z, b))$ for all $b \in Y$. Because if $\max(z, b) = b$ then $\min(z, b) = z$. And, if $\max(z, b) = z$ then $\min(z, z) = z$. So, $X \subseteq X \bigwedge (X \bigvee Y)$

5. Distributivity

   Let's $A, B, C \in \mathcal{A}_1$ where $supp(A) = \{x_1, \cdots, x_n\}$,$supp(B) = \{y_1, \cdots, y_m\}$ and $supp(C) = \{z_1, \cdots, z_k\}$ respectively. Then

   $$supp(A) \bigwedge (supp(B) \bigvee supp(C))$$

   is the set of natural numbers z such that they belong to the set

   $$\{x_1, \cdots, x_n\} \bigwedge \{\max(y_1, z_1), \cdots, \max(y_m, z_k)\}$$

   , i.e., it's the set of natural numbers z such that $z \in \{\min(x_1, \max(y_1, z_1)), \cdots, \min(x_n, \max(y_m, z_k))\} =$ (using the distributive property of natural numbers with the usual order)

   $$= \{z \in \{\max(\min(x_1, y_1), \min(x_1, z_1)), \cdots$$
   $$\cdots, \max(\min(x_n, y_m), \min(x_n, z_k))\}\}$$

   Where the previous set is an interval of consecutive natural numbers. On the other hand,

   $$(supp(A) \bigwedge supp(B)) \bigvee (supp(A) \bigwedge supp(C)) =$$
   $$= \{z \in \{\{\min(x_1, y_1), \cdots, \min(x_n, y_m)\} \bigvee$$
   $$\{\min(x_1, z_1), \cdots, \min(x_n, z_k)\}\} =$$
   $$= \{z \in \{\max(\min(x_1, y_1), \min(x_1, z_1)), \cdots$$
   $$\cdots, \max(\min(x_n, y_m), \min(x_n, z_k))\}.$$

   The proof of the second distributive law is analogous.

6. If $z \in X \bigwedge (Y \bigvee Z))$ then $z = \min(a, \max(b, c))$ where $a \in X$, $b \in Y$ and $c \in Z$. Using the distributive property of the minimum and the maximum of real numbers, $z = \max(\min(a, b), \min(a, c))$. Then we obtain that $z \in (X \bigwedge Y) \bigvee (X \bigwedge Z)$ and so $X \bigwedge (Y \bigvee Z)) \subseteq (X \bigwedge Y) \bigvee (X \bigwedge Z)$.

   Analogously for the other inclusion.

7. If $z \in X$ then $z = \max(z, \min(z, b))$ for all $b \in Y$. Because if $\min(z, b) = b$ then $\max(z, b) = z$. And, if $\min(z, b) = z$ then $\max(z, z) = z$. So, $X \subseteq X \bigvee (X \bigwedge Y)$

Remark 3.10 *In general, the absorption and the distributive laws of the supports do not hold. For example, if* $X = \{4, 7, 9\}$,$Y = \{4, 6, 7\}$ *and* $Z = \{8, 9, 10\}$ *represent the supports of the discrete fuzzy numbers* A, B *and* C *respectively, then*

a) $X \bigwedge (X \bigvee Y) = \{4, 6, 7, 9\}$ *but* $X = \{4, 7, 9\}$.

b) $X \bigwedge (Y \bigvee Z) = \{4, 7, 8, 9\}$, *but* $(X \bigwedge Y) \bigvee (X \bigwedge Z) = \{4, 6, 7, 8, 9\}$.

The following theorem establishes some properties of the binary operations MAX_w and MIN_w using the same conditions and results that the previous proposition 3.9:

Theorem 3.11 *Let* MAX_w *and* MIN_w *be the binary operations on* DFN *defined by the propositions 3.2 and 3.6, respectively. Then, for any,* $A, B, C \in DFN$ *the following properties hold:*

1. Commutativity: $MIN_w(A, B) = MIN_w(B, A)$, $MAX_w(A, B) = MAX_w(B, A)$

2. Associativity:

$$MIN_w(MIN_w(A, B), C) =$$
$$= MIN_w(A, MIN_w(B, C))$$

and

$$MAX_w(MAX_w(A, B), C) =$$
$$= MAX_w(A, MAX_w(B, C))$$

ISBN: 978-989-95079-6-8

3. Idempotence: $\begin{array}{l} MIN_w(A,A)=A \\ MAX_w(A,A)=A \end{array}$

If the supports of A and B are the same or $A,B,C \in \mathcal{A}_1$ then

4. Absorption: $\begin{array}{l} MIN_w(A, MAX_w(A,B))=A \\ MAX_w(A, MIN_w(A,B))=A \end{array}$

5. Distributivity:

$$MIN_w(A, MAX_w(B,C)) =$$
$$= MAX_w(MIN_w(A,B), MIN_w(A,C))$$

And

$$MAX_w(A, MIN_w(B,C)) =$$
$$= MIN_w(MAX_w(A,B), MAX_w(A,C))$$

Proof Let A, B and C be three dfn. Let's consider the α-cut sets: $A^\alpha = \{x_1^\alpha, ..., x_p^\alpha\}$, $B^\alpha = \{y_1^\alpha, ..., y_k^\alpha\}$, $C^\alpha = \{w_1^\alpha, ..., w_l^\alpha\}$ for A, B and C respectively.

1. We want to show that $MIN_w(A,B) = MIN_w(B,A)$

It is enough to prove that the discrete fuzzy numbers $MIN_w(A,B)$ and $MIN_w(B,A)$ are the same α-cut sets for each $\alpha \in [0,1]$.

By definition 3.4

$$MIN_w(A,B)^\alpha =$$
$$= \{z \in supp(A) \bigwedge supp(B) \text{ such that}$$
$$(\min A^\alpha \wedge \min B^\alpha) \le z \le (\max A^\alpha \wedge \max B^\alpha)\} =$$
$$= \{z \in supp(A) \bigwedge supp(B) | (x_1^\alpha \wedge y_1^\alpha) \le z \le (x_p^\alpha \wedge y_k^\alpha)\} =$$
$$= \{z \in supp(A) \bigwedge supp(B) | (y_1^\alpha \wedge x_1^\alpha) \le z \le (y_p^\alpha \wedge x_k^\alpha)\} =$$
$$= MIN_w(B,A)^\alpha$$

Analogously for the other commutative law.

2. We want to see that

$$MIN_w(MIN_w(A,B),C) = MIN_w(A, MIN_w(B,C))$$

By definition 3.4

$$MIN_w(MIN_w(A,B),C)^\alpha =$$
$$= \{z \in supp(MIN_w(A,B)) \bigwedge supp(C) \text{ such that}$$
$$\min MIN_w(A,B)^\alpha \wedge \min C^\alpha \le z \le \max MIN_w(A,B)^\alpha \wedge \max C^\alpha\} =$$
$$= \{z \in supp(MIN_w(A,B)) \bigwedge supp(C) \text{ such that}$$
$$(x_1^\alpha \wedge y_1^\alpha) \wedge w_1^\alpha \le z \le (x_p^\alpha \wedge y_k^\alpha) \wedge w_l^\alpha\} =$$
$$= \{z \in (supp(A) \bigwedge supp(B)) \bigwedge supp(C) \text{ such that}$$
$$(x_1^\alpha \wedge y_1^\alpha) \wedge w_1^\alpha \le z \le (x_p^\alpha \wedge y_k^\alpha) \wedge w_l^\alpha\} =$$
$$= \{z \in (supp(A) \bigwedge supp(B)) \bigwedge supp(C) \text{ such that}$$
$$x_1^\alpha \wedge (y_1^\alpha \wedge w_1^\alpha) \le z \le x_p^\alpha \wedge (y_k^\alpha \wedge w_l^\alpha\}) =$$
$$= \{z \in supp(A) \bigwedge (supp(B) \bigwedge supp(C)) \text{ such that}$$
$$x_1^\alpha \wedge (y_1^\alpha \wedge w_1^\alpha) \le z \le x_p^\alpha \wedge (y_k^\alpha \wedge w_l^\alpha\}) =$$
$$= MIN_w(A, MIN_w(B,C))^\alpha$$

The proof of the other associative law is similar.

3. We want to demonstrate that

$$MIN_w(A,A) = A$$

By definition 3.4

$$MIN_w(A,A)^\alpha =$$
$$= \{z \in supp(A) \bigwedge supp(A) \text{ such that}$$
$$(\min A^\alpha \wedge \min A^\alpha) \le z \le (\max A^\alpha \wedge \max A^\alpha)\} =$$
$$= \{z \in supp(A) | (x_1^\alpha \wedge x_1^\alpha) \le z \le (x_p^\alpha \wedge x_p^\alpha)\} =$$
$$= \{z \in supp(A) | x_1^\alpha \le z \le x_p^\alpha\} =$$
$$= A^\alpha$$

It is the same for the other idempotence law.

4. We want to demonstrate that

$$MIN_w(A, MAX_w(A,B)) = A$$

By definition 3.4

$$MIN_w(A, MAX_w(A,B))^\alpha =$$
$$= \{z \in supp(A) \bigwedge supp(MAX_w(A,B)) \text{ such that}$$
$$\min A^\alpha \wedge \min MAX_w(A,B)^\alpha \le z \le \max A^\alpha \wedge \max MAX_w(A,B)^\alpha\} =$$
$$= \{z \in supp(A) \bigwedge supp(MAX_w(A,B)) \text{ such that}$$
$$x_1^\alpha \wedge (x_1^\alpha \vee y_1^\alpha) \le z \le x_p^\alpha \wedge (x_p^\alpha \vee y_k^\alpha\}) =$$

(using the absorption law of real numbers)

$$= \{z \in supp(A) \bigwedge supp(MAX_w(A,B)) \text{ such that}$$
$$x_1^\alpha \le z \le x_p^\alpha\} = \{z \in supp(A) \text{ such that } x_1^\alpha \le z \le x_p^\alpha\} =$$
$$= A^\alpha.$$

(The previous equality is only possible if we use the same hypotheses and results of the proposition 3.9)

5. We want to show the first distributive law, ie,

$$MIN_w(A, MAX_w(B,C)) = MAX_w(MIN_w(A,B), MIN_w(A,C))$$

By definition 3.4

$$MIN_w(A, MAX_w(B,C))^\alpha =$$
$$= \{z \in supp(A) \bigwedge supp(MAX_w(B,C)) \text{ such that}$$
$$\min A^\alpha \wedge \min MAX_w(B,C)^\alpha \le z \le \max A^\alpha \wedge \max MAX_w(A,B)^\alpha\} =$$
$$= \{z \in supp(A) \bigwedge supp(MAX_w(B,C)) \text{ such that}$$
$$x_1 \wedge (y_1^\alpha \vee w_1^\alpha) \le z \le x_p^\alpha \wedge (y_k^\alpha \vee w_l^\alpha)\} =$$

(using the distributive law of real numbers)

$$= \{z \in supp(A) \bigwedge supp(MAX_w(B,C)) \text{ such that}$$
$$(x_1^\alpha \wedge y_1^\alpha) \vee ((x_1^\alpha \wedge w_1^\alpha) \le z \le (x_p^\alpha \wedge y_k^\alpha) \vee (x_p^\alpha \wedge w_l^\alpha)\} =$$
$$= \{z \in supp(A) \bigwedge (supp(B) \bigvee supp(C)) \text{ such that}$$
$$(x_1^\alpha \wedge y_1^\alpha) \vee ((x_1^\alpha \wedge w_1^\alpha) \le z \le (x_p^\alpha \wedge y_k^\alpha) \vee (x_p^\alpha \wedge w_l^\alpha)\} =$$
$$= \{z \in (supp(A) \bigwedge supp(B)) \bigvee (supp(A) \bigwedge supp(C))$$
$$\text{such that } (x_1^\alpha \wedge y_1^\alpha) \vee ((x_1^\alpha \wedge w_1^\alpha) \le z \le (x_p^\alpha \wedge y_k^\alpha) \vee (x_p^\alpha \wedge w_l^\alpha)\} =$$
$$= MAX_w(MIN_w(A,B), MIN_w(A,C))^\alpha$$

Analogously for the other distributive law. □

ISBN: 978-989-95079-6-8

Corollary 3.12 *The triplet ($\mathcal{A}_1$,MIN_w,MAX_w) is a distributive lattice, in which MIN_w and MAX_w represent the meet and join, respectively.*

Remark 3.13 *The lattice ($\mathcal{A}_1$,MIN_w,MAX_w) can also be expressed as the pair ($\mathcal{A}_1$,$\preceq$), where $\preceq$ is a partial ordering defined as:*

$A \preceq B$ if and only if $MIN_w(A, B) = A$
or, alternatively,
$A \preceq B$ if and only if $MAX_w(A, B) = B$
for any $A, B \in \mathcal{A}_1$.

Remark 3.14 *We can also define the partial ordering in terms of α-cuts:*

$A \preceq B$ if and only if $\min(A^\alpha, B^\alpha) = A^\alpha$
$A \preceq B$ if and only if $\max(A^\alpha, B^\alpha) = B^\alpha$
for any $A, B \in \mathcal{A}_1$ and $\alpha \in (0, 1]$, where A^α and B^α are finite sets (say , $A^\alpha = \{x_1^\alpha, \cdots, x_p^\alpha\}$,$B^\alpha = \{y_1^\alpha, \cdots, y_k^\alpha\}$). Then,

$$\min(A^\alpha, B^\alpha) = \{z \in supp(A) \bigwedge supp(B) \text{ such that}$$

$$\min(x_1^\alpha, y_1^\alpha) \leq z \leq \min(x_p^\alpha, y_k^\alpha)\} = A^\alpha \text{ and}$$

$$\max(A^\alpha, B^\alpha) = \{z \in supp(A) \bigvee supp(B) \text{ such that}$$

$$\max(x_1^\alpha, y_1^\alpha) \leq z \leq \max(x_p^\alpha, y_k^\alpha)\} = B^\alpha$$

Example 3.15
Let $u = \{0.4/1, 1/2, 0.8/3, 0.6/4, 0.5/5, 0.4/6, 0.3/7\}$ and $v = \{0.3/4, 0.6/5, 0.7/6, 0.8/7, 1/8, 0.8/9\}$ be two discrete fuzzy numbers. It is easy to prove that $u \preceq v$ because $MAX_w(u, v) = v$ or equivalently $MIN_w(u, v) = u$.

Corollary 3.16 *For each finite subset X of real numbers, let's consider the subset F_X of DFN such that any element of F_X has as support X. Then, the triplet (F_X,MIN_w,MAX_w) is a distributive lattice, in which MIN_w and MAX_w represent the meet and join, respectively.*

Example 3.17 *Let $u = \{0.3/3, 0.4/5, 1/7, 0.7/8\}$ and $v = \{0.3/3, 1/5, 0.7/7, 0.6/8\}$ be two discrete fuzzy numbers. It is easy to prove that $v \preceq u$ because $MAX_w(u, v) = u$ or equivalently $MIN_w(u, v) = v$.*

Remark 3.18 *It is possible to find two discrete fuzzy numbers $A, B \in \mathcal{A}_1$ such that $MAX_w(A, B) \neq A$ and $MAX_w(A, B) \neq B$ or equivalenty $MIN_w(A, B) \neq A$ and $MIN_w(A, B) \neq B$. This means that the two discrete fuzzy numbers, A and B, are not comparable. For example, let's consider $A = \{0.3/4, 0.5/5, 0.8/6, 1/7, 0.9/8, 0.7/9\}, B = \{0.5/6, 1/7, 0.9/8, 0.6/9, 0.5/10\} \in \mathcal{A}_1$. Then, $MIN_w(A, B) = \{0.3/4, 0.5/5, 0.8/6, 1/7, 0.9/8, 0.6/9\} \neq A, B$.*

Acknowledgment

We would like to express our thanks to anonymous reviewers who have contributed to improve this article.

References

[1] J.J. Buckley and Y. Qu. Solving systems of linear fuzzy equations. *Fuzzy Sets and Systems*, 43(2): 33–43, 1991.

[2] J.J. Buckley and E.Eslami. Fuzzy plane geometry I: Points and lines. *Fuzzy Sets and Systems*, 86(2): 179–188, 1997.

[3] J. Casasnovas and J. Vicente Riera. On the addition of discrete fuzzy numbers. *Wseas Transactions on Mathematics* 5(5): 549–554, 2006.

[4] J. Casasnovas and J. Vicente Riera. Discrete fuzzy numbers defined on a subset of natural numbers. *Theoretical Advances and Applications of Fuzzy Logic and Soft Computing:Advances in Soft Computing* , Springer, 42: 573–582, 2007.

[5] J. Casasnovas and J.Torrens. An Axiomatic Approach to the fuzzy cardinality of finite fuzzy sets. *Fuzzy Sets and Systems*, 133: 193–209, 2003.

[6] J. Casasnovas and J.Torrens. Scalar cardinalities of finite fuzzy sets for t-norms and t-conorms. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*,11 : 599–615, 2003.

[7] J. Casasnovas, J. Vicente Riera. "Maximum and Mininum of Discrete Fuzzy numbers". *Frontiers in Artificial Intelligence and Applications: artificial intelligence research and development*,IOS Press,163: 273–280, 2007.

[8] D. Dubois. A new definition of the fuzzy cardinality of finite sets preserving the classical additivity property. *Bull.Stud.Fuzziness Appl.(BUSEFAL)*, 5: 11–12, 1981.

[9] D.Dubois and H.Prade. *Fuzzy Sets and Systems:Theory and Applications*. Academic Press, New York, 1980.

[10] D.Dubois and H.Prade (Eds.).*Fundamentals of Fuzzy Sets*. The handbooks of Fuzzy Sets Series, Kluwer, Boston,2000.

[11] M.Hanns. On the implementation of fuzzy arithmetical operations for engineering problems.*Fuzzy Information Processing Society(NAFIPS)*:462–466, 1999.

[12] Wen-Ling Hung and Miin-Shen Yang, Fuzzy clustering on LR-type fuzzy numbers with an application in Taiwanese tea evaluation. *Fuzzy Sets and Systems* ,150: 561–577, 2005.

[13] George J. Klir and Yuan Bo. *Fuzzy sets and fuzzy logic (Theory and applications)*. Prentice Hall, 1995.

[14] G. Mayor et al. Multi-dimensional Aggregation of Fuzzy Numbers Through the Extension Principle. *Data Mining, Rought sets and Granular Computing*, Eds Lin, Yao, Zadeh, Physica-Verlag: 350-363, 2002.

[15] S. Miyamoto. Fuzzy Multisets and Their Generalizations. *Multisets Processing(LNCS)*, 2235 : 225–235, 2001.

[16] Daniel Rocacher, Patrick Bosc. *The set of fuzzy rational numbers and flexible querying*. Fuzzy Sets and Systems, 155(3): 317-339, 2005.

[17] D. Tadic. Fuzzy Multi-criteria approach to ordering policy ranking in a supply chain. *Yugoslav Journal of Operations Research*: 243–258, 2005.

[18] W. Voxman. Canonical representations of discrete fuzzy numbers. *Fuzzy Sets and Systems*, 54: 457–466, 2001.

[19] Guixiang Wang,Cong Wu, Chunhui Zhao. Representation and Operations of discrete fuzzy numbers. *Southeast Asian Bulletin of Mathematics* 28: 1003–1010, 2005.

[20] Kun-lun Zhang, Kaoru Hirota. On fuzzy number lattice $(\widetilde{R}, \leq)$. *Fuzzy Sets and Systems*, 92: 113–122, 1997.

ISBN: 978-989-95079-6-8

On Some Classes of Aggregation Functions that are Migrative

János Fodor Imre J. Rudas

Institute of Intelligent Systems, Budapest Tech
Budapest, Hungary
Email: {fodor,rudas}@bmf.hu

***Abstract*— *In this paper we introduce and describe two classes of aggregation functions with members that are migrative. First continuous triangular norms are studied that own the migrative property with respect to another fixed t-norm, in particular to the three prototypes minimum, product, and the Łukasiewicz t-norm. For classes of nilpotent and strict migrative t-norms the characterization and construction is carried out by solving functional equations for the generators. For the third case when the fixed t-norm is the minimum, an ordinal-sum-like construction is resulted. Then we formulate and study the migrative property for quasi-arithmetic means. The results are less permissive than for t-norms: we can hardly escape from the classical arithmetic mean.*

***Keywords*—** Triangular norms, quasi-arithmetic means, functional equations, migrative property.

1 Introduction

In this paper we study the migrative property for some classes of binary *aggregation functions*. These are increasing functions $F\colon [a,b]\times[a,b]\to[a,b]$ such that $F(a,a)=a$ and $F(b,b)=b$.

It is well-known that some operations in fuzzy logic (like t-norms, t-conorms, uninorms, nullnorms) can be considered aggregation functions on the closed unit interval in the above sense. In addition, these operations own a characteristic property of associativity that can be formulated as a well-known functional equation. A function $F\colon [a,b]\times[a,b]\to[a,b]$ is called *associative* if

$$F(x,F(y,z))=F(F(x,y),z),\quad x,y,z\in[a,b]. \tag{1}$$

Under some additional conditions, the above-mentioned logical operations are the general solutions of the associativity equation (1).

Another important class of aggregation functions consists of *means*. One of the first authors who introduced the term "mean" in a mathematical sense was Cauchy [1]. In that spirit, a function $F\colon [a,b]\times[a,b]\to[a,b]$ is called a *mean* in the interval $[a,b]\subset\mathbb{R}$ if

$$\min(x,y)\le F(x,y)\le\max(x,y)$$

for all $x,y\in[a,b]$. Well-known examples are the median, the arithmetic mean, the geometric mean, the root-power mean, the harmonic mean and the like.

Although these means and t-norms, t-conorms, uninorms, nullnorms have rather different properties, there is one particular common property they all share. This is called *bisymmetry*, introduced by Aczél [2], and is defined for a function $M\colon [a,b]\times[a,b]\to[a,b]$ by

$$F(F(x,y),F(u,v))=F(F(x,u),F(y,v)), \tag{2}$$

for all $x,y,u,v\in[a,b]$.

It is easy to see that if a function $F\colon [a,b]\times[a,b]\to[a,b]$ is symmetric (i.e., $F(x,y)=F(y,x)$ for all $x,y\in[a,b]$) and associative then F is bisymmetric. The converse statement is not true in general. If, however, F is symmetric and has a neutral element, then bisymmetry of F implies that F is associative [3].

We give the definition of α-migrative two-variable functions in Definition 1 just as it was proposed originally in [5]. Since that notion has something to do, in essence, with triangular norms, first we attempt to catch hold of the very meaning of it for t-norms. This is expressed by equation (6), and two more equivalent forms in Theorem 2. That is, for t-norms we can choose any of the three forms of (8), (9) or (10). Equation (8) is obtained by considering the classical associativity functional equation, then fixing the t-norm (T_0) inside the equation, and finally fixing one of the variables' value at α. Equation (9) can be obtained similarly: the only difference is that we fixed the t-norm outside the equation. Equation (10) is just equivalent to (8) and (9) for t-norms.

Therefore, when we realized that Definition 1 can never hold for t-conorms, and it is very restrictive for quasi-arithmetic means, we tried to develop new definitions that are context dependent. The idea is to use the characteristic composite functional equation of the function class under study (e.g., the bisymmetry equation for means), then fix the function inside (or outside), and finally fix one of the variables at some α. Or alternatively, consider the equation that corresponds to (10). While for t-norms these three approaches are equivalent, for quasi-arithmetic means this is not the case at all.

According to our sketched ideas and investigations, in this paper we introduce and describe two classes of aggregation functions (triangular norms and quasi-arithmetic means) with members that are migrative. The underlying compact interval is $[0,1]$ in the study. In our formulation the migrative property is context-dependent. First we show the original property and its extensions for triangular norms. Then migrativity is reformulated and studied for quasi-arithmetic means. In the conclusions we highlight the essential differences between the two classes.

2 The migrative property for t-norms

In this section we highlight some essential aspects and consequences of the original α-migrative property and its possible extensions.

A *triangular norm* (t-norm for short) $T:[0,1]^2\to[0,1]$ is an associative, commutative, non-decreasing function such that $T(1,x)=x$ for all $x\in[0,1]$. Prototypes of t-norms are

ISBN: 978-989-95079-6-8

the minimum $T_{\mathbf{M}}(x,y)=\min(x,y)$, the product $T_{\mathbf{P}}(x,y)=xy$, and the Łukasiewicz t-norm $T_{\mathbf{L}}(x,y)=\max(x+y-1,0)$. Obviously, the product t-norm $T_{\mathbf{P}}$ is α-migrative for any $\alpha\in]0,1[$.

As it is well-known, each continuous Archimedean t-norm T can be represented by means of a continuous additive generator (see e.g. [4]), i.e., a strictly decreasing continuous function $t\colon[0,1]\to[0,\infty]$ with $t(1)=0$ such that

$$T(x,y)=t^{(-1)}(t(x)+t(y)), \tag{3}$$

where $t^{(-1)}\colon[0,\infty]\to[0,1]$ is the pseudo-inverse of t, and is given by

$$t^{(-1)}(u)=t^{-1}(\min(u,t(0))).$$

2.1 The Class of α-Migrative Functions

In [5] the authors introduced a new term – α-migrative – for a class of two-variable functions (binary operations) as follows.

Definition 1. Let α be in $]0,1[$. A binary operation $T\colon[0,1]^2\to[0,1]$ is said to be *α-migrative* if we have

$$T(\alpha x,y)=T(x,\alpha y)\quad\text{for all } x,y\in[0,1]. \tag{4}$$

Although this definition seems to be rather general, the paper [5] deals with triangular norms and subnorms. Further progress can be seen in [6], where the authors studied a slightly modified definition of migrativity. They discussed also the associativity and bisymmetry of migrative aggregation functions.

Notice also that Definition 1 does not provide a meaningful notion for triangular conorms. Indeed, if S is a t-conorm then it is α-migrative if and only if $S(\alpha x,y)=S(x,\alpha y)$ holds for all $x,y\in[0,1]$. If we choose $y=0$ then we must have $\alpha x=x$ for all $x\in[0,1]$, because S is α-migrative. This is impossible when $\alpha\neq 1$. Similarly, if $y=1$ then we must have $S(x,\alpha)=1$ for all $x\in[0,1]$, which is again impossible unless $\alpha=1$. Therefore, even the correct definition of α-migrative t-conorms needs special care.

Obviously, the product t-norm $T_{\mathbf{P}}$ is α-migrative for any $\alpha\in]0,1[$. The following non-continuous t-norm T_β is also α-migrative:

$$T_\beta(x,y)=\begin{cases}\min(x,y) & \text{if } \max(x,y)=1,\\ \beta xy & \text{otherwise.}\end{cases} \tag{5}$$

It was provided in [8] as a counterexample to the question: Is a strictly monotone t-norm always continuous?

We provide all *continuous* solutions of the equation (4) in Theorem 1. More details and proofs can be found in our paper [7].

Theorem 1. *Let α be in $]0,1[$. Suppose T is a continuous t-norm. Then the following statements hold.*

(a) T is α-migrative if and only if we have

$$T(\alpha,y)=\alpha y\quad \textit{for all } y\in[0,1]. \tag{6}$$

(b) If T is α-migrative then T is strict.

(c) Suppose t is an additive generator of T. Then T is α-migrative if and only if there exists a continuous, strictly decreasing function t_0 from $[\alpha,1]$ to the non-negative reals with $t_0(0)<+\infty$ and $t_0(1)=0$ such that

$$t(x)=k\cdot t_0(\alpha)+t_0\left(\frac{x}{\alpha^k}\right)\quad \textit{if } x\in]\alpha^{k+1},\alpha^k], \tag{7}$$

where k is any non-negative integer.

In the next subsection we extend the above definition by allowing any fixed t-norm instead of the product in the defining equation.

2.2 The Class of (α,T_0)-Migrative Functions

Following our plan, the next definition extends the migrative property as follows. Some more details can be found in [9] and in [10] with proofs.

Definition 2. Let α be in $]0,1[$ and T_0 a fixed triangular norm. A binary operation $T\colon[0,1]^2\to[0,1]$ is said to be *α-migrative with respect to T_0* (shortly: (α,T_0)-migrative) if we have

$$T(T_0(\alpha,x),y)=T(x,T_0(\alpha,y)) \tag{8}$$

for all $x,y\in[0,1]$.

Obviously, Definition 1 is a particular case of Definition 2 with $T_0=T_{\mathbf{P}}$.

In the next theorem we give a simple but very useful characterization of (α,T_0)-migrative t-norms.

Theorem 2. *Let α be in $]0,1[$ and T_0 a fixed triangular norm. Then the following statements are equivalent for a t-norm $T\colon[0,1]^2\to[0,1]$.*

(i) T is α-migrative with respect to T_0;

(ii) T satisfies the following equation for all $x,y\in[0,1]$:

$$T_0(T(\alpha,x),y)=T_0(x,T(\alpha,y)); \tag{9}$$

(iii) T satisfies the following equation for all $x\in[0,1]$:

$$T(\alpha,x)=T_0(\alpha,x). \tag{10}$$

We do not recall the case when $T_0=T_{\mathbf{L}}$. The interested reader can find the related results in [9, 10].

For illustrating the non-representable case $T_0=T_{\mathbf{M}}$, we consider $(\alpha,T_{\mathbf{M}})$-migrative continuous triangular norms. In the present case the $(\alpha,T_{\mathbf{M}})$-migrative property is read as follows:

$$T(\min(\alpha,x),y)=T(x,\min(\alpha,y)),\ x,y\in[0,1]. \tag{11}$$

By Theorem 2 we know that the $(\alpha,T_{\mathbf{M}})$-migrative property is equivalent to the following functional equation:

$$T(\alpha,y)=\min(\alpha,y),\quad y\in[0,1]. \tag{12}$$

The description of all $(\alpha,T_{\mathbf{M}})$-migrative continuous triangular norms, as solutions to the equation (12), is given in the following theorem. For the proof see [10].

ISBN: 978-989-95079-6-8

Theorem 3. *A continuous t-norm T is $(\alpha, T_{\mathbf{M}})$-migrative if and only if there exist two continuous t-norms T_1 and T_2 such that T can be written in the following form:*

$$T(x,y) = \begin{cases} \alpha T_1\left(\frac{x}{\alpha}, \frac{y}{\alpha}\right) & x,y \in [0,\alpha], \\ \alpha + (1-\alpha)T_2\left(\frac{x-\alpha}{1-\alpha}, \frac{y-\alpha}{1-\alpha}\right) & x,y \in [\alpha,1], \\ \min(x,y) & \textit{otherwise}. \end{cases}$$

Notice that equations (8) and (10) has something to do with the *generalized associativity equation* that has also been studied and solved under some additional conditions, see [11, 12]. It can be written as follows:

$$F(G(x,y),z) = H(x,K(y,z)). \tag{13}$$

In this general framework the particular form of $H = F$ and $K = G$ in (13) correspond to (8).

3 The class of α-migrative quasi-arithmetic means

Let $[a,b]$ be a compact real interval. Aczél [2] proved that a function $M : [a,b]^2 \to [a,b]$ is continuous, symmetric, strictly increasing in each argument, idempotent and fulfils the bisymmetry equation (2) if and only if

$$M(x,y) = f^{-1}\left(\frac{f(x)+f(y)}{2}\right), \quad x,y \in [a,b] \tag{14}$$

with some continuous strictly monotonic function f. We also know that this result still holds for intervals of the form $]a,b]$, $[a,b[$, $]a,b]$ or even for any unbounded interval of the real line (see [13], pp 250–251, 280). A function M having a representation (14) is called a *quasi-arithmetic mean*.

When one tries to extend migrativity for quasi-arithmetic means, there are several options. One possibility is to choose a particular quasi-arithmetic mean M_0 (for instance, the arithmetic mean $M_{\mathbf{A}}(x,y) = \frac{x+y}{2}$), and consider the bisymmetry equation (2) extended in the following forms, with a fixed argument (say $y = \alpha$):

$$M(M_0(x,\alpha), M_0(u,v)) = M(M_0(x,u), M_0(\alpha,v)) \tag{15}$$
$$M_0(M(x,\alpha), M(u,v)) = M_0(M(x,u), M(\alpha,v)) \tag{16}$$

For motivations of these equations we refer to Theorem 2 and formulas (8)–(10) in case of t-norms.

We indeed chose the arithmetic mean, because of its key role in the representation of quasi-arithmetic means (14). In this case equations (15) and (16) are of the following respective forms:

$$M\left(\frac{x+\alpha}{2}, \frac{u+v}{2}\right) = M\left(\frac{x+u}{2}, \frac{\alpha+v}{2}\right), \tag{17}$$

$$\frac{M(x,\alpha)+M(u,v)}{2} = \frac{M(x,u)+M(\alpha,v)}{2}, \tag{18}$$

for a fixed $\alpha \in]0,1[$ and for all $x,u,v \in [0,1]$.

First we state that equation (18) is extremely restrictive: under the condition of idempotency, M must be the arithmetic mean.

Theorem 4. *Assume that an idempotent function $M\colon [0,1]^2 \to [0,1]$ satisfies equation (18). Then M is equal to the arithmetic mean.*

Therefore, if we want to define α-migrativity in a way that allows not only the arithmetic mean owning this property, we should try the other equation (17).

Lemma 1. *Let α be in $]0,1[$. Then*

$$M(\alpha,x) = \frac{\alpha+x}{2} \textit{ for all } x \in [0,1] \tag{19}$$

is a necessary condition for having (17) satisfied by an idempotent M.

This simple observation suggests that an appropriate definition of α-migrative quasi-arithmetic means might be the following one.

Definition 3. Let α be in $]0,1[$. A quasi-arithmetic mean $M\colon [0,1]^2 \to [0,1]$ is said to be *α-migrative* if (19) holds for all $x \in [0,1]$.

We solve equation (19) first, than we relate the solution to (17).

3.1 Quasi-Arithmetic Means that Satisfy (19)

Consider a quasi-arithmetic mean M in the form of (14) with a strictly increasing, continuous generator function f defined on $[0,1]$. Suppose M satisfies (19). That is, we have

$$f^{-1}\left(\frac{f(\alpha)+f(x)}{2}\right) = \frac{\alpha+x}{2} \quad \text{for all } x \in [0,1]. \tag{20}$$

Equation (20) implies that $f(0) = \lim_{x \downarrow 0} f(x)$ and $f(1) = \lim_{x \uparrow 1} f(x)$ are finite. Indeed, for instance, if $f(0) = -\infty$ then the left-hand side of (20) equals 0, while the right-hand side is $\frac{\alpha}{2}$, a contradiction. The other cases lead to a contradiction in a similar manner.

Therefore, (20) can be written in an equivalent form as

$$\frac{f(\alpha)+f(x)}{2} = f\left(\frac{\alpha+x}{2}\right) \quad \text{for all } x \in [0,1]. \tag{21}$$

The general solution of (21) is given in the next theorem.

Theorem 5. *Assume that $f : [0,1] \to \mathbb{R}$ is a strictly increasing, continuous function. Then f satisfies (21) if and only if there exist real numbers f_0, f_α, f_1 such that f can be written in the following form:*

$$f(x) = \begin{cases} \frac{f_\alpha - f_0}{\alpha} \cdot x + f_0 & \textit{if } x \in [0,\alpha] \\ \frac{f_1 - f_\alpha}{1-\alpha} \cdot x + \frac{f_\alpha - \alpha f_1}{1-\alpha} & \textit{if } x \in]\alpha,1] \end{cases}. \tag{22}$$

That is, the most general quasi-arithmetic means that satisfy (19) are generated by a piece-wise linear generator function. It is linear on $[0,\alpha]$ and on $[\alpha,1]$.

Next we show that the piecewise linear solution of (21) becomes linear when we feed it back into (17).

Theorem 6. *Let M be a quasi-arithmetic mean that satisfies (17). Then $M(x,y) = \frac{x+y}{2}$, i.e., M equals the arithmetic mean.*

Notice that this result is somewhat disappointing: we returned back again to quasi-arithmetic means.

ISBN: 978-989-95079-6-8

4 Conclusions

It is worth comparing results presented in this paper for t-norms and for quasi-arithmetic means. For t-norms a general form of (α, T_0)-migrativity resulted in two other equivalent formulations, and also infinite families of t-norms owning that property could be obtained. For quasi-arithmetic means idempotency alone implied the arithmetic mean as the only solution in (18). When we looked for quasi-arithmetic mean solutions of (17), it was also the arithmetic mean that satisfied the equation. The only exceptions are quasi-arithmetic mean solutions to (19) with piece-wise linear generator functions.

Notice that the general case when the fixed quasi-arithmetic mean $M_0(x,y) = f_0^{-1}\left(\frac{f_0(x)+f_0(y)}{2}\right)$ differs from the arithmetic mean can be handled in a similar way.

It might be worth studying other types of composite functional equations in the spirit of the present paper.

Acknowledgment

The authors thank the referees for their valuable suggestions.

The authors gratefully acknowledge the support obtained within the frames of Bilateral Romanian-Hungarian S&T Cooperation Project conducted by the Research and Technology Fund in Hungary. János Fodor has also been supported in part by OTKA K063405, and by COST Actions IC0602 and IC0702.

References

[1] Cauchy, A.L., *Cours d'analyse de l'Ecole Royale Polytechnique, 1ère partie; Analyse algébrique* (Paris) (1821).

[2] J. Aczél, On mean values, *Bulletin of the American Math. Society*, **54** (1948) 392–400.

[3] W. Sander, Some aspects of functional equations, in: E.P. Klement and R. Mesiar, Eds., Logical, algebraic, analytic and probabilistic aspects of triangular norms (Elsevier, Amsterdam-Boston-Heidelberg-London-New York-Oxford-Paris-San Diego-San Francisco-Sydney-Tokyo, 2005), pp. 143–187.

[4] E.P. Klement, R. Mesiar, E. Pap, *Triangular Norms*, Trends in Logic. Studia Logica Library, Vol. 8, Kluwer Academic Publishers, Dordrecht, 2000.

[5] F. Durante and P. Sarkoci, A note on the convex combinations of triangular norms, *Fuzzy Sets and Systems* 159 (2008) 77–80.

[6] H. Bustince, J. Montero and R. Mesiar, Migrativity of aggregation functions, Fuzzy Sets and Systems (2008), doi: 10.1016/j.fss.2008.09.018.

[7] J. Fodor and I.J. Rudas, On continuous triangular norms that are migrative, *Fuzzy Sets and Systems* 158 (2007) 1692 – 1697.

[8] M. Budinčević and M.S. Kurilić, A family of strict and discontinuous triangular norms, *Fuzzy Sets and Systems* 95 (1998) 381–384.

[9] J. Fodor and I.J. Rudas, Further results on migrative triangular norms, In: *Proc. 6th IEEE International Conference on Computational Cybernetics*, (November 27-29, 2008, Stará Lesná, Slovakia), pp. 123-126.

[10] J. Fodor and I.J. Rudas, On generalizations of the migrativity property for t-norms (2009, forthcoming).

[11] J. Aczél, V. D. Belousov and M. Hosszú, Generalized associativity and bisymmetry on quasigroups, Acta Math. Sci. Hung. 11 (1960) 127–136.

[12] Gy. Maksa, Quasisums and generalized associativity, Aequationes Math. 69 (2005) 6–27.

[13] J. Aczél, *Lectures on Functional Equations and Applications*, (Academic Press, New York, 1966).

ISBN: 978-989-95079-6-8

Intelligent Operation Support Method based on Time Change Fuzzy Sets

Seiji YASUNOBU[1] Mitsuru KOKUHUDA[1]

1.Department of Intelligent Interaction Technologies
University of Tsukuba, Tsukuba, Japan
Email: yasunobu@iit.tsukuba.ac.jp, kokuhuda_ ics@edu.esys.tsukuba.ac.jp

Abstract— *In the present paper, an effective support method for equipment operation under a dynamic environment is proposed. The proposed method presents operation instructions that the driver should be preparing to follow in the future. Using this method, the current and future states of the surrounding environment can be considered. The time change fuzzy set, which considers the state and time, to be universe of discourse. The time change fuzzy set is used for a computer as a buildup method with expert knowledge concerning the time change in the state. Experiments to support a lane change operation were conducted using a driving simulator. The obtained results indicated that the proposed method safely supports a lane change operation, and the time lag between the instruction being given by the system and operation being performed by the driver is decreased.*

Keywords— Dynamic environment, fuzzy inference, operation support, safety, time change fuzzy sets.

1 Introduction

When a human operates equipment in a dynamic environment, in which the surrounding conditions change with time, the operator should correctly understand complex changes in the surrounding environment. In addition, the operator must judge whether an operation is appropriate within a limited time. Therefore, an equipment operation support system is necessary in order to reduce the load on the operator and to achieve safe operation.

The equipment operation support system causes a time lag between the instruction by system and the operation by the driver. This is a significant problem in a dynamic environment, into which surrounding circumstances change with time.

Fuzzy sets can quantitatively manage human subjective assessments, such as ”He is *young*,” and thought processes, such as ”IF x is *small*, THEN y is *very large*” [1, 2]. Mamdani applied fuzzy control, which is an intelligent control, experimentally to a steam engine for the first time [3]. Yasunobu achieved the selection of the control order based on a multipurpose evaluation of an expert driver using the predictive fuzzy control and applied to the automatic train operation (ATO) system [4, 5, 6]. These fuzzy sets and control methods are applied to numerous intelligent cooperative control systems based on ”fuzzy instruction” [7, 9] and a ”fuzzy target”[8]. However, in fuzzy sets, it is difficult to build in expert knowledge concerning the time change. High-performance computers that have recently been developed have enabled equipment operation support, including expert knowledge, in real time.

In the present study, we construct an effective support method for use in a dynamic environment. When operating equipment in a dynamic environment, experts perform appropriate operations by considering not only the present state but also potential future states. In the present paper, time change fuzzy sets, which considers the state and time, to be universe of discourse. Time change fuzzy sets are used to provide expert knowledge to the computer regarding the time change of the states. Time change fuzzy sets can provide a sense of time change, such as ”Speed is continuous high” and ”Temperature change to cold from hot” to a computer.

The proposed method is applied to support a lane change operation using a driving simulator.

2 Support method in a dynamic environment

In this section, an effective support method for a dynamic environment is considered. The process of the proposed system is shown in Fig. 1.

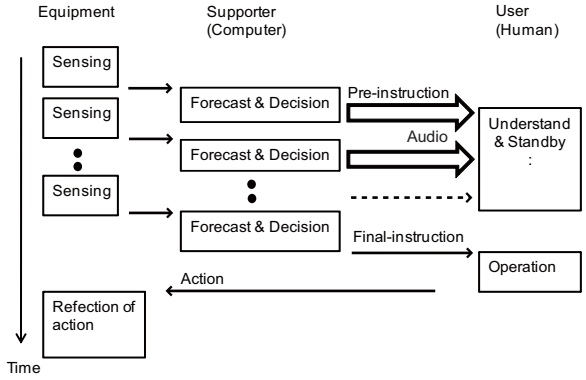


Figure 1: Process of the proposed operation support method.

The support (system) consists of 1) sensing equipment and surrounding states, 2) forecasting of future states, 3) instruction decision, and 4) presentation of instruction in advance and sending of a pre-instruction via audio. The driver (human) is 5) perceives the pre-instruction, and 6) interprets the intention of the system and waits for the final/pre-instruction. The support (system) continues 7) sensing the surrounding state and sends a final short signal. The driver (human) 8) starts an operation.

In the proposed method, the driver can confirm the operation instructions in advance. In addition, the time of 4) presentation of instruction in advance, 5) perception of instruction, 6) interpreting the intention of the system, can be sufficiently secured. The time lag can be related only to 8) operation.

The proposed method 2) forecasts future states and 3) decides pre-instruction or final-instruction, and the manner of action 2) and 3) is important.

ISBN: 978-989-95079-6-8

3 Fuzzy inference that considers future time

A computer can be made to emulate man's interpretive ability by building expert knowledge into a system using the fuzzy theory and fuzzy inference.

In a dynamic environment, an expert operates equipment by considering not only present states but also forecasting future states. It is necessary to define fuzzy sets concerning the change of states in order to build expert knowledge about the change of states and to perform inference by considering future states.

3.1 Time change fuzzy sets

In the present paper, we define a fuzzy set that changes with time. A fuzzy set to which a time axis has been added is referred to herein as a time change fuzzy set.

The total set of the state is assumed as R. The time change fuzzy set $\widetilde{X}fn$ in state $x(t)$ of the object is defined by the following expression (1):

$$\widetilde{X}_{fn(x,p)} = \int_{R\times P} \mu_{\widetilde{X}fn(x,p)}/(x,p), \quad x \in R, p \in (0,P). \tag{1}$$

Here, $\mu_{\widetilde{X}_{fn(x,p)}}$ is the membership value in universe of discourse the state x and the following time p, and P is the maximum following time.

An example of a time change fuzzy set is shown in Fig. 2. This fuzzy set describes a "change in distance from far to near". The axes of this figure indicate distance, x, the following time, p, and the membership degree of the fuzzy set μ.

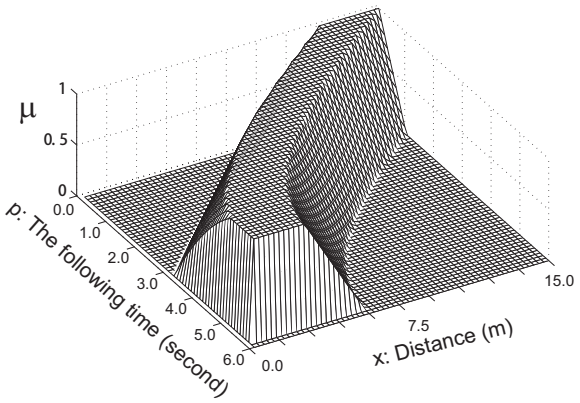


Figure 2: Example of a time change fuzzy set: change in distance from far to near.

Figure 3 shows the time change fuzzy set at $p = 0$ sec, $p = 3$ sec, and $p = 6$ sec.

Time change fuzzy sets indicate how a fuzzy set changes with time. Time change fuzzy sets can be used to build expert knowledge concerning the time change of states into a computer.

3.2 Fuzzy inference rules

Time change fuzzy sets are used to if-part and to then-part of fuzzy inference. To simplify explanation, three fuzzy rules for lane change support are described below.

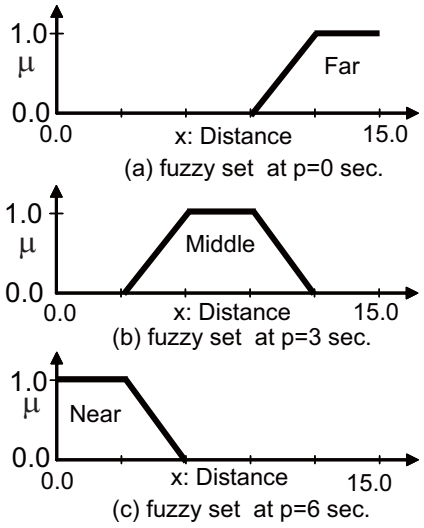


Figure 3: Time-sliced fuzzy sets of the time change fuzzy set: change in distance from far to near.

Rule 1: If forward vehicle (x) is *continuous far* and following vehicle (y) is *continuous far*, then lane change (z) is *continuous safe.*

Rule 2: If forward vehicle (x) is *continuous far* and following vehicle (y) is *change from far to near*, then lane change (z) is *change from safe to danger.*

Rule 3: If forward vehicle (x) is *change from near to far* and following vehicle (y) is *change from far to near*, then lane change (z) is *continuous middle.*

The terms x, y, and z in the above rules are defined by time change fuzzy sets, as shown in (1) . The above rules are then described as follows:

r1: $\widetilde{X}_{1(x,p)}$ *and* $\widetilde{Y}_{1(y,p)} \Rightarrow \widetilde{Z}_{1(z,p)}$,

r2: $\widetilde{X}_{2(x,p)}$ *and* $\widetilde{Y}_{2(y,p)} \Rightarrow \widetilde{Z}_{2(z,p)}$,

r3: $\widetilde{X}_{3(x,p)}$ *and* $\widetilde{Y}_{3(y,p)} \Rightarrow \widetilde{Z}_{3(z,p)}$.

3.3 Fuzzy inference process

Figure 4 shows an example of the fuzzy inference process. The if-part is given by time change fuzzy sets $\widetilde{X}_{(x,p)}$ and $\widetilde{Y}_{(x,p)}$, and the then-part is given by time change fuzzy set $\widetilde{Z}_{(x,p)}$. These fuzzy sets are in the form of time functions, as shown in Fig. 4. The input functions $\widehat{x}_{(p)}$ and $\widehat{y}_{(p)}$ are calculated by forecasting the future states from the present states.

The degree of membership of the input value to the time change fuzzy sets $\widetilde{X}'_{i(p)}$ and $\widetilde{Y}'_{i(p)}$ as shown in (2) and (3), respectively, is calculated by minimum operation, whereby time change fuzzy sets are removed by the input value. For example, when the time change fuzzy set $\widetilde{X}_{i(x,p)}$ is removed by the input value of $\widehat{x}_{(p)}$, the section area becomes $\widetilde{X}'_{i(p)}$.

$$\widetilde{X}'_{i(p)} = \widetilde{X}_{i(x,p)} \wedge \widehat{x}_{(p)} = \widetilde{X}_{i(\widehat{x}_{(p)},p)} \quad i = 1,3, \tag{2}$$

$$\widetilde{Y}'_{i(p)} = \widetilde{Y}_{i(y,p)} \wedge \widehat{y}_{(p)} = \widetilde{Y}_{i(\widehat{y}_{(p)},p)} \quad i = 1,3. \tag{3}$$

Each rule ri is evaluated to $\widetilde{Z}'_{i(z,p)}$, as shown in (4).

ISBN: 978-989-95079-6-8

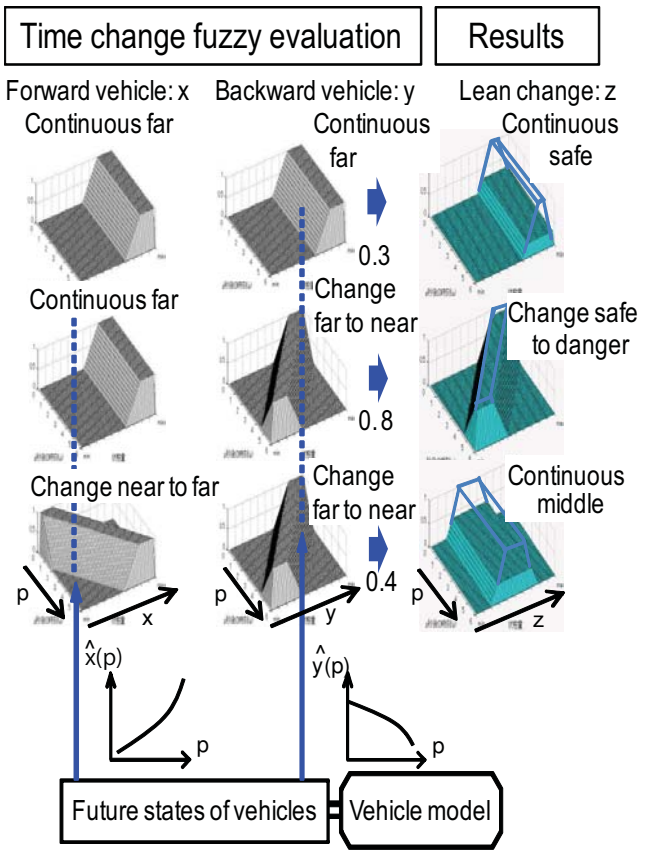


Figure 4: Fuzzy inference using time change fuzzy sets.

$$\widetilde{Z}'_{i(z,p)} = \widetilde{Z}_{i(z,p)} \wedge \widetilde{X}'_{i(p)} \wedge \widetilde{Y}'_{i(p)} \quad i = 1, 3. \tag{4}$$

3.4 *Result of fuzzy inference*

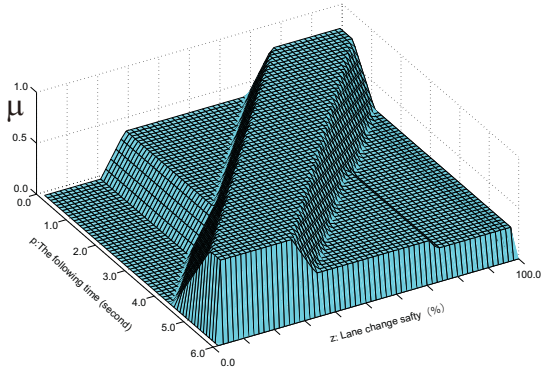
When the evaluation values of rules $\widetilde{Z}'_1$, $\widetilde{Z}'_2$, and $\widetilde{Z}'_3$ are 0.3, 0.8, and 0.4, respectively, as shown in Fig. 4, the result of fuzzy inference, i.e., $\widetilde{Z}_{ans(z,p)}$, as given by (5), is as shown in Fig. 5(a).

$$\widetilde{Z}_{ans(z,p)} = \widetilde{Z}'_{1(z,p)} \cup \widetilde{Z}'_{2(z,p)} \cup \widetilde{Z}'_{3(z,p)}. \tag{5}$$

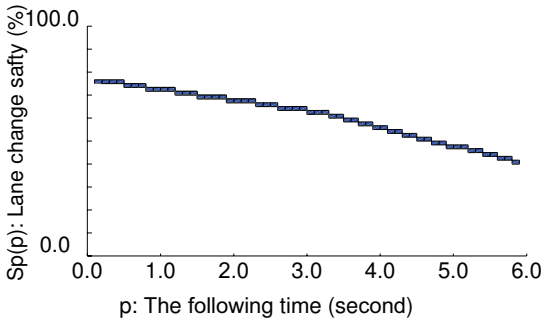
In the present study, $Sp_{(p)}$ is the lane change safety at time p obtained using the center of gravity method (defuzzification), as given by (6). Figure 5(b) shows the lane change safety at time p.

$$Sp_{(p)} = \frac{\int z \cdot \mu_{\widetilde{Z}ans(z,p)} dz}{\int \mu_{\widetilde{Z}ans(z,p)} dz} \tag{6}$$

Thus, the future output states are inferred from the future states of input, which are forecast from the present states. The system determines the operation that the driver should perform in the future based on the results of inference. The operation instruction is presented to the driver in advance, and the support method described in Section 2 is achieved. The shape of this fuzzy set looks like a type-2 fuzzy set[10]. The type-2 fuzzy set extends the membership grade to fuzzy value. Other side, this time change fuzzy set based on universe of discourse time and dynamic state, and simple membership grade from 0.0 to 1.0.



(a) Fuzzy set of lane change safety: $\widetilde{Z}_{ans}$



(b) Lane change safety at p: $Sp_{(p)}$

Figure 5: Result of fuzzy inference.

4 Application supporting a vehicle lane change

4.1 *Outline of the support system*

The support method proposed in Section 2 and fuzzy inference that considers the time change proposed in Section 3 are applied to support the lane change of a vehicle.

An expert driver can forecast the future states of other vehicles and judge intuitively whether a lane change can be performed safely.

In the present study, the knowledge of the expert driver is referred to as driving knowledge. A support system is constructed that forecasts future traffic states from present traffic states, judges whether a lane change can be performed safely and presents instructions to the driver.

It is preferable to maintain the velocity of the vehicle constant during a lane change so that the driver can operate the vehicle safely. Therefore, constant velocity is a requirement for safe lane changing. In addition, it is necessary to keep safety from the presentation of system's instruction to the operation of driver. This requirement is judged based on whether a sufficient distance between the driver's vehicle and the other vehicles will be maintained during a lane change.

The lane change support process is as follows. (a) Assume that the velocity of one's own vehicle is constant, (b) Forecast changes in the relative positions of the other vehicles, and (c) Present the instruction to change lanes if it appears that safety will be maintained.

4.2 *Forecast future states*

Other vehicles are assumed to maintain their velocities. The relative positions of the other vehicles are calculated as follows based on assumption (a) above.

ISBN: 978-989-95079-6-8

Predictive states after p seconds are as follows:

$$\widehat{x}_{(p)} = \widetilde{x}_{(t)} + v_x p \quad (7)$$

$$\widehat{y}_{(p)} = \widetilde{y}_{(t)} + v_y p, \quad (8)$$

where $\widetilde{x}_{(t)}$ and $\widetilde{y}_{(t)}$ denote the relative position of forward and backward vehicle at the present time, t, and v_x and v_y denote the relative velocity of these vehicles respectively.

4.3 Judge lane change appropriateness

In order to judge whether a lane change is possible based on the prediction of future states, expert knowledge concerning the future state of traffic is necessary. In the present study, the appropriateness of the lane change is referred to as the lane change appropriateness and is expressed as a number from 0 to 10. These numbers correspond to the following meanings: 0; dangerous(unsafe), 5; will be safe soon, and 10; safe.

Fuzzy inference is performed by using the predictive values of the relative positions of other vehicles as the input calculated in Section 4.2, and the lane change appropriateness is given as the output. The instruction to change lanes is presented to the driver based on the lane change appropriateness.

4.4 Presentation of audio instructions

In the proposed system, the driver is presented with audio instructions. The instruction is decided according to the lane change appropriateness and the change. In the present study, the instruction is output as follows.

Audio instruction:

· Present: Safety is high. Future: Safety is high.
⟶ Instruction: ”You can safely change lanes.”

· Present: Safety is high. Future: Safety is low.
⟶ Instruction: ”It’s getting dangerous.”

· Present: Safety is low. Future: Safety is high.
⟶ Instruction: ”You will be able to change lanes soon.”
After a few seconds.
⟶ Instruction: ”Change lanes now.” = Go signal

· Present: Safety is low. Future: Safety is low.
⟶ Instruction: ”You can’t change lanes.”

5 Driving simulator experiments

Experiments using the lane change support system are performed using proposed support method. The case in which the vehicle in the left-hand lane on a two-lane roadway enters the right-hand lane is assumed. The only vehicle in the left-hand lane is the driver’s own vehicle, and two or more other vehicles are driving in the right-hand lane. Fig. 6 shows a photograph of the driving simulator. The test drivers are three men in their twenties.

The following two situations are assumed in the experiments.

Situation 1) Approach from the rear: A vehicle approaches from the rear (Fig. 7(a)).



Figure 6: Driving simulator.

Situation 2) Nearby vehicle moves away: A nearby vehicle moves away gradually (Fig. 8(a)).

First, an experiment is performed using the support method in which only the present state is considered. Then, the experiment is performed using the proposed method (Section 2) in which the future states are considered. In the proposed method, time change fuzzy sets (Section 3) are used. Audio instructions are presented as described in Section 4.4.

5.1 Situation 1: Approach from the rear

5.1.1 Results obtained by considering only the present state

The results of an experiment conducted with one test driver in Situation 1 without considering future states is shown in Fig. 7(b). In this method, in which only the present state is considered, the experimental system can only judge whether the present state is ”safe” or ”unsafe”, and the safety of future states cannot be predicted. Fig. 7(b) indicates that when the driver started to operate the steering wheel, the system reported that the state changed from safe to unsafe. In this example, the time lag between the instruction of the system and the operation by the driver caused a problem.

5.1.2 Result obtained by considering the future states

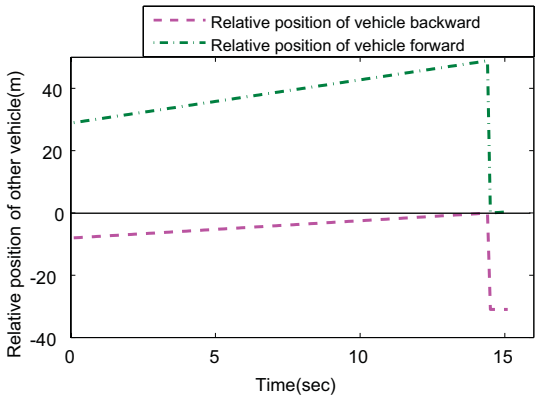
The results of an experiment conducted with one test driver in Situation 1 considering future states is shown in Fig. 7(c). As the results indicate, the experimental system did not indicate that the state became safe, and driver did not operate the steering wheel. In Situation 1, the experimental system can predict the safety of future states. In addition, the experimental system judges that indicating that the state will become safe is not appropriate considering the time of driver’s perception, understanding, and operation.

5.2 Situation 2: Nearby vehicle moves away

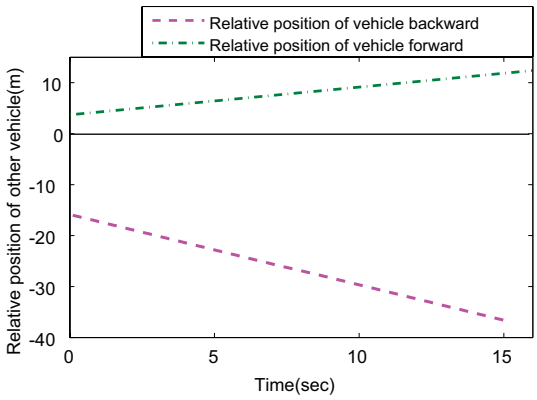
5.2.1 Results obtained by considering only the present state

The results of an experiment conducted with one test driver in Situation 2 are shown in Fig. 8(b). This figure shows the time lag between the time at which the “safe” message is provided by the system and the time at which the driver begins to operate the steering wheel. The average of time lag of the three testers is $1.70sec.$

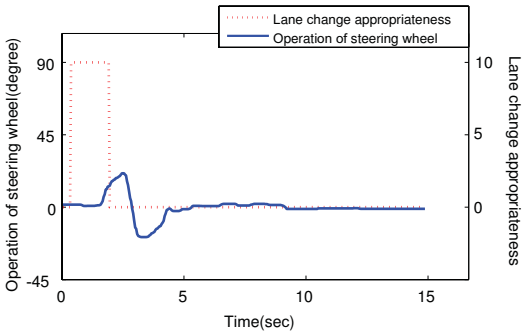
ISBN: 978-989-95079-6-8



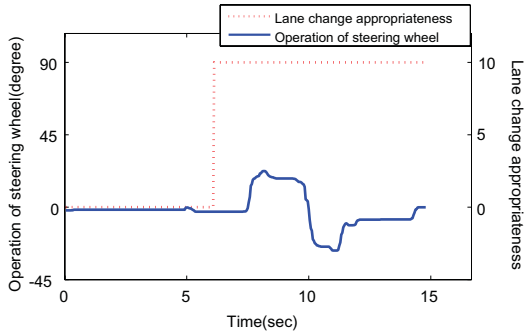
(a) Situation 1:Approach from the rear



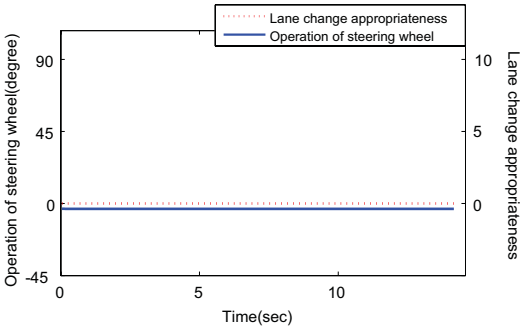
(a) Situation 2:Nearby vehicle moves away



(b) Results obtained by considering only the present state: Situation 1

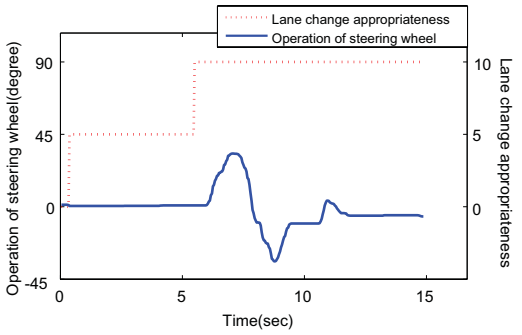


(b) Results obtained by considering only the present state: Situation 2



(c) Results obtained by considering future states: Situation 1

Figure 7: Experiments results obtained for Situation 1.



(c) Results obtained by considering future states: Situation 2

Figure 8: Experiments results obtained for Situation 2.

5.2.2 *Results obtained by considering future states*

The results of an experiment conducted with one test driver considering future states in Situation 2 are shown in Fig.8(c). The difference from Fig. 8(b) is that the message "It will be safe to change lanes soon" is presented to the driver, because this method considers not only the present state but also predicts future states. The "safe" message is then presented at the time when system judges that it will be safe to change lanes for a while. The final signal is then presented to the driver.

In this way, the driver is presented with the knowledge that he will be able change lanes soon, and can prepare to change lanes. In addition, the time required to provide the audio instruction can be shortened by using a signal to indicate that it is safe to change lanes.

As shown in Fig. 8(c), the time lag between the time at which the "safe" message is provided by the system and the time at which the driver begins to operate the steering wheel became shorter. The average time lag of three testers is $0.63sec$.

5.3 *Discussion*

The results of the experiments reveal the following.

- In Situation 1, the proposed support system can predict danger in future states, whereas the system that considers only the present state cannot predict danger in future states.
- In Situation 2, in the proposed support system, the time lag between the time at which the system provides an instruction and the time at which the driver performs an operation is shorter than that in the system that considers only the present state.

ISBN: 978-989-95079-6-8

Therefore, the effectiveness of the proposed system as a support system for use in a dynamic environment is demonstrated.

6 Conclusions

An effective support method for use in a dynamic environment is proposed in the present paper. The proposed method forecasts future states that are calculated from present states, and the driver is provided in advance with instructions to perform various operations. In order to achieve the proposed method, time change fuzzy sets are proposed as a method of building expert knowledge concerning the time change of the states into a computer.

The proposed method was applied to a lane change support system and experiments were performed. The obtained results indicate that it is possible to achieve a safety support system that can predict danger in future states and can reduce the time lag between the system providing an instruction and the driver performing an operation. In addition, the results indicate the effectiveness of the proposed support method.

References

[1] Lotfi A Zadeh. Fuzzy sets. *Information and Control*, 8: 338-353, 1965.

[2] Lotfi A Zadeh. Outline of a New Approach to the Analysis of Complex Systems and Decision Processes. *IEEE Trans. SMC*, 3-1: 28-44, 1973.

[3] E. H. Mamdani. Applications of Fuzzy Algorithms for Control of a Simple Dynamic Plant. *Proc. of IEEE*, 121-12: 1585-1588, 1974.

[4] S. Yasunobu, S. Miyamoto and H. Ihara. Fuzzy Control for Automatic Train Operation System. *4th IFAC/ IFIP/ IFRS Int. Conf. on Transportation Systems*, 39-45, 1983.

[5] S. Yasunobu and S. Miyamoto. Automatic Train Operation System by Predictive Fuzzy Control. *Industrial Applications of Fuzzy Control (M.Sugeno ed.)*, North-Holland, 1-18, 1985.

[6] S. Yasunobu, Fuzzy Engineering, SHOKODO, 1991(in Japanese).

[7] Y. Okamoto and S. Yasunobu. An Intelligent Cooperative Control System Based on Fuzzy Instruction. *Proc. of the 6th International Symposium on Advanced Intelligent Systems(ISIS2005)*, 341-344, 2005.

[8] Y. Chen and S. Yasunobu. Soft Target Based Obstacle Avoidance for Car-like Mobile Robot in Dynamic Environment. *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE2007)*, 1351-1356, 2007.

[9] S.H. Zhou and S. Yasunobu. An Intelligent Cooperative System using Fuzzy Instruction for Car-like Driving. *SICE Journal of Control, Measurement, and System Integration*, 1-4: 293-298, 2008.

[10] J.M.Mendel. Type-2 Fuzzy Sets and Systems: An Overview. *Computational Intelligence Magazine, IEEE*, 2-2: 20-29, 2007.

ISBN: 978-989-95079-6-8

Semantical evaluators

Antoon Bronselaer[1] Guy De Tré[1]
1. Department of Telecommunications and Information Processing, Ghent University
Ghent, Belgium
Email: {antoon.bronselaer,guy.detre}@ugent.be

Abstract— *In the context of a possibilistic framework for detection of object co-reference, evaluators have been defined as operators that compare two values and express the belief that such values are co-referent. Hereby, co-reference of two values means that these values describe the same entity in the real world. In this paper, a class of evaluators is investigated that determines the belief of co-reference based on semantical connections between values of the universe. These semantical connections are modeled by means of binary relations. In case these binary relations are not a-priori given, they can be (partially) learned in an iterative co-reference detection schema.*

Keywords— Possibility theory, co-reference, binary relations

Table 1: Sample of 'restaurant' dataset

name	address	city	type
campanile	624 s. la brea ave.	los angeles	american
campanile	624 s. la brea ave.	los angeles	californian
grill on the alley	9560 dayton way	los angeles	american
grill the	9560 dayton way	beverly hills	american

1 Introduction

In order to deal with the increasing amount of (partially) independent information sources, detection of duplicate information storage is of key importance to avoid inefficient storage management and inconsistencies. In the context of this paper, data are discussed in terms of *objects*. An object is a structured description of a real world entity. Hereby, structured refers to a complex user defined data structure that is used to model data. Such is the case with, among others, relational databases, XML-documents and Object Oriented environments. Objects that refer to the same entity are called *co-referent* objects. A typical property of an object is that it can be decomposed into sub-objects with a *well-defined* domain. These sub-objects are called *attributes*. Note that attributes are not bound to have an atomic data structure, i.e. any collection type is a valid datatype for an attribute. In case of relational databases, the attributes are the tuple fields.

As a running example throughout this paper, the 'restaurant' dataset [1] is used, which is a table from a relational database that describes restaurants. Each restaurant is represented by an object that has four attributes: name, address, city and type. Next to the attributes, the dataset contains a source field, which indicates the restaurant guide from which the information was taken ('fodor' or 'zagat') and an identifier field 'oid'. The goal of this dataset is to find the co-referent restaurants. The dataset is chosen as an example in this paper because syntactical evaluators are not suited for attributes 'city' and 'type'. It is shown in this paper how semantical evaluators can be used for these attributes. A sample of the dataset is shown in Table 1, where objects 1 and 3 have source 'fodor' and objects 2 and 4 have source 'zagat'. Objects 1 and 2 are co-referent objects and so are objects 3 and 4.

In a possibilistic setting, an evaluator E_U determines the possibility (or belief) of (non) co-reference about two values from a universe U. This universe can be the domain of complex objects or the domain of less complex attributes. The subject of this paper is to investigate a special class of evaluators called *semantical evaluators*. These evaluators use binary relations on the universe of interest to determine the belief that two values are co-referent. The transfer of knowledge from a relational level to a comparative level has already been studied in literature. The first notion of semantical connections between linguistic terms is pointed out by Quillian [2]. Resnik uses a network of semantical connections to construct a *semantical similarity measure* [3]. The key idea in this model is to compare the information that is held by two concepts in order to compute the similarity between two words. Further on, Resnik uses an existing taxonomy as knowledge base. This interesting approach has led to several further enhancements, which are all based on information theory [4, 5]. The knowledge base T reflects relations between elements of a universe of concepts. Each approach then uses T to construct a similarity measure. The research presented in this paper is in line of these approaches and provides a theoretical framework for semantical evaluators. The case in which the the taxonomy T is not a-priori known, will be treated in this paper, leading to an unsupervised schema for construction of binary relations. Unsupervised approaches for co-reference have been studied in [6, 7, 8]. The idea proposed here is in fact complementary with the key idea behind these approaches. In recent research, the term co-reference has become relevant in the field of computational linguistics [9, 10, 11]. The search for co-referent *non-structured* objects, i.e. flat text documents, has important applications in text clustering and information aggregation. In this recent field, semantics is of crucial importance to obtain good cluster accuracy. However, it is emphasized that semantical evaluators as introduced in this paper operate in a framework for *structured* objects, in contradiction to the frameworks for *non-structured* objects in [9, 10, 11]. The remainder of the paper is structured as follows. Section 2 summarizes some basic notations and preliminary knowledge. In Section 3, semantical evaluators are defined and properties are investigated. Section 4 introduces the iterative co-reference

ISBN: 978-989-95079-6-8

detection scheme, which is an unsupervised learning schema in which binary relations used by semantical evaluators are increasingly constructed. Finally, Section 5 summarizes the main contributions of this paper.

2 Preliminaries

Within the context of this paper, a (structured) object is a (structured) description of a real world entity that is decomposable into *attributes* and *co-referent* objects are objects that describe the same entity. The possibilistic model for detection of co-referent objects is based on the concept of *possibilistic truth values* [12, 13, 14]. A possibilistic truth value $\tilde{p}$ is a normalized possibility distribution over the domain $\mathbb{B} = \{T, F\}$ and represents uncertainty about the truth value of a proposition p. Hence $\tilde{p} = \{(T, \mu_{\tilde{p}}(T)), (F, \mu_{\tilde{p}}(F))\}$. A short notation thereof is $\tilde{p} = (\mu_{\tilde{p}}(T), \mu_{\tilde{p}}(F))$. The domain of all possibilistic truth values is denoted $\tilde{\wp}(\mathbb{B})$ which is complete lattice. Total order relations $\tilde{R}$ on this domain are defined as follows, with R a total order relation on the unit interval $[0, 1]$:

$$\tilde{p}_1 \ \tilde{R} \ \tilde{p}_2 \Leftrightarrow \begin{cases} \mu_{\tilde{p}_2}(F) \ R \ \mu_{\tilde{p}_1}(F), & \mu_{\tilde{p}_1}(T) = \mu_{\tilde{p}_2}(T) = 1 \\ \mu_{\tilde{p}_1}(T) \ R \ \mu_{\tilde{p}_2}(T), & \text{otherwise} \end{cases}$$

Possibilistic truth values describe available knowledge (or belief) about a proposition. It is emphasized that they do not describe a state of graded truth. The most important operators for knowledge aggregation in the domain of possibilistic truth values are generalized *conjunction*

$$\tilde{\wedge} : \tilde{\wp}(I)^2 \to \tilde{\wp}(I) : \tilde{p} \tilde{\wedge} \tilde{q} \mapsto \{ \ (T, \min\left(\mu_{\tilde{p}}(T), \mu_{\tilde{q}}(T)\right)), \\ (F, \max(\mu_{\tilde{p}}(F), \mu_{\tilde{q}}(F))) \ \}$$

and *disjunction*

$$\tilde{\vee} : \tilde{\wp}(I)^2 \to \tilde{\wp}(I) : \tilde{p} \tilde{\vee} \tilde{q} \mapsto \{ \ (T, \max(\mu_{\tilde{p}}(T), \mu_{\tilde{q}}(T))), \\ (F, \min(\mu_{\tilde{p}}(F), \mu_{\tilde{q}}(F))) \ \}$$

In the context of co-reference detection, possibilistic truth values are generated by *evaluators*. These are formally defined as follows:

Definition 1 (Evaluator)
Assume a universe U of values describing entities in the real world. An evaluator on U is a commutative function E_U such that:

$$E_U : U^2 \to \tilde{\wp}(\mathbb{B}) : (u, v) \mapsto E_U(u, v) = \tilde{p} \qquad (1)$$

with $p =$"u and v are co-referent", $\mu_{\tilde{p}}(T)$ is the possibility that p is true and $\mu_{\tilde{p}}(F)$ is the possibility that p is false.

An evaluator is *reflexive* if $\forall (u, u') \in U^2 : u = u' \Rightarrow E_U(u, u') = \{(T, 1)\}$ and is *strong reflexive* if $\forall (u, u') \in U^2 : u = u' \Leftrightarrow E_U(u, u') = \{(T, 1)\}$.

This paper deals with evaluators that extract knowledge from binary relations. A binary relation R on U is a subset of the Cartesian product $U \times U$. The elements of R are thus couples of elements from U. $(u, v) \in R$ is also denoted as $u \ R \ v$. An equivalence relation is a binary relation that is *reflexive* ($\forall u \in U : u \ R \ u$), *symmetrical* ($\forall (u, v) \in U^2 : u \ R \ v \Leftrightarrow v \ R \ u$) and *transitive* ($\forall (u, v, w) \in U^3 : (u \ R \ v \wedge v \ R \ w) \Rightarrow u \ R \ w$). A partial order relation is a binary relation that is reflexive, transitive and anti-symmetrical ($\forall (u, v) \in U^2 : (u \ R \ v) \wedge (v \ R \ u) \Rightarrow (u = v)$). A *strict* partial order relation is anti-reflexive ($\forall u \in U : \neg(u \ R \ u)$), anti-symmetric and transitive.

3 Semantical evaluators

3.1 Definitions

Evaluators have been introduced and investigated in [15]. In general, three classes of evaluators are distinguished: syntactical, semantical and hybrid evaluators. *Syntactical* evaluators determine possibility based on syntactical differences and are useful to deal with typographical errors in string data and inaccuracies with numerical data. *Semantical* evaluators are based on *binary relations* between elements of a universe U. The name 'semantical evaluator' reflects that in many cases, such binary relations model a semantical connection between elements of the universe. However, the use of the framework presented here is not restricted to semantical connections only. In fact, binary relations can very well express a syntactical connection, rather than a semantical one, for example the *equality* relation. Nevertheless, the applications of interest assume in most cases relations that are semantical in nature, which explains the name of this type of evaluators. *Hybrid* evaluators are a combination of several evaluators on the same domain. Syntactical evaluators are studied in [15], whereas semantical evaluators are the subject of this paper.

The rationale of a semantical evaluator emerges from the fact that co-reference of objects is nothing more than equality of the entities described by the objects. As such it can be seen that co-reference of objects can in fact be modeled by a symmetrical reflexive binary relation (say R_c for convenience) in the domain $O \times O$ with O the set of objects. Moreover, the domain O can be decomposed into sub-domains U_i which are the domains of attributes a_i. Hence, if a binary relation R_i is defined in $U_i \times U_i$, a key question is then how the information given by relation R_i can be translated into information about R_c. It is thus studied in this paper how a given binary relation R on a universe U can be used to generate the possibility that elements of U belong to co-referent objects. An important assumption hereby made is that R contains *positive* knowledge about co-reference, which means that connection of two elements u and v through a binary relation is regarded as evidence for their co-reference. If elements are not connected through R, it is concluded that u and v are not co-referent. This is formalized as follows.

Definition 2 (Semantical evaluator)
Assume a universe U and a binary relation R. A semantical evaluator $E_{U,R}$ is defined by means of a function f_R called the knowledge transfer function, such that $\forall (u, v) \in U^2 : E_{U,R}(u, v) = f_R(u, v)$ with:

$$f_R : U^2 \to \tilde{\wp}(\mathbb{B}) \qquad (2)$$

such that

$$f_R(u, v) = \begin{cases} \tilde{p}_{u,v} & u \ R \ v \vee v \ R \ u \\ (0, 1) & \neg(u \ R \ v) \wedge \neg(v \ R \ u) \end{cases} \qquad (3)$$

The condition $u \ R \ v \vee v \ R \ u$ in Definition 2 is a consequence of the assumption of commutativity in the definition of evaluators. As an example, let $\mathcal{S}$ be the domain of strings and consider the relation R='city-part-of'. If, for the sake of simplicity, it is assumed that $\tilde{p}_{u,v}$ is independent of u and v and equal to $(1, 0.3)$, then:

$$E_{\mathcal{S},R}(\text{"Manhattan"}, \text{"New York"}) = (1, 0.3)$$
$$E_{\mathcal{S},R}(\text{"New York"}, \text{"Chicago"}) = (0, 1)$$

ISBN: 978-989-95079-6-8

The assignment of possibilities by a semantical evaluator is constrained by the specificity of the relation R. Assume two binary relations R_1 and R_2 on universe U such that $R_1 \subset R_2$, then R_2 applies for more couples of elements than does R_1. In an extreme case, a relation R can apply for each couple of elements, which means that R is an equivalence relation on U with 1 equivalence class. Clearly such a relation offers no information about co-reference at all, which means that:

$$R = U^2 \Rightarrow \forall (u, v) \in U^2 : E_R(u, v) = (1, 1) \quad (4)$$

Thus, if more couples satisfy a relation R, R is less specific and less knowledge about co-reference can be derived from R. Hence, a knowledge transfer function must satisfy:

$$(R_1 \subseteq R_2) \Rightarrow \forall (u, v) \in R_1 : f_{R_1}(u, v) \tilde{\geq} f_{R_2}(u, v) \quad (5)$$

The principle of specificity can also be applied inside a relation R. Consider a binary relation R defined over the universe U. The projection of R over an element $u \in U$ is given by:

$$\text{Proj}_u(R) = \{(x, y) | x \; R \; y \wedge (x = u \vee y = u)\} \quad (6)$$

The knowledge transfer from R to $\tilde{\wp}(\mathbb{B})$ modeled by f_R is *monotone* if and only if for any triple $(u, v, w) \in U^3$ that satisfies $uRw \vee wRu$ and $vRw \vee wRv$, we have that:

$$|\text{Proj}_u(R)| \geq |\text{Proj}_v(R)| \Leftrightarrow f_R(u, w) \tilde{\leq} f_R(v, w) \quad (7)$$

It can be verified that monotone knowledge transfer is obtained if f_R is constructed as follows:

$$\begin{array}{l} u \; R \; v \vee v \; R \; u \Rightarrow \\ f_R(u, v) = \left(1, g\left(\frac{|\text{Proj}_u(R)|}{2|R|-1}, \frac{|\text{Proj}_v(R)|}{2|R|-1}\right)\right) \end{array} \quad (8)$$

with g a monotonic increasing, commutative function such that $g(1, 1) = 1$. Monotone transfer knowledge implies that it is *more* certain that u and v are co-referent if they are connected by R and if there is *less* evidence that u or v are co-referent with other elements.

A common problem with semantical evaluators is that relations are a-priori information. In other words, relations are constructed without observing the actual values on which the evaluator must operate. It is possible that the values on which R is constructed are not *equal* to values of an observed problem, but rather co-referent. This is caused by the fact that values considered for a-priori construction of R are typically standardized, which is not the case for values that are the subject of co-reference detection. For example, with R='city-part-of', it is known that "Manhattan" R "New York". Nevertheless:

$$E_{\mathcal{S},R}(\text{"Manhatan"}, \text{"New York"}) = (0, 1)$$

due to the spelling error in "Manhatan". To deal with such a situation, an evaluator can be used to detect co-references between values from an observed problem and values from the a-priori universe on which R is constructed. Typically, such an evaluator will be *syntactical*. A semantical evaluator is then used in a *chain* of evaluators.

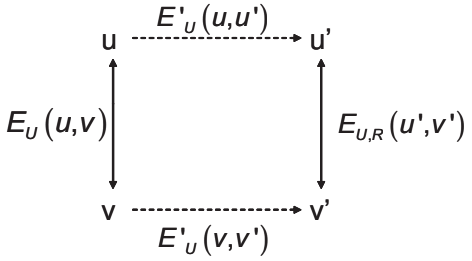


Figure 1: Evaluator chain

Definition 3 (Evaluator chain)
Assume a universe U equipped with a semantical evaluator $E_{U,R}$ and a reflexive evaluator E'_U. An evaluator chain is an evaluator E_U such that:

$$E_U(u, v) = \max_{(u', v')} E'_U(u, u') \tilde{\wedge} E_{U,R}(u', v') \tilde{\wedge} E'_U(v', v) \quad (9)$$

The principle of chaining is schematically depicted in Figure 1. Due to reflexivity of E'_U it can be seen that:

$$(u \; R \; v \vee v \; R \; u) \Rightarrow E_U(u, v) = E'_U(u, v) \quad (10)$$

Moreover, if E'_U is strong reflexive:

$$(u \; R \; v \vee v \; R \; u) \Leftrightarrow E_U(u, v) = E'_U(u, v) \quad (11)$$

In the context of iterative co-reference detection (Section 4), an important concept is that of a hybrid evaluator.

Definition 4 (Hybrid evaluator)
Assume a universe U and a set of evaluators $\mathcal{E} = \{E_{U,1}, ..., E_{U,n}\}$ on U. A hybrid evaluator is an evaluator on U such that:

$$E_U = c(E_{U,1}, ..., E_{U,n}) \quad (12)$$

with c a commutative, associative and monotonic increasing function called the combinatory function.

3.2 Examples

The most obvious example of a semantical evaluator is an evaluator that is based on the equality relation. In many cases, reflexivity of an evaluator is required which means that $\forall (u, v) \in U^2 : u = v \Rightarrow E_U(u, v) = (1, 0)$. Hence, in order to support this constraint, equality of elements must imply full impossibility of non co-reference. As mentioned in Section 3.1, this evaluator can be regarded as syntactical evaluator, because the binary relation is not semantical in nature. A related case is a semantical evaluator where R is relaxed to an *equivalence* relation, which has interesting applications such as detection of synonyms, i.e. words with the same meaning. Another example is the case where two words don't have the same meaning, but they both fully qualify for description of a given entity. This is because the entity can not be described exactly and more descriptions are possible. In that case, words are equivalent with respect to their appropriateness for description of an entity, despite the fact that these words do not have the same meaning. This is the case with the attribute 'type' of the 'restaurant' dataset, where for example 'steakhouse' and 'american' are possible (and thus equivalent) descriptions of the type of restaurant.

ISBN: 978-989-95079-6-8

Assume a universe U and assume a partition of U into n disjunct subsets U_i. Consider $n-1$ relations R_i such that each R_i is a surjective function between U_i and U_{i+1}. The set of relations $\{R_i\}_{i=1,2,...,n-1}$ is called a *hierarchy* on U. It can be seen that, by definition, each R_i from a hierarchy is anti-reflexive and anti-symmetrical. Due to the fact that $\forall i \in \{1,...,n-1\} : im(R_i) = dom(R_{i+1})$, it is possible to compose relations R_i and R_{i+1}.

$$\forall (u,w) \in U^2 : u\ R_1R_2\ w \Leftrightarrow (\exists v \in U : u\ R_1\ v \wedge v\ R_2 w) \tag{13}$$

Hence, $R_1R_2...R_n$ denotes the n-ary composition of n relations and thus, elements of non subsequent partitions are related through composition. For each hierarchy $\{R_i\}_{i=1,2,...,n-1}$, consider the set of relations $\{R_i^*\}$ such that:

$$u\ R_i^*\ v \Leftrightarrow \exists k \leq l \leq i : u\ R_k...R_l v \tag{14}$$

By definition:

$$\forall i,j \leq n : i < j \Rightarrow R_i^* \subset R_j^* \tag{15}$$

It is then possible to construct $n-1$ semantical evaluators with knowledge transfer function $f_{R_i^*}$. Due to the consistency constraint:

$$\forall i,j \leq n : i < j \Rightarrow f_{R_i^*} \tilde{\geq} f_{R_j^*} \tag{16}$$

It can be seen that composing all relations, leads to a strict partial order relation. Hence, each partial order relation can be decomposed into relations with higher granularity and thus higher belief of co-reference. This principle can be applied on the 'city' attribute of the restaurant dataset, because the values are connected by a strict partial order relation.

4 Iterative co-reference detection

4.1 Approach

In this Section, attention is given to the case where R is not a-priori given. This is the case when dealing with subjective information, for example the 'type' attribute of the restaurant dataset. There is no a-priori information about which types can be used to denote the same restaurant. In that case a semantical evaluator can be (partially) constructed based on a training set. However, in what follows a method is proposed to construct a semantical evaluator in cases where no training set is given. Assume objects consisting of n attributes a_i. For each attribute a_i, an evaluator $E_{dom(a_i)} = E_i$ must be constructed. The set of attributes can thus be partitioned into attributes that require a syntactical evaluator (A_{syn}) and attributes that require a semantical evaluator (A_{sem}), such that $A = A_{syn} \cup A_{sem}$ and $A_{syn} \cap A_{sem} = \emptyset$. For the 'restaurant' dataset, $A_{syn} = \{$'name', 'street'$\}$ and $A_{sem} = \{$'type', 'city'$\}$.

The key idea of iterative co-reference detection, is to find the co-referent objects in m iterations. Semantical evaluators for which no binary relation R is available are gradually constructed throughout the different iterations. An evaluator is called *unavailable* if it is a semantical evaluator for which the binary relation R is unknown. In each iteration j, a subset of the available evaluators is selected. With these evaluators, a candidate set C_j of co-referent object couples is generated. For each attribute a_i that requires a semantical evaluator, a binary relation R_j is constructed by projecting each object couple from C_j over the attribute a_i. In other words, if the object couple (o_1, o_2) is contained in the candidate set C_j and attribute a_i requires a semantical evaluator, then the couple $(\text{Proj}_{a_i}(o_1), \text{Proj}_{a_i}(o_2))$ is considered as a couple of co-referent values for attribute a_i. As a consequence, this couple of values is added to the binary relation R_j. For each attribute a_i that requires semantical evaluation, a novel relation is R_j is constructed in each iteration j. The reason to do this is explained next. Typically, the generation of the first candidate set is based on very stringent criteria. If this is not the case, many false positives are included in the candidate set, which implies that constructed binary relations contain a large amount of couples. As explained before, a binary relation with large cardinality implies inferred knowledge with much uncertainty (Eq. 4, Eq. 5). Throughout sequential iterations, the candidate set is expanded by (i) relaxation of syntactical evaluators and (ii) additional knowledge of semantical evaluators. As a consequence, there is a high probability that for $k \leq l$, $C_k \subset C_l$. If this is true, relations constructed in iteration k are subsets of the corresponding relations in iteration l. Due to the consistency constraint (Eq. 5), relations constructed in iteration l imply more uncertainty. Therefor, in order to preserve the level of (un)certainty in relations constructed in early iterations, each iteration constructs a new relation. Hence, for each attribute a_i that requires a semantical evaluator, m binary relations are constructed. Each of these relations $R_{i,j}$ yields a semantical evaluator $E_{i,R_{i,j}}$. The constructed evaluator for attribute a_i is a hybrid evaluator (Eq. 12) with $\tilde{\vee}$ as combinatory function:

$$E_i(u,v) = E_{i,R_{i,1}}(u,v)\tilde{\vee}...\tilde{\vee}E_{i,R_{i,m}}(u,v) \tag{17}$$

Hence, the constructed evaluator tries to find the most stringent binary relation by which the values are connected as this relation yields the most belief in co-reference. After iteration m, the last candidate set C_m of co-referent couples is generated which gives us the possible co-referent objects. It is noted that iterative co-reference detection is a stand-alone comparison process: it produces a set of possible co-referent couples. However, it can also be used to construct evaluators and their corresponding binary relations. In other words, the proposed method can be used to find co-referent objects, but it's purpose

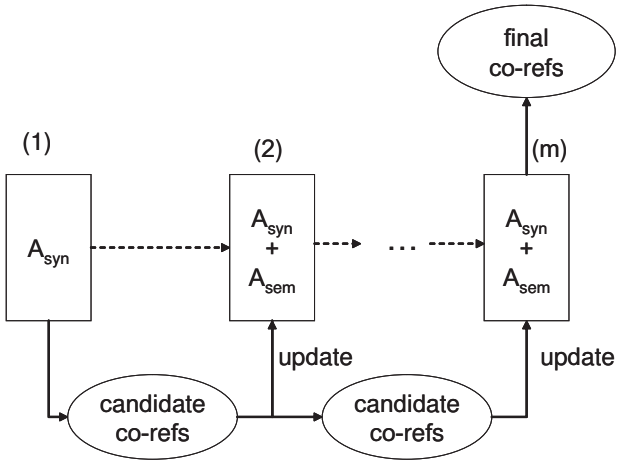


Figure 2: Iterative learning of semantical evaluators

ISBN: 978-989-95079-6-8

Table 2: Statistics of restaurant test

'type' evaluator	Co-refs	Non Co-refs
Equality	22	5018
Equality + R	100	27811

can also be limited to construction of binary relations, which makes it compatible with other (possibilistic) frameworks for co-reference detection. Figure 2 shows a schematical representation of iterative co-reference detection.

4.2 Example

The idea of iterative co-reference detection is illustrated on the 'restaurant' dataset. The focuss in this example lies on illustrating the construction of binary relations. The dataset consists of two lists ($|L_1| = 533$ and $|L_2| = 331$). Among the 176423 couples of restaurants, 112 couples are co-referent. For convenience, only two attributes are initially considered: 'name' and 'type'. It is assumed that a syntactical evaluator is used for 'name' and a semantical evaluator must be constructed for 'type'. For the first iteration, consider an evaluator for attribute 'name' based on equality (i.e. values are co-referent if they are equal). Among all possible couples, 83 couples have an equal name, of which 82 are identified as co-referent, based on ground truth (i.e. equal oid's). Hence, candidate set C_1 contains 83 couples of restaurants. Projecting this candidate set over the 'type' attribute and filtering identical couples out, yields 40 couples of possible co-referent restaurant types. These couples determine the binary relation R_1. Table 2 learns how R_1 allows for assignment of belief to 100 co-referent couples (89.3% of all co-referent couples). However, 27811 non-coreferent couples are also assigned belief (15.8% of all non co-referent couples). This is partially due to the type of data. As can be seen, 5018 non-coreferent couples have equal values for type, which is an indication that the 'type' attribute is prone to many false positives. It must be remembered that the goal is to construct an evaluator that is used in a larger object comparison system. Hence, the contribution of semantical evaluators lies in the assignment of belief to co-referent values, thereby trying to minimize the belief attached to non co-referent values. As can be seen in Table 2, the fraction of non-coreferent values that are assigned belief can still be large in absolute numbers. However, this fraction can be filtered out using an intelligent aggregation function to combine the belief assigned by the different evaluators. The constructed relation for 'type' assigns belief to 89.3% of the co-referent couples, after one iteration. Assume in a second iteration, that the 'street' attribute is involved, which requires a syntactical evaluator. For the sake of simplicity, assume an evaluator for 'street' based on equality. If the evaluations for 'name' and 'street' are combined through disjunction, i.e. two restaurants are co-referent if their names are equal *or* their streets are equal, then the candidate set C_2 contains 106 object couples of which 101 are actually co-referent. Suppose that iteration 2 would be the last iteration, then the model can be refined by adding semantical evaluation of the 'type' attribute, i.e. two restaurants are co-referent if their types are connected by R_1 *and* (names are equal *or* their streets are equal). Using this model results in a candidate set of 95 couples of which 94 are co-referent, which means that 83.9% of all co-referent couples are detected with a precision of 98.95%. If iteration 2 is not the last iteration, then C_2 can be used to construct a second binary relation R_2. Note that by construction $R_1 \subset R_2$.

An advantage of iterative co-reference detection is that the binary relations R are dynamically adapted throughout the different iterations. If an approach with a training set is used, the relation R can be partially constructed only on the observations in the training set. With iterative co-reference detection, binary relations are constructed by generating a candidate set from *all* known observations. Hence, the relations constructed in an iterative co-reference detection scheme will be more complete.

5 Conclusion

In the possibilistic framework for detection of complex co-referent objects, a central concept is that of an evaluator, which is a function that estimates belief/uncertainty about co-reference of (sub)-objects. This paper introduces a formal framework for *semantical* evaluators E_U that model the transfer of knowledge from binary relations on the universe, to knowledge about co-reference of elements. This transfer is modeled by means of a knowledge transfer function f_R. It is shown that, depending on the application, such a function must satisfy consistency constraints. It is also shown how semantical evaluators can be used in an evaluator chain and how to combine evaluators in a hybrid evaluator. Finally, when relations on U are unknown, it is investigated how to (partially) construct such relations without use of a pre-labeled training set. This construction process is called iterative co-reference detection.

References

[1] Sheila Tejada, Craig Knoblock, and Steven Minton. Learning object identification rules for information integration. *Information Systems*, 26(8):607–633, 2001.

[2] Ross Quillian. *Semantic Information Processing*, chapter Semantic memory. MIT Press, 1968.

[3] Philip Resnik. Using information content to evaluate semantic similarity in a taxonomy. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pages 448–453, 1995.

[4] Jay Jiang and David Conrath. Semantical similarity based on corpus statistics and lexical taxonomy. In *Proceedings of International Conference Research on Computational Linguistics*, pages 15–33, 1997.

[5] Yuhua Li. An approach for measuring semantic similarity between words using multiple information sources. *IEEE Transactions on Knowledge and Data Engineering*, 15(4):871–882, 2003.

[6] Matthew Jaro. Advances in record linking methodology as applied to the 1985 census of tampa florida. *Journal of the American Statistical Society*, 84(406):414–420, 1989.

[7] William Winkler. Improved decision rules in the fellegi-sunter model of record linkage. Technical Report RR93/12, Statistical Research Report Series, 1993.

[8] William Winkler. Methods for record linkage and bayesian networks. Technical Report RRS2002/05, Statistical Research Report Series, 2002.

[9] Wee Meng Soon, Hwee Tou Ng, and Daniel Chung Yong Lim. A machine learning approach to coreference resolution of noun phrases. *Computational Linguistics*, 27(4):521–544, 2001.

ISBN: 978-989-95079-6-8

[10] Vincent Ng and Claire Cardie. Improving machine learning approaches to coreference resolution. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 104–111, Philadelphia, 2002.

[11] Xiaoqiang Luo and Imed Zitouni. Multi-lingual coreference resolution with syntactic features. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 660 – 667, 2005.

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

[13] Henri Prade. Possibility sets, fuzzy sets and their relation to lukasiewicz logic. In *Proceedings of the International Symposium on Multiple-Valued Logic*, pages 223–227, 1982.

[14] Gert De Cooman. Towards a possibilistic logic. in: Fuzzy set theory and advanced mathematical applications, edited by da ruan, kluwer academic, pp. 89–133, boston., 1995.

[15] Antoon Bronselaer, Axel Hallez, and Guy De Tré. Evaluation in the possibilistic framework for object matching. In *Proceedings of the IPMU 08 Congres*, pages 1701–1708, Malaga, 2008.

ISBN: 978-989-95079-6-8

A Hybrid Meta-Heuristic to Solve the Portfolio Selection Problem

José M. Cadenas[1] Juan V. Carrillo[1] M. Carmen Garrido[1]
Carlos Ivorra[2] Teresa Lamata[3] Vicente Liern[2]

1.Dep. Ingeniería de la Información y las Comunicaciones. Universidad de Murcia.
2.Dep. Matemàtiques Economia i Empresa, Universidad de Valencia.
3.Dep. Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada.
Emails: jcadenas@um.es, jvcarrillo@um.es, carmengarrido@um.es,
carlos.ivorra@uv.es, mtl@decsai.ugr.es, vicente.liern@uv.es

Abstract— *We present a fuzzy version of the efficient portfolio selection problem which adds to the Markowitz's classical model the vagueness of the investor's preferences about the assumed risk. This is done by adapting some techniques previously applied by the authors in other logistic problems. We provide several procedures to solve it: an exact one and a hybrid meta-heuristic procedure more adequate for medium and large-sized problems. To show the usefullness of the model and the algorithms, we provide an example based on data from the Spanish index IBEX35.*

Keywords— Fuzzy programming, Heuristic strategies, Portfolio selection.

1 Introduction

The modern portfolio selection problem is a classical model for determining the optimal composition of a portfolio according to the preferences of an investor. The model is credited to Markowitz [17, 18], who is rightfully regarded as the founder of modern portfolio theory [6].

In a first approach, given n securities S_1, $S_2, \cdots, S_n$ in which we can invest non-negative quantities $x_1, x_2, \cdots, x_n$, a portfolio consists of a subset of securities in which all the capital should be invested satisfying investor's preferences. These preferences must take into account the risk that the investor is willing to assume, the expected return and, to a lesser extent, some diversification criteria.

One possibility is taking the return as objective function and including the risk as a constraint, i.e.

$$\begin{array}{ll} \text{Max.} & \text{Portfolio return} \\ \text{s.t.} & \text{Risk of portfolio} \leq R \\ & \sum_{i,=1}^{n} x_i = 1, \\ & l_i \leq x_i \leq u_i, \qquad i = 1, \ldots, n. \\ & l_i \geq 0 \qquad\qquad i = 1, \ldots, n. \end{array} \tag{1}$$

where R is the maximum risk accepted by the investor, and the last constraints express that all the capital must be invested as well as diversification conditions.

However, we could also consider the dual version (in an economic sense) in which the risk is minimized taking the expected return as a constraint.

Obviously, the portfolio selection, like most financial problems, is related with uncertainty because it consists in taking a decision about future events. Therefore, we do not have at our disposal more than historical data which usually are managed with statistical methods [18, 6]. Moreover, it is not easy to model the investor's preferences. After the seminal work by Markowitz, attention has been given in the study of alternative models [13, 14]. Most of these models are based on probability distributions, which are used to characterize risk and return. For instance, in the Markowitz model, variance and mean were generally deemed satisfactory measures of risk and return, respectively [18, 21, 7, 12]. However, according to the chosen modellization of the expected risk and return, different models coexist to select the best portfolio. Some of them propose to describe the risk by means of mean-absolute deviation ([11], [24]), other authors propose linear programming [22] or multiobjective [15] models.

Another way of dealing with uncertainty is working with models based on soft computing. This can be done by means of vague goals for the expected return rate and risk [23, 16], or by using possibility distributions to model the returns. This fact permits the incorporation of expert knowledge by means of a possibility grade, to reflect the degree of similarity between the future state of stock markets and the state of previous periods.

In this paper, we present several ways of dealing with the uncertainty associated to the portfolio selection problem. Namely, we are going to show that the techniques used by the authors in logistic problems[2, 3, 4] can be generalized to deal with the risk and return fuzziness in the portfolio selection problem. These techniques allows us to include in our fuzzy model the vagueness related with the fact that the investor fixes arbitrarily the risk that he/she will assume. The uncertainty concerning to the valuation of risk and returns is modellized by means of the usual variance and mean.

To solve our model, meta-heuristic techniques are often needed because of the size of the problem or because a quick solution is required in order to reflect real-time cases. Therefore, we propose a hybrid meta-heuristic procedure.

The paper is organized as follows: The next section introduces the framework for portfolio selection used in this paper. Section 3 examines the portfolio selection with flexible constraints, incorporating this flexibility in a fuzzy model. In section 4, we present a hybrid meta-heuristic to solve the portfolio selection problem with flexible constraints. Finally, Section 5 outlines the most important conclusions.

ISBN: 978-989-95079-6-8

2 Portfolio selection

We could consider the problem of an investor who intends to invest some money in securities in such a way that the rate of return is maximized and a given level of risk cannot be surpassed. If we have n securities, r_i is the return of the i-th security and x_i is the proportion of total investment funds devoted to the i-th security, then we have the constraint

$$\sum_{i=1}^{n} x_i = 1.$$

The risk of investment can be measured by the variance $x^t S x$, where $S = [s_{ij}]$ is the variance-covariance matrix. If R is the maximum risk accepted by the investor, we have the constraint

$$\sum_{i,j=1}^{n} s_{ij} x_i x_j \leq R.$$

Then, one of the simplest models for the portfolio selection problem is:

$$\begin{array}{rl} \text{Max.} & \displaystyle\sum_{i=1}^{n} r_i x_i \\ \text{s.t.} & \displaystyle\sum_{i=1}^{n} x_i = 1, \\ & \displaystyle\sum_{i,j=1}^{n} s_{ij} x_i x_j \leq R, \\ & x_i \geq 0, \;\; i = 1, \ldots, n. \end{array} \tag{2}$$

However, choosing the best investment options for a portfolio is a difficult task due to economic environment uncertainty and the problem of suitably reflecting decision maker desires in the model. Both stochastic and fuzzy programming provide different ways of handling the first kind of uncertainty [9]. This paper presents a fuzzy approach for the second kind, in which the portfolio problem includes the subjective criteria of the decision maker to determine the level of risk that he or she is able to support and the level of satisfaction to be assigned to a possible increase in return.

3 Flexible Portfolio selection

The portfolio selection problem has two data concerning decision maker preferences, namely the capital to be invested and the risk to be assumed. The investor can be assumed to know with certainty the capital that he/she would like to consider, and in fact, in the model this quantity has been normalized to the unit. However, determining the risk to be assumed could be more flexible. As a result, it is worth incorporating this flexibility in a fuzzy model.

The main idea is to consider partially feasible solutions involving slightly greater risk than that fixed by the decision maker, and to study the possibilities that they offer in order to improve the expected return. When compared with the logistic models (actually with the p-median case [2, 3, 4]), this problem happens to be more complicated, because the p-median problem is linear whereas the risk constraint in the portfolio model is quadratic. Moreover, in the p-median case, a small reduction in covered demand affected optimal cost in a simple linear way, whereas the way in which the maximum expected return depends on the accepted risk is rather more complicated.

A fuzzy set $\tilde{S}$ of partially feasible solutions is defined so that portfolio selection belongs to $\tilde{S}$ with a degree of membership that depends on how much it exceeds the risk R_0 fixed by the investor. On the other hand, a second fuzzy set $\tilde{G}$ is defined, whose membership function reflects the improvement of the return provided by a partially feasible solution with respect to the optimal crisp return z^*. In practice, we consider piecewise linear membership functions

$$\mu_{\tilde{S}}(x) = \begin{cases} 1 & \text{if } r \leq R_0, \\ 1 - \frac{r - R_0}{p_f} & \text{if } R_0 < r < R_0 + p_f, \\ 0 & \text{if } r \geq R_0 + p_f, \end{cases}$$

$$\mu_{\tilde{G}}(x) = \begin{cases} 0 & \text{if } z \leq z^*, \\ \frac{z - z^*}{p_g} & \text{if } z^* < z < z^* + p_g, \\ 1 & \text{if } z \geq z^* + p_g, \end{cases}$$

where r and z are the risk and the return provided by the portfolio x (which is assumed to satisfy the constraints of (2), except the second one); the parameter p_f is the maximum increment in the risk that the decision maker can accept, and p_g is the increment of the return that the decision maker would consider completely satisfactory. From this, we can define a global degree of satisfaction

$$\lambda(x) = \min\{\mu_{\tilde{G}}(x), \mu_{\tilde{S}}(x)\},$$

which is the membership degree to the fuzzy intersection of $\tilde{S} \cap \tilde{G}$. The fuzzy portfolio model becomes

$$\begin{array}{rl} \text{Max.} & \lambda(x) \\ \text{s.t.} & x \in \tilde{S}. \end{array} \tag{3}$$

In order to solve it, the optimal solution of (2) will be calculated for each risk level R. Therefore, we solve explicitly the Kuhn-Tucker conditions for the problem. To carry out the computations in a generic framework, we start making a change of variables to diagonalize the risk matrix. As we are interested in small variations of R, the variables x_i that are zero in the optimal crisp portfolio can be removed and hence assume that non-negative conditions are not active. We can also assume that the risk constraint is active.

Standard linear algebra theory ensures us that we can decompose

$$S = A^t D A, \tag{4}$$

where the matrix D is diagonal and A is regular. Then the change of variables $y = Ax$ transforms the problem to

$$\begin{array}{rl} \text{Max.} & \displaystyle\sum_{i=1}^{n} r'_i y_i \\ \text{s.t.} & \displaystyle\sum_{i=1}^{n} b_i y_i = 1, \\ & \displaystyle\sum_{i=1}^{n} d_i y_i^2 = R, \\ & A^{-1} y \geq 0, \end{array} \tag{5}$$

where $r' = r^t A^{-1}$, $b = (1, \ldots, 1) A^{-1}$.

ISBN: 978-989-95079-6-8

As nonnegative conditions are inactive, the optimal solution must satisfy

$$\sum_{i=1}^{n} b_i y_i = 1, \quad \sum_{i=1}^{n} d_i y_i^2 = R, \quad r'_i - b_i\lambda - 2d_i y_i \mu = 0,$$

where λ and μ are the Lagrange multipliers of the capital and risk constraints respectively. Hence:

$$y_i = \frac{r'_i - b_i\lambda}{2d_i\mu},$$

and, from the first constraint, we get:

$$\sum_{i=1}^{n} \frac{b_i r'_i}{2d_i} - \sum_{i=1}^{n} \frac{b_i^2}{2d_i}\lambda = \mu.$$

Call

$$K = \sum_{i=1}^{n} \frac{b_i r'_i}{2d_i}, \qquad L = \sum_{i=1}^{n} \frac{b_i^2}{2d_i},$$

so that $\mu = K - L\lambda$. Hence:

$$y_i(\lambda) = \frac{r'_i - b_i\lambda}{2d_i(K - L\lambda)}.$$

The second constraint gives us:

$$\sum_{i=1}^{n} \frac{(r'_i - b_i\lambda)^2}{4d_i} = R(K - L\lambda)^2$$

or

$$\sum_{i=1}^{n} \frac{r'^2_i}{4d_i} - \sum_{i=1}^{n} \frac{r'_i b_i}{2d_i}\lambda + \sum_{i=1}^{n} \frac{b_i^2}{4d_i}\lambda^2 = R(K^2 - 2KL\lambda + L^2\lambda^2),$$

$$\sum_{i=1}^{n} \frac{r'^2_i}{4d_i} - K\lambda + \frac{L}{2}\lambda^2 = R(K^2 - 2KL\lambda + L^2\lambda^2).$$

If

$$M = \sum_{i=1}^{n} \frac{r'^2_i}{4d_i} \tag{6}$$

we get

$$(RL^2 - L/2)\lambda^2 + (K - 2KLR)\lambda + RK^2 - M = 0.$$

This is a second degree equation on λ with discriminant

$$\Delta(R) = K^2 - 2ML + (4ML^2 - 2K^2L)R.$$

Hence,

$$\lambda(R) = \frac{2KLR - K \pm \sqrt{\Delta(R)}}{2L^2R - L},$$

and so

$$\mu(R) = K - L\lambda(R) = \frac{\pm\sqrt{\Delta(R)}}{1 - 2LR}.$$

The sign must be chosen so that $\mu(R) \geq 0$ (the Kuhn-Tucker sign condition). Once the right sign is chosen, this function $\lambda(R)$ allows us to calculate $y_i(R)$ and, from this, we get the optimal portfolio $x_i(R)$ for each R such that $x_i(R) > 0$. This reduces the range of R to a certain interval that can be computed. When we reach one end point of this interval, we need to start again with the problem corresponding to the new set of positive variables.

We can also compute the return

$$F(R) = \sum_{i=1}^{n} r'_i y_i(R).$$

Since

$$y_i(\lambda) = \frac{r'_i - b_i\lambda}{2d_i\mu} = \mp\frac{2LR-1}{2d_i\sqrt{\Delta(R)}}\left(r'_i - b_i\frac{2KLR - K \pm \sqrt{\Delta(R)}}{2L^2R - L}\right), \quad 1 \leq i \leq n,$$

we have

$$F(R) = \mp\frac{2LR-1}{2\sqrt{\Delta(R)}}\sum_{i=1}^{n}\left(\frac{r'^2_i}{d_i} - \frac{r'_i b_i}{d_i}\frac{2KLR - K \pm \sqrt{\Delta(R)}}{2L^2R - L}\right).$$

This expression allows us to calculate the degree of improvement of the goal $\mu_g(R)$ of the best portfolio with risk R, whereas its degree of feasibility $\mu_f(R)$ is trivially computed.

Now we can determine the risk R^* such that $\mu_f(R^*) = \mu_g(R^*)$, which is easily shown to be the risk of the portfolio maximizing λ. The portfolio $x(R^*)$ corresponding to $y(R^*)$ by the change of variables is the optimal solution of (3).

4 A hybrid meta-heuristic to solve the Portfolio Selection problem

When the size of the problem increases, traditional methods become useless as they would need too much time, due to the combinatorial explosion in the solution space. Genetic Algorithms [19, 5] are meta-heuristic methods that have already shown their kindness to solve optimization problems.

In this section we describe a hybrid meta-heuristic for the portfolio selection problem with flexible constraints, called GAFUZ-PF (Genetic algorithm + simulated Annealing + FUZzy PortFolio). This meta-heuristic uses a hybrid scheme, mixing the ideas of Simulated Annealing technique with the classic Genetic Algorithm. Note that we need the optimal crisp return z^* of the portfolio selection problem (2) because $\mu_{\tilde{G}}(x)$ depends on it. Hence, solving the crisp problem must be considered as a part of the process of solving the fuzzy problem (this means to solve two NP-hard problems). In order to realize this idea, we use three different populations on three stages of the genetic algorithm. First of all, it uses a population to achieve feasible solutions to the problem (with a risk less than or equal to R fixed) which will be part of the population of the algorithm. Once feasible solutions are found, in the second stage, the algorithm uses other population that evolves searching the best return. The third stage uses another population, which looks for the best fitness $\lambda(x)$.

Simulated Annealing [1, 10] is a meta-heuristic used for optimization problems. This heuristic replaces the current solution by a random "nearby" solution, chosen with a probability that depends on the difference between the corresponding

ISBN: 978-989-95079-6-8

function values and a global parameter T (called the temperature), that is gradually decreased during the process. The use of this idea with Genetic Algorithm is suggested by Mitchell [19], using it with the individual selection function. We apply Simulated Annealing to mutation function. Having a mutation function with a big mutation rate helps us to keep diversity in the population, but it introduces distortion on the population when the algorithm approaches to the optimum. So, it seems a good idea to use Simulated Annealing with a temperature depending on the number of generations and the fitness of the solution in order to make possible a big mutation rate at the beginning, and reduce it while increasing the number of generations and the fitness. Now we specify our options and we give a general outline of our approach. Figure 1 shows the scheme of `GAFUZ-PF` hybrid meta-heuristic.

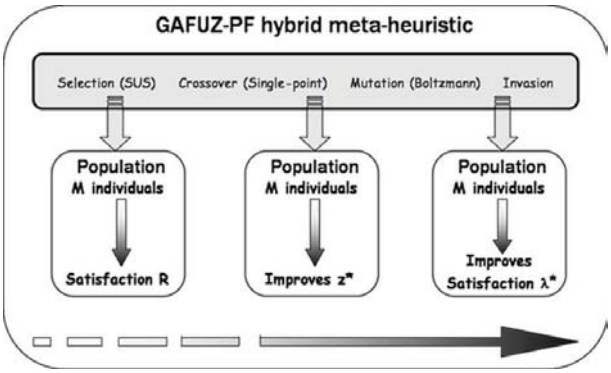


Figure 1: Scheme of hybrid meta-heuristic

4.1 Encoding, Fitness function and Population size

Each element of the population of the algorithm represents a solution to the described portfolio problem. The population encoding is made having an array of n-elements, where each gene represents the invest on one asset. The size of the population is fixed to 50 individuals, as it is known to be the best average size to avoid a too slow algorithm [19]. Furthermore, we use three different populations against time: the first one evolves searching for individuals to satisfy the second constraint of the problem 2. When it founds a suitable solution, it changes to the second stage, where it improves the return of the portfolio, taking account of the flexible constraint of R. Once the algorithm gets a state with $\lambda(x) > 0$, it changes the population and evolves taking into account p_f and p_g. The fitness function is the objective function of the problem (3).

The algorithm scheme is the classical one, with selection, crossover, mutation, invasion and having the number of generations as stop criteria.

4.2 Selecting the individuals

In order to select individuals for the crossover operator, we use the Stochastical Universal Sampling (SUS), as it has been shown to be better than Roulette Wheel [5]. Based on their fitness, SUS places the individuals on a Roulete and spuns it once with n selectors equally separated to select n individuals.

4.3 Crossover: generating new individuals

The crossover operator is the single-point crossover [19] adapted to the problem constraints: there is a fixed part from one parent that goes directly to the son; the genes from the second part (from the second parent) are added to the son (choosing them randomly) while the sum of its genes is ≤ 1.

4.4 Mutation

When we talk about mutation, there are some parameters used. The mutation rate (MUR) refers to the percentage of population individuals that will be mutated. On the other hand, we have the number of genes that will change (NR). As a mutation scheme we use Boltzmann mutation which is based on Simulated Annealing [5], with the following annealing function:

$$r = 1 - e^{-1/(n*max(\lambda(x)))}$$

It defines the mutation rate across the algorithm, reducing it as the algorithm evolves in order not to disturb the good solutions achieved when the algorithm is finishing. This rate specifies the percentage of individuals that would be mutated, since the number of mutations per individual is 1, as it is known to be the best choice [20]. A mutation consists in substract a random value from a gene and add this amount to other gene, both of them choose randomly.

4.5 Invasion

Finally, invasion is another genetic operator that helps to keep diversity in the population. It introduces a number of new randomly generated individuals, which replace some individuals at random in the population (except the best one). This helps to introduce new genes in the population, like mutation does. The invasion rate (IR) selected is 1% .

4.6 Stop criteria

The detention criterion is the number of generations (NG).

4.7 General outline of the `GAFUZ-PF` meta-heuristic

Now we have shown the configuration of the algorithm, we will see its general outline in algorithm 1.

4.8 Experiments

To test `GAFUZ-PF` hybrid meta-heuristic we have used various test problems. The PC used for the executions has the following features: Intel Pentium IV 3.00 GHz 2048 MB RAM.

• We have considered the returns on 20 assets from the Spanish index IBEX35. The set of assets included in the experiment represents Acesa (ACE), Arcelor (ACR), ACS, Altadis (ALT), BBVA, Bankinter (BKT), Dragados (DRC), Endesa (ELE), FCC, Iberdrola (IBE), Metrovacesa (MVC), NH Hoteles (NHH), Banco Popular (POP), Repsol (REP), SCH (SAN), Telefónica (TEF), Unión Fenosa (UNF), Vallermoso (VAL), Acerinox (ACX), Acciona (ANA) data respectively. We have considered the observations of the Wednesday prices as an estimate of the weekly prices. Hence, the return on the j- th asset during the k-th week is defined as $r_{kj} = (p_{(k+1)} - p_{kj})/p_{kj}$, where p_{kj} is the price of the j-th asset on the Wednesday of the k-th week. The used data base covers the period from January 1998 to March 2003. Table 1 contains the obtained results for two different risk levels R

ISBN: 978-989-95079-6-8

Algorithm 1 - GAFUZ-PF: Hybrid meta-heuristic for the portfolio selection

GAFUZ-PF( )
Require: r_i, s_{ij}, R, p_f, p_g
Ensure: z^*, $\overline{x}^*$, R^*, $\lambda(x)$, z, $\overline{x}$, R
begin
Generate initial population with 50 individuals at random.
Calculate the fitness of each individual.
while Number of generations $< NG$ **do**
*S*ELECT 50 individuals
if $\nexists$ individual | risk of the individual $\leq R$ **then**
select using risk
else
if the sum of individual fitness is = 0 **then**
select using z^*
else
select using fitness
end if
end if
*C*ROSS the 50 individuals to obtain 50 new individuals
Replace 49 individuals of the population with the new individuals, keeping the best (elitism)
MUR=Boltzmann(max.($\lambda(x)$), generation_number)
*M*UTATE $M \times MUR$ individuals, changing NR genes
Replace $50 \times IR$ individuals with random generated individuals
Calculate the new individuals' fitness, return and risk
end while
end

with the same parameters p_f and p_g. The first column contains the assets apearing either in the crisp or the fuzzy solution. The other four columns contain the crisp and the fuzzy solution in both cases. The first one has a medium satisfaction level ($\lambda = 0.45$) whereas the second one has a very low one ($\lambda = 0.14$). This means that, in the first case, the fuzzy solution is an alternative solution that the investor should take into account, whereas, in the second one, the crisp solution cannot be substantially improved. The last two rows of the table contain the crisp and fuzzy returns and risks for the two cases. These risks measure also the efficiency of our metaheuristic: the optimal crisp solution should have a risk equal to the given one, and the near-optimal crisp solutions given by our metaheuristic have a risk level differing by a 0.54% and 0.19%, respectively, from the optimal one. The time used for the algorithm to obtain the solution to this problem is 1.61 seconds on average.

5 Conclusions

In this paper we have presented the problem of portfolio selection, which is a difficult problem to solve due to economic environment uncertainty and the problem of suitably reflecting decision maker desires in the model. We present an approach where the problem of fuzzy portfolio includes the subjective criteria of the decision maker to determine the level of risk that he or she is able to support and the level of satisfaction to be assigned to a possible increase in return.

We have proposed an exact method to solve the fuzzy model of the portfolio problem with the main idea of finding partially

	$R = 0.000494$		$R = 0.000839$	
	$p_f = 0.0002,\ p_g = 0.002$		$p_f = 0.0002,\ p_g = 0.002$	
	Crisp	Fuzzy	Crisp	Fuzzy
ACE	0.316	0.200	0	0
ACR	0.011	0	0	0
ALT	0.168	0.199	0.207	0.162
DRC	0.019	0.205	0.353	0.461
IBE	0.165	0.014	0	0
MVC	0.050	0.102	0.157	0.098
NHH	0.011	0.033	0.056	0.065
POP	0.62	0	0	0
REP	0.031	0	0.001	0
UNF	0	0.077	0.0395	0.052
VAL	0.036	0	0	0
ACX	0.053	0.069	0.051	0.017
ANA	0.078	0.100	0.139	0.145
λ	0	0.454078	0	0.145928
Return	0.00202704	0.0029352	0.00353188	0.00382374
Risk	0.000491306	0.0006031840	0.00083643	0.0010098

Table 1: Obtained solutions by GAFUZ-PF

feasible solutions involving slightly greater risk than that fixed by the decision maker, and to study the possibilities that they offer in order to improve the expected return.

We have also proposed a hybrid meta-heuristic to solve the fuzzy model for problems with medium or big size where the traditional methods become useless as they would need too much time due to the combinatorial explosion in the solution space. The proposed meta-heuristic uses a hybrid scheme, which combines ideas from the Simulated Annealing technique and genetic algorithms. To test the proposed meta-heuristic we have used several test problems based on dates of IBEX35, the best-known index of Spanish Stock Markets. The results allow us to verify that with the small problems the solutions are very similar to those obtained with the exact model. Moreover, our proposal allows us to work with big problems (large time series, intra-day data, etc.) and the time taken to get the results is about 1.6 seconds, achieving the goal we wanted to get with that meta-heuristic.

Acknowledgements

This work has been partially supported by the Ministerio de Ciencia e Innovación of Spain, TIN2008-06872-C04-02, TIN2008-06872-C04-03, TIN2008-06872-C04-04.

References

[1] E. Aarts and J. Korst, "Simulated annealing and Boltzmann machines: a stichastic approach combinatorial to optimization and neural computing," John Wiley and Sons, New York, 1980.

[2] M. J. Canós, C. Ivorra, and V. Liern, *The fuzzy p-median problem: A global analysis of the solutions*, European Journal of Operational Research, **130** (2001), 430–436.

[3] M. J. Canós, C. Ivorra, and V. Liern, *Marginal analysis for the fuzzy p-median problem*, European Journal of Operational Research, **130** (2008), 430–436.

ISBN: 978-989-95079-6-8

[4] J.M. Cadenas, M.J. Canós, M.C. Garrido, C. Ivorra and V. Liern, *Soft-computing based heuristics for location on networks: The p-median problem*, Applied Soft Computing, (2008), doi:10.1016/j.asoc.2008.03.015

[5] J.M. Cadenas, M.J. Canós, J.V. Carrillo, M.C. Garrido, C. Ivorra and V. Liern *A hybrid algorithm for the fuzzy p-median problem.* IADIS International Conference Intelligent Systems and Agents 2008, Amsterdam (The Netherlands).

[6] E.F. Fama, "Foundations of Finance," Basil Blackwell, Oxford, 1977.

[7] A. Fernández-Carazo, *Un estudio holístico de la selección y planificación temporal de carteras de proyectos*, Rect@, **9** (2008), 5–44.

[8] P. Guo and H. Tanaka, *Decision analysis based on fused double exponential possibility distributions*, European Journal of Operational Research, **148** (2003), 467–479.

[9] M. Inuiguchi and J. Ramík, *Possibilistic linear programming: a brief review of fuzzy mathematical programming and a comparison with stochastic programming in portfolio selection problem*, Fuzzy Sets and Systems, **111** (2000), 3–28.

[10] S. Kirkpatrick, C.D. Gelat and M.P. Jr Vecchi, *Optimization by Simulated Annealing*, Science, **220** (1983), 671–680.

[11] H. Konno and H. Yamazaki, *Mean-absolute deviation portfolio optimization model and its applications Tokyo Stock Market*, Management Science, **37** (1991), 519–531.

[12] H. Konno and A. Wijayanayake, *Portfolio optimization problem under concave transaction costs and minimal transactions unit constraints*, Mathematical Programming, **89** (2001), 233–250.

[13] Y. J. Lai and C. L. Hwang, "Fuzzy Mathematical Programming: Theory and Applications," Springer, Berlin, 1992.

[14] T. León, V. Liern and E. Vercher, *Viability of infeasible portfolio selection problems: A fuzzy approach*, European Journal of Operational Research, **139** (2002), 178–189.

[15] T. León, V. Liern and E. Vercher, *Two fuzzy approaches for solving multiobjective decision problems*, Computational Economics, **19** (2002), 273–286.

[16] T. León, V. Liern, P. Marco, J. V. Segura and E. Vercher, *A downside risk approach for the portfolio selection problem with fuzzy returns*, Fuzzy Economic Review, **9** (2004), 61–77.

[17] H.M. Markowitz, *Portfolio selection*, Journal of Finance, **7** (1952), 79–91.

[18] H.M. Markowitz, P. Kall and S. T. Wallace, "Portfolio selection: Efficient Diversification of Investments," John Wiley, New York, 1959.

[19] M. Mitchell, "An Introduction to Genetic Algorithms," The MIT Press, London, 1996.

[20] C.R. Reeves, Genetic Algorithms, In F. Glover and G.A. Kochenberger (editors)Handbook of Metaheuristics, Int. Series in Operations Research & Management Science 57, Kluwer Academic Publishers, 62–77, 2002.

[21] W.F. Sharpe, "Portfolio theory and capital market," McGraw-Hill, New York, 1970.

[22] M.G. Speranza, *Linear programming model for portfolio optimization*, Finance, **14** (1993), 107–123.

[23] J. Watada, *Fuzzy portfolio selection and its applications to decision making*, Tatra Mountains Mathematical Publications, **13** (1997), 219–248.

[24] S.A. Zenios and P. Kang, *Mean-absolute deviation portfolio optimization for mortgage-backed securities*, Annals of Operations Research **45** (1993), 433–450.

ISBN: 978-989-95079-6-8

Information System for the valuation of Universities in Spain

M. Socorro Garcia-Cascales[1], M. Teresa Lamata[2]

[1] Dpto de Electrónica, Tecnología de Computadoras y Proyectos.
Universidad de Politécnica de Cartagena. Murcia, España
[2] *Dpto. Ciencias de la Computación e Inteligencia Artificial.*
Universidad de Granada. 18071- Granada. España
Email: socorro.garcia@upct.es, mtl@decsai.ugr.es

Abstract— *The Spanish National Agency for Quality Assessment and Accreditation as part of its evaluation activities has established a procedure for evaluating both teaching and institutions, by means of the Institutional Assessment Programme. In this communication we shall focus on the external assessment phase for qualifications in the field of Industrial Engineering and specifically on the structures of the database for a Decision Support System on the universities' rankings. In particular, this paper will focus on obtaining the weight of the criteria and definition of the linguistic labels used in the external assessment phase.*

Keywords— Analytical Hierarchy Process (AHP), Linguistic Labels, Qualification evaluation, Weighting Criteria

1 Introduction

The university rankings which just 20 years ago were an innovation are today a normal characteristic in the majority of countries with extensive systems of higher education [7, 13, 16, 17, 18]. These lists have an increasing impact not only between the universities themselves, but also between different social sectors.

University rankings are lists of certain groups of institutions, classified in comparative form, and according to a common set of indicators, in descending order.

Likewise, the reorganisation at European scale of university studies as a result of the Bologna Process, will contribute in an active way to the harmonisation of basic European academic aspects. This aspect will allow, on the one hand, for a better connection between universities and on another hand, for an easier comparison between them; for which certain indicators of functioning will be necessary.

All the above leads us to think about the need for the existence of global rankings of universities as an instrument to measure their quality rigorously [7, 16]. In this context, the Decision Support Systems (DSS), seem to be useful for the evaluation of the qualifications in the universities' ranking.

In this paper, we shall focus on the area of industrial engineering within the Spanish university system and on the structure of the Database as a fundamental element of the DSS.

The aim of the paper is the use of the fuzzy AHP process to obtain the weight of the criteria within a DSS in order to facilitate the process of evaluation on the universities' rankings.

Another aim is to obtain the fuzzy numbers associated with the linguistic labels used to evaluate the ANECA in the external assessment phase.

The paper is organised as follows: The next section introduces the ANECA process. Section 3 introduces the linguistic variables and fuzzy sets. In section 4, the framework for the AHP method is defined. Section 5 describes the database. Section 6 shows the aggregation results and finally we outline the most important conclusions.

2 The ANECA Process

In this sense and within the framework of the Institutional Assessment Programme (*Programa de Evaluación Institucional*, "PEI"), the Spanish National Agency for Quality Assessment and Accreditation (*Agencia Nacional de Evaluación de la Calidad y Acreditación "*ANECA") presents this Guide [2] with the purpose of showing the steps in the Process of Institutional Assessment

The primary objective of the Institutional Assessment Programme is to facilitate an assessment process to officially improve the quality of education leading to obtain university degrees throughout the national territory, through self-diagnosis and the external view brought by experts.

The development of this programme is intended to promote assessment processes favouring the establishment or continuity of processes guaranteeing quality in teaching, as well as providing information to the students and their families, to society, to the governing bodies of the universities and to public administrations, regarding the quality of university teaching and their action plans.

This process is organised in three phases, see figure 1:

- *Self-assessment:* the "Self-assessment Report" describes and evaluates the situation of the assessed degree with respect to the criteria established, identifying strengths and weaknesses and enhancement proposals forming the basis of execution of the action plans that must be initiated on conclusion of the entire process. The report is written by the Self-assessment Committee.

ISBN: 978-989-95079-6-8

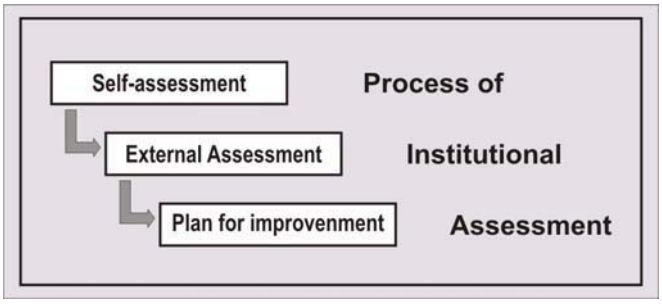


Figure 1: Phases of the Institutional Assessment Programme

- *External Assessment:* a group of external assessors to the teaching institution assessed, appointed by ANECA, and under its guidelines and supervision, analyzes the Self-Assessment Report, both through the study of documents and by means of a visit to the unit assessed, issues its recommendations, and proposes improvements. The result of this phase is the "External Assessment Report."

- *Plan for improvement*: it collects the main results of the Assessment process. This phase concludes with the plan for improvements of the degree, describing the proposed improvement actions in the Self-assessment and External Assessment phases, once their viability is analyzed. The tasks to be performed are determined according to their accomplishment, the persons responsible, resources involved, deadlines for their implementation, the indicators monitoring the actions proposed, and the benefits expected from them.

3 Linguistic variable and fuzzy sets

3.1 Linguistic variable

Most of the times, the decision-maker is not able to define the importance of the criteria or the goodness of the alternatives with respect to each criterion in a strict way. In many situations, we use measures or quantities which are not exact but approximate.

Since Zadeh [19] introduced the concept of fuzzy set and subsequently went on to extend the notion via the concept of linguistic variables, the popularity and use of fuzzy sets has been extraordinary. We are particularly interested in the role of linguistic variables as an ordinal scale and their associated terms, in this case triangular fuzzy number, as used in the multi-criteria decision making.

By a *linguistic variable,* [20,21,22], we mean a variable whose values are words or sentences in a natural or artificial language. For example *Age* is a linguistic variable if its values are linguistic rather than numerical, i.e., *young, not young, very young, quite young, old, not very old and not very young*, etc., rather than numbers as 20, 21,22, 23,.... .

Definition.1- A linguistic variable is characterized by a quintuple

$$\{X;T(X);U;G;M\}$$

in which

1. X is the name of the variable,
2. $T(X)$ is the term set of X, that is, the collection of its linguistic values
3. U is a universe of discourse,
4. G is a syntactic rule for generating the elements of $T(X)$ and
5. M is a semantic rule for associating meaning with the linguistic values of X.

In general for the decision-maker it is easier when he/she evaluates their judgments by means of linguistic terms. In those cases, the concept of fuzzy number is more adequate than that of real number.

3.2 Fuzzy sets

Then we have identified the linguistic variable with a fuzzy set [3,10,11]. The fuzzy set theory, introduced by Zadeh [19] to deal with vague, imprecise and uncertain problems has been used as a modelling tool for complex systems that can be controlled by humans but are hard to define precisely. A collection of objects (universe of discourse) X has a fuzzy set A described by a membership function f_A with values in the interval $[0,1]$.

$$f_A : X \to [0,1]$$

Thus A can be represented as $A = \{f_A(x) \mid x \in X\}$. The degree that u belongs to A is the membership function $f_A(x)$.

The basic theory of the triangular fuzzy number is described in Klir [12].

With regard to the fuzzy numbers, we will show only the mathematical operations that will be used throughout the development of the paper.

Definition 2. If T_1 and T_2 are two triangular fuzzy numbers defined by the triplets (a_1,b_1,c_1) and (a_2,b_2,c_2), respectively. Then, for this case, the necessary arithmetic operations with positive fuzzy numbers are:

a) Addition

$$T_1 \oplus T_2 = [a_1 + a_2, b_1 + b_2, c_1 + c_2] \quad (1)$$

b) Subtraction

$T_1 \Theta T_2 = T_1 + (-T_2)$ when the opposite

$-T_2 = (-c_2, -b_2, -a_2)$

$$\text{then} \quad T_1 \Theta T_2 = [a_1 - c_2, b_1 - b_2, c_1 - a_2] \quad (2)$$

c) Multiplication

$$T_1 \otimes T_2 = [a_1 \times a_2, b_1 \times b_2, c_1 \times c_2] \quad (3)$$

d) Division

$$T_1 \varnothing T_2 = [[a_1,b_1,c_1] \cdot [1/c_2, 1/b_2, 1/a_2]], \quad 0 \neq [a_2,b_2,c_2] \quad (4)$$

e) Scalar Multiplication

$$k \circ T_1 = (k \circ a_1, k \circ b_1, k \circ c_1) \quad (5)$$

ISBN: 978-989-95079-6-8

4 The Analytic Hierarchy Process Method (AHP)

The Analytic Hierarchy Process (AHP methodology [14,15] has been accepted by the international scientific community as a robust and flexible multi-criteria decision making tool for dealing with complex decision problems. AHP has been applied to numerous decision problems such as energy policy [1,7], project selection [6], measuring business performance [1], and evaluation of advanced manufacturing technology [4,5].

- Basically, AHP has three underlying concepts: Structuring the complex decision problem as a hierarchy of goal, criteria and alternatives,
- Pair-wise comparison of elements at each level of the hierarchy with respect to each criterion on the preceding level,
- and finally vertically synthesizing the judgements over the different levels of the hierarchy.

AHP attempts to estimate the impact of each one of the alternatives on the overall objective of the hierarchy. In this case, we only apply the method in order to obtain the criteria's weights.

We assume that the quantified judgements provided by the decision-maker on pairs of criteria (C_i, C_j) are represented in an *nxn* matrix as in the following:

$$C = \begin{array}{c} \\ C_1 \\ C_2 \\ . \\ . \\ C_n \end{array} \begin{array}{c} \begin{array}{cccc} C_1 & C_2 & \cdots & C_n \end{array} \\ \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \ldots & c_{2n} \\ . & . & \ldots & . \\ . & . & \ldots & . \\ c_{n1} & c_{n2} & \ldots & c_{nn} \end{bmatrix} \end{array} \quad (6)$$

The c_{12} value is supposed to be an approximation of the relative importance of C_1 to C_2, i.e., $c_{12} \approx (w_1/w_2)$. This can be generalized and the statements below can be concluded:

1. $c_{ij} \approx (w_i/w_j)$ $i,j=1, 2, ..., n$
2. $c_{ii}=1$, $i=1,2,...,n$
3. If $c_{ij}=\alpha$, $\alpha \neq 0$, then $c_{ji}=1/\alpha$, $i=1,2,...,n$
4. If C_i is more important than C_j then $c_{ij} \cong (w_i/w_j)>1$

This implies that matrix *A* should be a positive and reciprocal matrix with 1's in the main diagonal and hence the decision-maker needs only to provide value judgements in the upper triangle of the matrix. The values assigned to c_{ij} according to Saaty scale are usually in the interval of 1-9 or their reciprocals.

It can be shown that the number of judgements (*L*) needed in the upper triangle of the matrix are:

$$L = n(n-1)/2 \quad (7)$$

where *n* is the size of the matrix *C*.

The matrixes associated to the AHP approach are reciprocal, thus:

- The maximum eigenvalue (λ_{max}) is a positive real number and such that $\lambda_{max} \geq n$.
- Associated with this eigenvalue is a vector whose components are also positive. If this vector is normalized the vector of weights associated with the matrix is obtained.

Where the values are fuzzy, not crisp; instead of using lambda as an estimator to the weight, we will use the geometric normalized average, expressed by the following expression:

$$w_i = \frac{\prod_{j=1}^{n}(a_{ij}, b_{ij}, c_{ij})}{\sum_{i=1}^{m}\prod_{j=1}^{n}(a_{ij}, b_{ij}, c_{ij})} \quad (8)$$

where, (a_{ij}, b_{ij}, c_{ij}) is a fuzzy number.

5 The database

The Database has two parts. The first part consists of the summary of the different Reports of External Evaluation, published by ANECA. These reports are given by the members of the External Assessment Committee who are persons qualified by ANECA; the valuation given by the external assessors is impartial.

The second part of the database, which is the focus of this paper, corresponds with obtaining the weights of the criteria, as well as the numerical representation of the labels. See Figure 2.

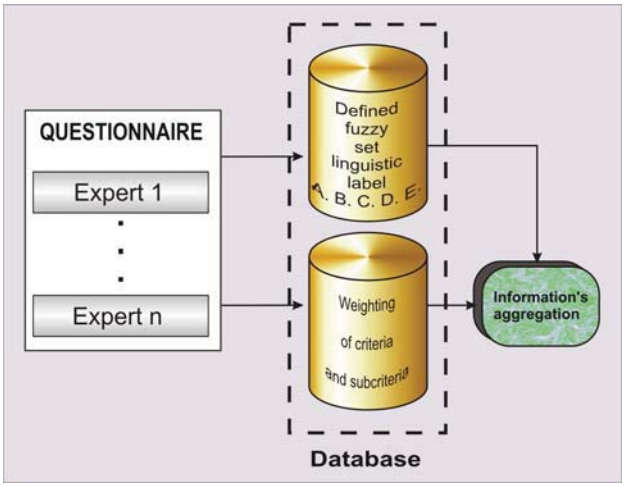


Figure 2: The database corresponding to this part of the work.

The study carried out has been based on a questionnaire designed to such effect. ANECA itself sent the questionnaire to the experts that it determined as proper assessors. The questionnaire was sent / answered by e-mail.

This questionnaire consists of two clearly differentiated parts;

- The first one has been produced on the basis of the hierarchic structure of the criteria and subcriteria (Table 1), using the AHP methodology to do so.
- The second part refers to the linguistic labels used in the ANECA survey for the external evaluation of university qualifications.

ISBN: 978-989-95079-6-8

Table 1. Weighting criteria/subcriteria

CRITERIA			General Weighting		
1st LEVEL	2nd LEVEL	3rd LEVEL	3th. LEVEL	2nd LEVEL	1st LEVEL
1. Educational programme (E.P)	1.1. Aims of the programme	Aims of the E.P.	(3.39,**5.26**,8.01)		
		Admission profile	(3.39,**5.26**,8.01)	(6.77,**10.52**,16.02)	
	1.2. Studies plan and its structure	Curricular content	(1.50,**2.32**,3.54)		
		Curricular coherence	(1.50,**2.32**,3.54)		
		Curricular consistency	(1.50,**2.32**,3.54)		
		Curricular updating	(1.50,**2.32**,3.54)		
		Aims of E.P.	(1.50,**2.32**,3.54)	(7.49,**11.61**,17.72)	(16.61,**22.13**,29.14)
2. Organization of teaching	2.1. Management and planning	Planning	(5.27,**7.94**,11.81)	(5.27,**7.94**,11.81)	
	2.2. Management and organization	Communication	(1.59,**2.39**,3.57)		
		Organisation of teaching	(1.59,**2.39**,3.57)		
		Improvement processes	(1.59,**2.39**,3.57)	(4.78,**7.16**,10.70)	(11.33,**15.10**,20.04)
3. Human Resources	3.1. Academic staff (AS)	Appropriateness AS	(5.11,**7.95**,12.15)		
		Implication AS	(5.11,**7.95**,12.15)	(10.21,**15.90**,24.29)	
	3.2. Administration and service staff (ASS)	Adaptation ASS	(2.50,**3.79**,5.94)	(2.50,**3.79**,5.94)	(14.44,**19.70**,26.70)
4. Material Resources	4.1 Classrooms	Appropriateness for numbers of students	(1.58,**2.50**,3.94)	(1.58,**2.50**,3.94)	
	4.2 Work Spaces	Appropriateness for numbers of students	(0.51,**0.81**,1.31)		
		Appropriateness (AS and ASS)	(0.51,**0.81**,1.31)		
		Infrastructures: practical	(0.51,**0.81**,1.31)	(1.53,**2.44**,3.93)	
	4.3. Laboratories, workshops and experimental spaces	Appropriateness for number of students	(1.82,**2.86**,4.53)	(1.82,**2.86**,4.53)	
	4.4. Library and document banks	Correctly furnished	(0.69,**1.09**,1.78)		
		Quality, quantity, …	(0.69,**1.09**,1.78)	(1.38,**2.19**,3.56)	(7.55,**9.99**,13.46)
5. Training Process	5.1. Student assistance and integral training	Capture	(0.54,**0.84**,1.33)		
		Student welcome actions	(0.54,**0.84**,1.33)		
		Support programmes	(0.54,**0.84**,1.33)		
		Professional orientation programmes	(0.54,**0.84**,1.33)		
		Tutorial action programmes	(0.54,**0.84**,1.33)		
		Integral training	(0.54,**0.84**,1.33)	(3.24,**5.06**,8.01)	
	5.2. Teaching-learning process	Methodology	(1.55,**2.46**,3.84)		
		Evaluation	(1.55,**2.46**,3.84)		
		External practical	(1.55,**2.46**,3.84)		
		Mobility	(1.55,**2.46**,3.84)	(6.21,**9.84**,15.34)	(11.04,**14.90**,20.10)
6. Results	6.1. Results of educational programme	Effectiveness of E.P.	(1.56,**2.73**,4.67)		
		Student satisfaction	(1.56,**2.73**,4.67)	(3.13,**5.47**,9.35)	
	6.2. Graduate results	Compliance with the graduate profiles	(3.39,**5.99**,10.34)	(3.39,**5.99**,10.34)	
	6.3. Academic staff results	Academic staff satisfaction	(1.82,**3.21**,5.70)	(1.82,**3.21**,5.70)	
	6.3. Results in society	Employers and other groups	(0.10,**1.76**,3.08)		
		Social link	(0.10,**1.76**,3.08)	(0.20,**3.52**,6.17)	(8.54,**18.18**,24.75)
Source: own production			**100.00**	**100.00**	**100.00**

Likewise, we will take into account the possible valuations of the experts, expressed in linguistic terms (A, B, C, D) These labels are arranged from largest to smallest as follows: A > B > C > D and with the semantics that we will see later.

It should be remembered that the results obtained correspond to a problem of group decision-making, formed by *n* experts, from whom we will obtain:

1.-The criteria weight as a result of the consensus

2.- The membership functions of the fuzzy numbers that represent the linguistic labels A,B,C and D.

5.1 Obtaining the weightings of the criteria/subcriteria

According to Table 1, the hierarchy structure has three levels. Considering expression (7), the large number of questions might lead to the survey not being answered, since L = 74, as we have 6 in 1st level criteria, subcriteria 16 in 2nd level and 37 in 3rd level indicators. This results in a questionnaire which is not feasible for experts to answer.

For this, and since the influence of the third level on the weight of the previous ones is practically null, a uniform distribution on them was supposed.

ISBN: 978-989-95079-6-8

For the first and second level, the procedure is as follows: firstly, we ask if all the criteria/subcriteria have the same weight, if so, we pass to the following level.

On the contrary, if they do not have the same weight, it will be continued by the following question of the questionnaire.

In this part, there is a question that asking about the order of importance of the criteria/subcritera; and finally, using the linguistic labels defined by Saaty [14], the criteria/subcriteria are compared at the same level, taking into account the order established before.

We chose to ask only for one row of the pairwise matrix and from here to generate the rest of the information matrix, which was carried out completely consistently. Thus, this part of the questionnaire had only 21 questions.

5.2 Obtaining the linguistic evaluations

The second part of the questionnaire is based on the semiquantitative survey of the Evaluation of the Education inside the protocol to elaborate the Report of External Evaluation developed by ANECA, in which the following labels are used: A: Excellent, B: Good, C: Average, D: Deficient. For the valuation of these labels we use the interval [0, 10].

These linguistic variables, by the uncertainty of their nature, justified the use of fuzzy numbers associated with each linguistic term.

6 The results of the aggregation for the group decision experts

By means of a primary group decision-making process, it is possible to see the results of the weighting of the criteria and sub-criteria for all the experts (Table 1).

Table 2. Fuzzy numbers associated with the labels A,B,C and D.

Semi quantitative labels	General
A: Excellent	(8.1354, 9.4054, 10.0000)
B: Good	(5.8108, 7.1081, 8.4054)
C: Average	(3.5090, 4.8108, 6.1126)
D: Deficient	(0.7355, 2.5135, 4.2916)

Now, taking these weightings as the base; and the definition of semiquantitative labels (Table 2), obtained by ANECA's own experts, it is possible for us, attending to these six criteria, to order the different Universities. We make reference, in this case, only to the area of industrial engineering.

For the evaluation of the alternatives the methodology used has been the fuzzy weighted sum model as:

$$FWSM = \sum_{j=1}^{n} \tilde{w}_j \cdot \tilde{a}_{ij} \qquad (9)$$

where, $\tilde{w}_j$ and $\tilde{a}_{ij}$ are fuzzy numbers.

The defuzzification method used is described in [8] when:

$$I_{\beta,\lambda}(A_i) = \beta S_M(A_i) + (1-\beta)\lambda S_R(A_i) + (1-\beta)(1-\lambda) S_L(A_i)$$

In this way, we have defined a fuzzy number as a function of the three integrals, $S_L(A_i)$, $S_M(A_i)$ and $S_R(A_i)$, where $S_R(A_i)$ represents the upper mean value associated with the inverse function of $f_A^R(x)$, $S_L(A_i)$ is the lower mean value of the $g_A^L(x)$ function and $S_M(A_i)$ is the area of the core of the fuzzy number, $\beta \in [0,1]$ is the index of modality that represents the importance of the central value against the extreme values, and $\lambda \in [0,1]$ is the degree of optimism of the decision maker.

We have considered the case in which the three areas have the same weight and it would correspond to the neutral decision maker, when $\alpha = 1/2$ y $\beta = 1/3$

Taking as an example (Table 3), the outcome results of the DSS, for the real case of the evaluation of five universities in the qualifications of industrial engineering, where we have obtained results both for the principal criteria and for the global evaluation.

Table 3: Ranking result for five Universities.

Criteria	C1 Educational Programme	C2 Teaching Organizatio	C3 Human Resources	C4 Material Resources	C5 Educational Process	C6 Results	IAP Gobal
U1	0.062	0.033	0.071	0.040	0.042	0.056	0.306
U2	0.100	0.067	0.106	0.049	0.055	0.089	0.466
U3	0.129	0.086	0.105	0.067	0.098	0.119	0.604
U4	0.118	0.080	0.088	0.038	0.075	0.081	0.482
U5	0.125	0.089	0.105	0.047	0.082	0.100	0.548

7 Conclusions

For the assessors/experts it is simpler to express their knowledge by means of linguistic labels, instead of having to do so by means of numerical values. For that reason, it is preferable to prepare a questionnaire to obtain the experts' knowledge, in which the answers are in the form of linguistic variables. These linguistic variables have been modelized by means of fuzzy triangular numbers and from a methodology widely accepted by the scientific community, since it is the Analytical Hierarchic Process; developed by Saaty in 1980.

Taking Table 1into account, we conclude that in general the most important criterion is the Educational Programme and that the lowest weighting is obtained for the Material Resources.

For future work, it would be interesting to carry out a study of the aggregation of the information as a secondary process. In this process the weight of the criteria and subcriteria are obtained for every expert. Later all this information would be aggregated.

Moreover, a comparative primary and secondary study of both types of processes would be desirable.

ISBN: 978-989-95079-6-8

Similarly, as future work, it would be interesting to make a comparative study of the evaluation criteria according to the area of the different experts. It is possible that the valuation of an expert from the area of engineering does not value the criteria in the same way as another from the humanities.

Acknowledgment

This work is partially supported by the DGICYT under project TIN2008-06872-C04-04 and Junta de Andalucía (P07-TIC02970), respectively.

References

[1] K. M. A. Al Harbi. Application of the AHP in project management. International Journal of Project Management;,19(1), 19-27, 2001 .

[2] Agencia Nacional de Evaluación de la Calidad y Acreditación (ANECA) (2007). Programa de Evaluación Institucional. Anualidad 2006/2007. Guía de evaluación externa 2007. http://www.aneca.es/active/active_ense_pei0607.asp

[3] R.E. Bellman, L.A. Zadeh. Decision-making in a fuzzy environment. Management Sci., 17, 141-164.,1970.

[4] Chan F.T.S, Chan M.H, Tang N.K.H. Evaluation methodologies for technology selection. Journal of Materials Processing Technology 2000a ; 107: 330–337.

[5] F.T.S Chan, B.Jiang, N.K.H. Tang. The development of intelligent decision support tools to aid the design flexible manufacturing systems. International Journal of Production Economics, 65 (1): 73–84, 2000.

[6] C.-H. Cheng, K.-L. Yang, C.-L. Hwang. Evaluating attack helicopters by AHP based on linguistic variable weight. European Journal of Operation al Research 116 (2),423–435, 1999.

[7] G. Federkeil.. Rankings and Quality Assurance in Higher Education. Higher Education in Europe, 33, 2/3, 219-230, 2008.

[8] M.S, Garcia-Cascales,., M.T. Lamata,: A modification to the index of Liou and Wang for ranking fuzzy number, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 411-424, 2007.

[9] M. M. Kablan. Decision support for energy conservation promotion: an analytic hierarchy process approach. Energy Policy, 32 (10): 1151-1158, 2004.

[10] J. Kacprzyk, R.R. Yager. Linguistic summaries of data using fuzzy logic. International Journal of General Systems. 30, 2, 133-154, 2001.

[11] E.E. Kerre. The use of fuzzy set theory in electrocardiological diagnostics, in: M.M. Gupta and E. Sanchez, Eds., Approximate Reasoning in Decision Analysis Amsterdam; North-Holland 277-282, 1982.

[12] G.J. Klir and B. Yuan, Fuzzy sets and Fuzzy Logic, Prentice Hall PTR, New Jersey e d. 1995

[13] P. Montesinos, J. M.Carot, J. M. Martinez, F. Mora. Third Mission Ranking for World Class Universities: Beyond Teaching and Research. Higher Education in Europe, 33, 2 & 3, 259 – 27, 2008.

[14] T.L. Saaty. The Analytic Hierarchy Process. Macgraw Hill. 1980.

[15] T.L. Saaty. Group Decision Making and the AHP. Springer Verlag. New York. 1989.

[16] J. Sadlak J. Merisotis, N.C. Liu. University Rankings: Seeking Prestige, Raising Visibility and Embedding Quality. Higher Education in Europe, 33, 2 & 3, 195 – 199, 2008.

[17] M.E. Smith. Change and innovation in higher education: a role for corporate strategy?. Higher Education 16:37-52, 1987.

[18] E. de Weert. A macro-analysis of quality assessment in higher education. Higher Education 19:57-72, 1990.

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

[20] L. A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning: Part 1. Information Sciences, 8, 199-249, 1975.

[21] L. A. Zadeh, J. Kacprzyk (eds). Computing with Words in Information / Intelligent Systems 1. Foundations. Studies in Fuzziness and Soft Computing, Physica-Verlag (Springer-Verlag), Heidelberg and New York. Vol. 34 1999.

[22] L. A. Zadeh, J. Kacprzyk (eds). Computing with Words in Information / Intelligent Systems 2. Applications. Studies in Fuzziness and Soft Computing. Physica-Verlag (Springer-Verlag), Heidelberg and New York. Vol. 33 1999.

ISBN: 978-989-95079-6-8

Multiple-Criteria Fuzzy Evaluation: The FuzzME Software Package

Jana Talašová[1] Pavel Holeček[2]

1., 2. Faculty of Science, Palacký University Olomouc
tř. Svobody 26, 771 46 Olomouc, Czech Republic
Email: talasova@inf.upol.cz, holecekp@inf.upol.cz

Abstract— *This paper introduces a new software product FuzzME. It was developed as a tool for creating fuzzy models of multiple-criteria evaluation and decision making. The type of evaluations employed in the fuzzy models fully corresponds with the paradigm of the fuzzy set theory; the evaluations express the (fuzzy) degrees of fulfillment of corresponding goals. The FuzzME software takes advantage of linguistic fuzzy modeling to the maximum extent. In the FuzzME software, both quantitative and qualitative criteria can be used. The basic structure of evaluation is described by a goals tree. Within the goals tree, aggregation of partial fuzzy evaluations is done either by one of fuzzified aggregation operators (fuzzy weighted average, ordered fuzzy weighted average) or by a fuzzy expert system.*

The paper contains an illustrative example of the software usage. The application concerns a soft-fact-rating problem that was solved in one of Austrian banks.

Keywords— Fuzzy expert system, Fuzzy OWA operator, Fuzzy weighted average, Multiple-criteria fuzzy evaluation, Normalized fuzzy weights, Software.

1 Introduction

In practice, sophisticated models of multiple-criteria evaluation are required (e.g. rating of clients in banks, evaluation of hospitals or universities, comparison of alternative solutions to ecological problems). For creating the evaluating models, setting some of their input data and interpreting their outputs, expert's knowledge is needed (e.g. evaluations of alternatives according to qualitative criteria, partial evaluating functions for quantitative criteria, a choice of a suitable type of aggregation, criteria weights, or eventually, a rule base describing the relation between criteria values, the overall evaluation and a linguistic description of obtained results). Because uncertainty is the typical feature of any expert information, the fuzzy set theory is a suitable mathematical tool for creating such models. For the practical use of the fuzzy models of multiple-criteria evaluation, their user-friendly software implementation is necessary. But a good theoretical basis of the used models is crucial, too. The clear and well-elaborated theory of multiple-criteria fuzzy evaluation makes it possible to create an understandable methodics for the software user. And a good methodics is essential for correct application of any software to solving real problems.

There is a large number of papers and books dealing with the theory and methods of multiple-criteria evaluation that make use of the fuzzy approach (e.g. [1], [2], [3], [4]). Multiple-criteria evaluation (as a basis of decision making) was even one of the earliest applications of fuzzy sets (see [1]). In more than 40 years of existence of the fuzzy sets theory, several software products for multiple-criteria decision making, which use the fuzzy modeling principles in different degrees and in different ways, have been developed. In practice, FuzzyTECH (see [5]) is probably the best-known of these. It enables to use the specific instruments of the fuzzy set theory for solving multiple-criteria evaluation and decision making problems. Generally, FuzzyTECH is a universal software product which makes it possible to create and use fuzzy expert systems (or fuzzy controllers). It also includes the possibility to derive fuzzy rule bases from given data by means of neural network algorithms. In the book [6], there were described many successful applications of this software to solving multiple-criteria evaluation and decision making problems in the areas of business and finance. Similarly, fuzzy toolboxes of general mathematical systems such as Matlab can be used for multiple-criteria decision making.

The FuzzME software (**Fuzz**y models of **M**ultiple-criteria **E**valuation), presented in this paper, is based on a theoretical concept of evaluation which is very close to the original Zadeh's ideas. Similarly to his paper [1], the evaluations of alternatives according to particular criteria represent their degrees of fulfillment of the corresponding partial goals. Besides evaluations expressed by real numbers in $[0, 1]$, fuzzy evaluations modeled by fuzzy numbers on the same interval are employed in the software. They represent, analogously, the fuzzy degrees of fulfillment of the partial goals which are connected to the criteria. Resulting fuzzy evaluations, which are obtained by aggregation, have a similar clear interpretation. This theoretical approach to (fuzzy) evaluation was published in the book [7] and in the paper [8].

The predecessor of the FuzzME software package in terms of the used theoretical basis was the NEFRIT software. This software for multiple-criteria evaluation and decision making, which is also based on fuzzy technologies, was developed in about 2000 by the Czech software company TESCO SW Inc. The fuzzy model of evaluation applied there is described in detail in [8] and in the book [7] (a demo version of NEFRIT is enclosed in the book). NEFRIT makes it possible to work with expert fuzzy evaluations of alternatives according to qualitative criteria. Values of the quantitative criteria can either be crisp or fuzzy. Evaluating functions for quantitative criteria represent membership functions of corresponding partial goals. The main evaluation structure is expressed by a goals tree. For aggregation of the partial fuzzy evaluations the weighted average method is used. The weights (crisp only, not fuzzy) express the shares of partial evaluations in the

ISBN: 978-989-95079-6-8

aggregated one. Fuzzy evaluations on all levels of aggregation express the fulfillment of the corresponding goals. The NEFRIT software was originally developed for the Czech National Bank (decision making about granting a credit). Further, it was used e.g. by the Czech Tennis Association, the Czech Basketball Association and in other institutions. Nowadays it is tested by the Supreme Audit Office of the Czech Republic.

In contrast to NEFRIT, the FuzzME software makes it possible to use also uncertain weights in the aggregation by means of the weighted average method. The theory of normalized fuzzy weights, procedures for their setting, and an effective algorithm for calculation of fuzzy weighted average are taken from [9], [10] and [11]. Another fuzzy aggregation operator, available in the FuzzME software, is a fuzzified OWA operator. Again, it works with normalized fuzzy weights. The fuzzy OWA operator and the used algorithm for its calculation are described in [12]. In the FuzzME software, multiple-criteria evaluating functions can also be defined by means of fuzzy rule bases. In accordance with [7], two algorithms are offered for the approximate reasoning - the standard Mamdani algorithm and a modified Sugeno algorithm.

There are also software tools for multiple-criteria decision making based on other mathematical methods. But they are usually designed for solving a particular assignment. Our investigation by means of Internet did not result software fully comparable to FuzzME. Its universality and comprehensiveness make it unique.

2 Preliminaries

A fuzzy set A on a universal set X is characterized by its membership function $A : X \to [0,1]$. $Ker\ A$ denotes a kernel of A, $Ker\ A = \{x \in X \mid A(x) = 1\}$. For any $\alpha \in [0,1]$, A_α denotes an α-cut of A, $A_\alpha = \{x \in X \mid A(x) \geq \alpha\}$. A support of A is defined as $Supp\ A = \{x \in X \mid A(x) > 0\}$. The symbol $hgt\ \ A$ denotes a height of the fuzzy set A, $hgt\ A = \sup\{A(x) \mid x \in X\}$. An intersection and a union of the fuzzy sets A and B on X are defined for all $x \in X$ by the following formulas: $(A \cap B)(x) = \min\{A(x), B(x)\}$, $(A \cup B)(x) = \max\{A(x), B(x)\}$.

A fuzzy number is a fuzzy set C on the set of all real numbers $\Re$ which satisfies the following conditions: a) the kernel of C, $Ker\ C$, is not empty, b) the α-cuts of C, C_α, are closed intervals for all $\alpha \in (0,1]$, c) the support of C, $Supp\ C$, is bounded. A fuzzy number C is called to be defined on $[a,b]$, if $Supp\ C \subseteq [a,b]$. Real numbers $c^1 \leq c^2 \leq c^3 \leq c^4$ are called significant values of the fuzzy number C if the following holds: $[c^1, c^4] = Cl(Supp\ C), [c^2, c^3] = Ker\ C$, where $Cl(Supp\ C)$ denotes a closure of $Supp\ C$.

Any fuzzy number C can be characterized by a pair of functions $\underline{c} : [0,1] \to \Re$, $\overline{c} : [0,1] \to \Re$ which are defined by the following formulas: $C_\alpha = [\underline{c}(\alpha), \overline{c}(\alpha)]$ for all $\alpha \in (0,1]$, and $Cl(Supp\ C) = [\underline{c}(0), \overline{c}(0)]$. The fuzzy number C is called to be linear if both the functions $\underline{c}$, $\overline{c}$ are linear. A linear fuzzy number is fully determined by its significant values because $\underline{c}(\alpha) = (c_2 - c_1) \cdot \alpha + c_1$, $\overline{c}(\alpha) = (c_3 - c_4) \cdot \alpha + c_4$. For that reason, we can denote it as $C = (c^1, c^2, c^3, c^4)$.

An ordering of fuzzy numbers is defined as follows: a fuzzy number C is greater than or equal to a fuzzy number D, if $C_\alpha \geq D_\alpha$ for all $\alpha \in (0,1]$.

A fuzzy scale makes it possible to represent a closed interval of real numbers by a finite set of fuzzy numbers. Let $T_1, T_2, ..., T_s$ be fuzzy numbers defined on $[a,b]$, forming a fuzzy partition on the interval, i.e., for all $x \in [a,b]$ the following holds

$$\sum_{i=1}^{s} T_i(x) = 1, \tag{1}$$

then the set of the fuzzy numbers can be linearly ordered (see [7]). If the fuzzy numbers $T_1, T_2, ..., T_s$ are defined on $[a,b]$, form a fuzzy partition on the interval and are numbered according to their linear ordering, then they are said to form a fuzzy scale on $[a,b]$.

An uncertain division of the whole into m parts can be modeled by normalized fuzzy weights. Fuzzy numbers $V_1, ..., V_m$ defined on $[0,1]$ are normalized fuzzy weights if for any $i \in \{1, ..., m\}$ and any $\alpha \in (0,1]$ it holds that for any $v_i \in V_{i\alpha}$ there exist $v_j \in V_{j\alpha},\ j = 1, ..., m,\ j \neq i$, such that

$$v_i + \sum_{j=1, j\neq i}^{m} v_j = 1. \tag{2}$$

3 The FuzzME software

The mathematical models of the FuzzME software are based primarily on the theory and methods of multiple-criteria evaluation that were published in [7] and [8]. The theory of normalized fuzzy weights, the definition of fuzzy weighted average, and the algorithm for its computation were taken from [9], [10] and [11]. The fuzzified OWA operator and the algorithm for its calculation published in [12] are also used in the software.

In the FuzzME software, the basic structure of the fuzzy model of multiple-criteria evaluation is expressed by a goals tree. The root of the tree represents the overall goal of evaluation and each branch corresponds to a partial goal. The goals at the ends of branches are connected either with quantitative or qualitative criteria.

When an alternative is evaluated, evaluations with respect to criteria connected with the terminal branches are calculated first. Independently of the criterion type, each of the evaluations is described by a fuzzy number defined on the interval $[0,1]$. It expresses the fuzzy degree of fulfillment of the corresponding partial goal.

These partial fuzzy evaluations are then aggregated according to the defined type of the tree node. Three types of aggregation are available: a fuzzy weighted average (fuzzy WA), an ordered fuzzy weighted average (fuzzy OWA) or aggregation by means of a fuzzy expert system. For aggregation by fuzzy weighted average or ordered fuzzy weighted average, normalized fuzzy weights must be set. The weights express uncertain shares of the partial evaluations in the aggregated one. For the fuzzy expert system, the fuzzy rule base must be defined and an inference algorithm must be chosen (either the Mamdani algorithm or the generalized Sugeno algorithm of approximate reasoning).

The overall evaluation reflects the degree of fulfillment of the main goal. A verbal description of the overall evaluation can be obtained by means of the implemented linguistic approximation algorithm.

ISBN: 978-989-95079-6-8

The overall evaluations can be compared within the frame of a given set of alternatives. By this comparison the best of the alternatives can be chosen. That is why the FuzzME software can be also used as a decision support system.

The import and export of data is supported by the software, too. The FuzzME software is available in the Czech and English versions.

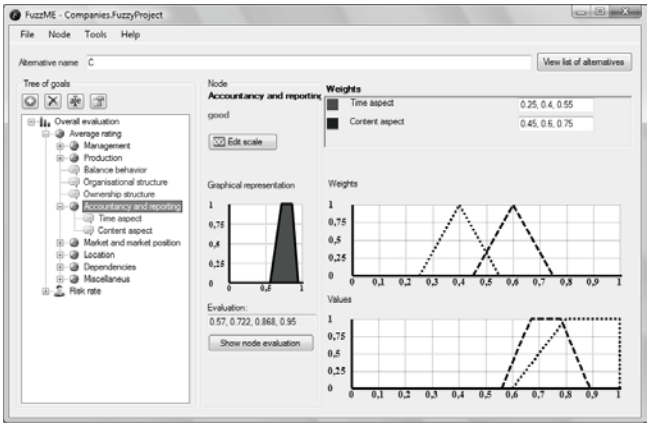


Figure 1: The main window of the software

3.1 Goals tree

Goals trees represent the basic structure of fuzzy models of multiple-criteria evaluation in the FuzzME software. When a goals tree is designed, the main goal is consecutively divided into goals of progressively lower levels. The process of division is stopped when such goals are reached whose fulfillment can be assessed by means of some known characteristics of alternatives (i.e. quantitative or qualitative criteria).

The design of a tree structure in the goals-tree editor is the first step in forming a fuzzy evaluation model in FuzzME. In the next step, the type of each node in the tree must be specified. For the nodes at the ends of tree branches the user defines if the node is connected with a quantitative or qualitative criterion. For the other nodes he/she sets the type of aggregation - fuzzy weighted average, ordered fuzzy weighted average or fuzzy expert system. An example of a goals tree is illustrated in Fig. 2.

3.2 Criteria of evaluation

In the models of evaluation created by the FuzzME software, qualitative and quantitative criteria can be combined arbitrarily.

3.2.1 Qualitative criteria

According to qualitative criteria, alternatives are evaluated verbally, by means of values of linguistic variables of special kinds - linguistic scales, extended linguistic scales and linguistic scales with intermediate values.

A linguistic variable is defined as a quintuple $(\mathcal{V}, \mathcal{T}(\mathcal{V}), X, G, M)$, where $\mathcal{V}$ is a name of the variable, $\mathcal{T}(\mathcal{V})$ is a set of its linguistic values, X is a universal set on which the meanings of the linguistic values are defined, G is a syntactic rule for generating values in $\mathcal{T}(\mathcal{V})$, and M is a semantic rule which maps each linguistic value $\mathcal{C} \in \mathcal{T}(\mathcal{V})$ to its mathematical meaning, $C = M(\mathcal{C})$, which is a fuzzy set on X.

A linguistic scale on $[a, b]$ is a special case of the linguistic variable $(\mathcal{V}, \mathcal{T}(\mathcal{V}), X, G, M)$, where $X = [a, b]$, $\mathcal{T}(\mathcal{V}) = \{\mathcal{T}_1, \mathcal{T}_2, ..., \mathcal{T}_s\}$ and the meanings of the linguistic values $\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_s$ are modeled by fuzzy numbers $T_1, T_2, \ldots, T_s$ which form a fuzzy scale on $[a, b]$. As the set of linguistic values of the scale is defined explicitly, it is not necessary to include the grammar G into the scale notation.

In the FuzzME software, the user defines a linguistic scale for each qualitative criterion in the fuzzy-scale editor. For example, the linguistic scale *quality of a product* can contain linguistic values *poor*, *substandard*, *standard*, *above standard* and *excellent*. The evaluating linguistic scale is usually defined on $[0, 1]$; in other cases, it has to be transformed to this interval.

The extended linguistic scale contains, besides elementary terms of the original scale, $\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_s$, also derived terms in the form $\mathcal{T}_i$ *to* $\mathcal{T}_j$, where $i < j, i, j \in \{1, 2, \ldots, s\}$. For example, the user can evaluate *quality of a product* by the linguistic term *standard to excellent*. The meaning of the linguistic value $\mathcal{T}_i$ *to* $\mathcal{T}_j$ is modeled by $T_i \cup_L T_{i+1} \cup_L \cdots \cup_L T_j$, where $\cup_L$ denotes the union of fuzzy sets based on the Lukasiewicz disjunction; e.g. $(T_i \cup_L T_{i+1})(x) = \min\{1, T_i(x) + T_{i+1}(x)\}$ for all $x \in \Re$.

The linguistic scale with intermediate values is the original linguistic scale enriched with derived terms *between* $\mathcal{T}_i$ *and* $\mathcal{T}_{i+1}$, $i \in \{1, 2, \ldots, s-1\}$. The meaning of the derived term *between* $\mathcal{T}_i$ *and* $\mathcal{T}_{i+1}$ is modeled by the arithmetic average of the fuzzy numbers T_i and T_{i+1}.

In the FuzzME software, the user evaluates a given alternative according to a qualitative criterion by selecting a proper linguistic evaluation from a drop-down list box. He/she can choose the value from a standard linguistic scale, extended scale or scale with intermediate values.

The three mentioned structures of linguistic values are also applied when resulting fuzzy evaluations are approximated linguistically.

3.2.2 Quantitative criteria

The evaluation of an alternative with respect to a quantitative criterion is calculated from the measured value of the criterion by means of the evaluating function expertly defined for the criterion. The evaluating function is the membership function of the corresponding partial goal. The FuzzME software admits both crisp and fuzzy values of quantitative criteria. The fuzzy values represent inaccurate measurements or expert estimations of the criteria values. In the case of a fuzzy value, the corresponding partial fuzzy evaluation is calculated by the extension principle.

In the FuzzME software, the evaluating function of a quantitative criterion is formally set by means of a fuzzy number. For example, if the evaluating function is defined by a linear fuzzy number $F = (f_1, f_2, f_3, f_4)$, then f_1 is the lower limit of all at least partly acceptable values of the criterion, f_2 is the lower limit of its fully satisfactory values, f_3 is the upper limit of the fully satisfactory values, and f_4 is the upper limit of the acceptable values.

For example, when a bank evaluates expected profitability of projects, the evaluating function can be defined by a linear fuzzy number with significant values 10, 30, 500, 500. In that case, values lower than 10% are not satisfying at all (the client

ISBN: 978-989-95079-6-8

would not be able to pay the money back to the bank). For the values from 10% to 30% the satisfaction of the bank is growing linearly. Values greater than 30% are fully satisfactory from the bank's point of view. Values greater than 500% are not supposed to occur. This way we can define a monotonous evaluating function, which is the most common in the evaluating models, by a fuzzy number.

3.3 Methods of aggregation of partial evaluations

The calculated partial fuzzy evaluations are then consecutively aggregated according to the structure of the goals tree. With respect to the defined type of the tree node, the fuzzy weighted average method, the ordered fuzzy weighted average method or the fuzzy expert system method is used for the aggregation. Each of the aggregation methods is suitable for a different situation:

The fuzzy weighted average is used if the goal corresponding with the node of interest is fully decomposed into disjunctive goals of the lower level. The normalized fuzzy weights represent uncertain shares of these lower-level goals in the goal corresponding with the considered node.

Again, the ordered fuzzy weighted average requires that the goal corresponding with the given node is decomposed into disjunctive goals of the lower level. In contrast to the fuzzy weighted average, the usage of this aggregation operator supposes special user's requirements concerning the structure of partial fuzzy evaluations. The normalized fuzzy weights again represent uncertain shares of the partial evaluations in the aggregated one. But the normalized fuzzy weights are not linked to the individual partial goals; the correspondence between the weights and the partial evaluations is given by the ordering of partial evaluations of the alternative of interest. It means, evaluations with respect to the same partial goal can have different weights for different alternatives.

If the relationship between the evaluations of the lower level and the evaluation corresponding with the given node is more complex (if neither of the two previous methods can be used), and if expert knowledge about the relationship is available, then the aggregation function is described by a fuzzy rule base of a fuzzy expert system. The approximate reasoning is used to calculate the resulting evaluation. In particular, evaluating function described by a fuzzy expert system is used if the fulfillment of a goal at the end of a tree branch depends on several mutually dependent criteria (i.e., if combinations of criteria values bring synergic or disynergic effects to the resulting multiple-criteria evaluation).

3.3.1 Fuzzy weighted average

If the fuzzy weighted average is used for aggregation of partial fuzzy evaluations, then the uncertain weights of the corresponding partial goals, which express their shares in the superior goal, must be set. To define consistent uncertain weights, a special structure of fuzzy numbers, normalized fuzzy weights, must be used.

In the FuzzME software, both real and fuzzy normalized weights can be used. Normalized real weights, i.e., real numbers $v_1, ..., v_m$, $v_j \geq 0$, $j = 1, ..., m$, $\sum_{j=1}^{m} v_j = 1$, represent a special case of the normalized fuzzy weights.

The fuzzy weighted average of the partial fuzzy evaluations, i.e., of fuzzy numbers $U_1, ..., U_m$ defined on $[0, 1]$, with the normalized fuzzy weights $V_1, \ldots, V_m$, is a fuzzy number U on $[0, 1]$ whose membership function is defined for any $u \in [0, 1]$ as follows

$$U(u) = \max\{\min\{V_1(v_1), ..., V_m(v_m), U_1(u_1), ..., U_m(u_m)\} \mid \sum_{i=1}^{m} v_i u_i = u,\ \sum_{i=1}^{m} v_i = 1,\ v_i, u_i \in [0, 1], i = 1, ..., m\}. \quad (3)$$

For an expert who sets the fuzzy weights, it is not so easy to satisfy the condition of normality. That is why the FuzzME software allows to set only an approximation to the normalized fuzzy weights - fuzzy numbers $W_1, ..., W_m$ on $[0, 1]$ satisfying the following weaker condition

$$\exists w_i \in Ker\ W_i,\ i = 1, ..., n : \sum_{i=1}^{n} w_i = 1. \quad (4)$$

The software removes the potential inconsistence in $W_1, ...,$ W_m and derives the normalized fuzzy weights $V_1, ..., V_m$ from them.

The structure of normalized fuzzy weights and the fuzzy weighted average operation are studied in detail in [9], [10] and [11]. Conditions for verifying normality of fuzzy weights, an algorithm for normalization of fuzzy weights satisfying the condition (4), and an algorithm for calculating fuzzy weighted average, which are all used in the FuzzME software, can be found there. Let us notice, that the used algorithm of fuzzy weighted average calculation is very effective.

3.3.2 Ordered fuzzy weighted average

The fuzzy OWA operator is used in case that the evaluator has special requirements concerning the structure of the partial evaluation. For example, he/she does not want any partial goal to be satisfied poorly. Then the weight of the minimum partial evaluation of any alternative equals 1, and the weights of all its other partial evaluations equal 0. The aggregated fuzzy evaluations then represent the guaranteed fuzzy degrees of fulfillment of all the partial goals (the fuzzy MINIMAX method). Another example of the fuzzy OWA operator usage could be the evaluation of subjects who can choose in which of the three areas they will be mostly involved. The evaluation algorithm should take into account their right of choice. Then, e.g., the results in the area where the subject performs best contribute to the overall evaluation by about one half, results from the second area by one third and results from the area in which the subject was least involved contribute to the overall evaluation only by one sixth. A practical application of such a fuzzy evaluation model could be the overall evaluation of the academic staff with respect to their results in the areas of research, education, and management of education and science.

The ordered fuzzy weighted average represents a fuzzification of the crisp OWA operator by means of the extension principle. Uncertain weights are modeled by normalized fuzzy weights as in the case of fuzzy weighted average.

The following notation will be used to define the ordered fuzzy weighted average: if $(x_1, ..., x_m)$ is a vector of real numbers, then $(x^{(1)}, ..., x^{(m)})$ is a vector in which for all $j \in \{1, \ldots, m\}$, $x^{(j)}$ is the j-th greatest number of $x_1, ..., x_m$.

The ordered fuzzy weighted average of the partial fuzzy evaluations, i.e., of fuzzy numbers $U_1, ..., U_m$ defined on

ISBN: 978-989-95079-6-8

$[0,1]$, with the normalized fuzzy weights $V_1, \ldots, V_m$, is a fuzzy number U on $[0,1]$ whose membership function is defined for any $u \in [0,1]$ as follows

$$U(u) = \max\{\min\{V_1(v_1), ..., V_m(v_m), U_1(u_1), ..., U_m(u_m)\} \mid \sum_{i=1}^{m} v_i u^{(i)} = u,\ \sum_{i=1}^{m} v_i = 1,\ v_i, u_i \in [0,1], i = 1, ..., m\}. \quad (5)$$

The algorithm used to calculate the ordered fuzzy weighted average in the FuzzME software was taken from [12], where fuzzification of the OWA operator is described in detail. The used algorithm is an analogy to the one used for the fuzzy weighted average.

3.3.3 Fuzzy expert system

The fuzzy expert system is used if the relationship between the criteria (or the partial evaluations) and the overall evaluation is complicated. Theoretically, it is possible to model, with an arbitrary precision, any Borel measurable function by means of a fuzzy rule base (properties of Mamdani and Sugeno fuzzy controllers, see e.g. [13]) In reality, the quality of the approximation is limited by the expert's knowledge of the relationship.

If the fuzzy rule base models the relation between values of criteria and the fulfillment of the corresponding partial goal, then the evaluation function is of the following form

$$\begin{aligned} &\text{If } \mathcal{C}_1 \text{ is } \mathcal{A}_{1,1} \text{ and } \ldots \text{ and } \mathcal{C}_m \text{ is } \mathcal{A}_{1,m}, \text{then } \mathcal{E} \text{ is } \mathcal{U}_1 \\ &\text{If } \mathcal{C}_1 \text{ is } \mathcal{A}_{2,1} \text{ and } \ldots \text{ and } \mathcal{C}_m \text{ is } \mathcal{A}_{2,m}, \text{then } \mathcal{E} \text{ is } \mathcal{U}_2 \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &\text{If } \mathcal{C}_1 \text{ is } \mathcal{A}_{n,1} \text{ and } \ldots \text{ and } \mathcal{C}_m \text{ is } \mathcal{A}_{n,m}, \text{then } \mathcal{E} \text{ is } \mathcal{U}_n \end{aligned} \quad (6)$$

where for $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$, $(\mathcal{C}_j, \mathcal{T}(\mathcal{C}_j), V_j, M_j)$ are linguistic scales representing the criteria, $\mathcal{A}_{i,j} \in \mathcal{T}(\mathcal{C}_j)$ are their linguistic values, $(\mathcal{E}, \mathcal{T}(\mathcal{E}), [0,1], M_e)$ is a linguistic scale representing the evaluation of alternatives and $\mathcal{U}_i \in \mathcal{T}(\mathcal{E})$ are its linguistic values.

In the FuzzME software, rule bases are defined expertly. The user defines such a rule base by assigning a linguistic evaluation to each possible combination of linguistic values of criteria.

For given values of criteria, a resulting fuzzy evaluation is calculated either by the Mamdani fuzzy inference algorithm or by the generalized Sugeno inference.

In the case of the Mamdani fuzzy inference, the degree h_i of correspondence between the given m-tuple of fuzzy values $(A_1', A_2', \ldots, A_m')$ of criteria and the mathematical meaning of the left-hand side of the *i*-th rule is calculated for any $i = 1, \ldots, n$ in the following way

$$h_i = \min\{hgt(A_1' \cap A_{i,1}), \ldots, hgt(A_m' \cap A_{i,m})\}. \quad (7)$$

Then for each of the rules, the output fuzzy value U_i', $i = 1, \ldots, n$, corresponding to the given input fuzzy values, is calculated as follows

$$\forall y \in [0,1] : U_i'(y) = \min\{h_i, U_i(y)\}. \quad (8)$$

The final fuzzy evaluation of the alternative is given as the union of all the fuzzy evaluations that were calculated for the particular rules in the previous step, i.e.,

$$U' = \bigcup_{i=1}^{n} U_i'. \quad (9)$$

Generally, the result obtained by the Mamdani inference algorithm need not be a fuzzy number. So, for further calculations within the fuzzy model, it must be approximated by a fuzzy number.

The advantage of the generalized Sugeno inference algorithm (see [7]) is that the result is always a fuzzy number. In its first step, the degrees of correspondence h_i, $i = 1, \ldots, n$, are calculated in the same way as in the Mamdani fuzzy inference algorithm.

The resulting fuzzy evaluation U is then computed as a weighted average of the fuzzy evaluations U_i, $i = 1, 2, \ldots, n$, which model the mathematical meanings of linguistic evaluations on the right-hand sides of the rules, with the weights h_i. This is done by the following formula

$$U = \frac{\sum_{i=1}^{n} h_i . U_i}{\sum_{i=1}^{n} h_i}. \quad (10)$$

3.4 Overall fuzzy evaluations, the optimum alternative

The final result of the consecutive aggregation of the partial fuzzy evaluations is an overall fuzzy evaluation of the given alternative. The obtained overall fuzzy evaluations are fuzzy numbers on $[0,1]$. They express uncertain degrees of fulfillment of the main goal by the particular alternatives.

The FuzzME software compares alternatives according to the centers of gravity of their overall fuzzy evaluations. A center of gravity of a fuzzy number U on $[0,1]$ that is not a real number, is defined as follows

$$t_U = \frac{\int_0^1 U(x).x\ dx}{\int_0^1 U(x)\ dx}. \quad (11)$$

If $U = u$ and $u \in \Re$, then $t_U = u$. In the FuzzME software, the optimum alternative is the one whose overall fuzzy evaluation has the largest center of gravity.

At present, the FuzzME software is aimed above all at solving multiple-criteria evaluation problems. To ensure high performance in choosing the optimum alternative, it will be necessary to include in the software other methods of ordering of the fuzzy evaluations in the future. Some approaches are proposed in [7] and further research in this area is planned.

4 Example of a practical application of the FuzzME software

The FuzzME software was tested e.g. on a soft-fact-rating problem of one of the Austrian banks. The problem was solved in co-operation with the Technical University in Vienna (see [14]). The fuzzy model of evaluation represents a part of the creditability evaluation of companies carried out by the bank - the evaluation according to soft (qualitative) data, which complements the evaluation according to hard (quantitative) data. The previous practical experience of the bank

ISBN: 978-989-95079-6-8

showed that it is not good to restrict the evaluation to hard data only.

In total, 62 companies were evaluated by the fuzzy model. The goals tree of the model contained 27 qualitative criteria.

During the testing, two approaches were compared - the original soft-fact-rating model used by the bank and the fuzzy models created in the FuzzME software.

The original evaluation model used simple discrete numeric scales with intuitively set linguistic descriptors for the evaluation according to the particular qualitative criteria. The aggregation of partial evaluation was done by the standard weighted average.

In testing by the FuzzME software, the applied linguistically described numeric scales were analyzed. It was found that in some cases the correspondence between the linguistic and numerical values was not perfect. Two new fuzzy models were formed. The first one used uniform fuzzy scales representing a simple fuzzification of the original numeric scales. The other worked with fuzzy values which tried to model, as closely as possible, the linguistic descriptors used in the original evaluation model. The results of the two models were quite different. At the same time, the normalized crisp weights were replaced by normalized fuzzy weights which correspond better to the expert's knowledge about the importance of criteria.

The subsequent discussion on results of the soft-fact-rating showed that there exist criteria values and combinations of criteria values which signalize a substantial danger that the company will go bankrupt or at least will have problems acquitting the debt. That is why, besides the evaluation of companies based on fuzzy weighted average ("average rating"), a fuzzy expert system was applied to calculate another evaluation ("a risk rate of the company"). The particular rules of the base identified the dangerous combinations of criteria values and assigned to them the corresponding risk rates. The solely use of the original fuzzy model without the fuzzy expert system would have lead to a rating score, which may have underestimated the risk inherent to this company. The use of the fuzzy expert systems offers the possibility to visualize and calculate such additional risk combinations.

Finally, both evaluations were aggregated with the fuzzy MINIMAX method. This method is a special case of a fuzzy OWA operator.The resulting evaluation is the infimum of the fuzzy numbers representing the partial evaluations.

The obtained results showed that the solid theoretical basis of the evaluation fuzzy models formed in the FuzzME software improves the quality of evaluations. Positive experiences with such fuzzy models of evaluation could win over the present-day opponents to the soft-fact-rating in the future.

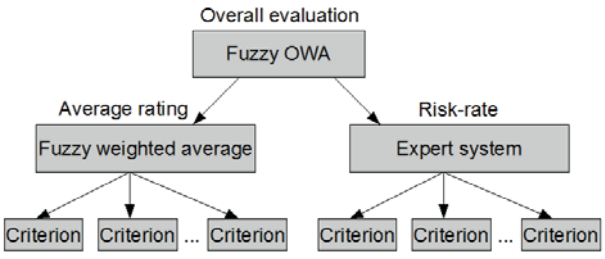


Figure 2: The simplified structure of the used goals tree

5 Conclusion

The software product FuzzME is a result of many years of research in the area of the theory and methods of multiple-criteria fuzzy evaluation. The type of evaluation consistently used in the software corresponds well to the fuzzy sets theory paradigm; the evaluations of alternatives express the fuzzy degrees of fulfillment of given goals. In the FuzzME software, several new methods, algorithms and tools of fuzzy modeling were implemented, e.g.: a structure of normalized fuzzy weights, fuzzy weighted average and ordered fuzzy weighted average operations and algorithms for their calculation, linguistic scales and linguistic variables derived from them. Well-elaborated theoretical basis of the FuzzME software provides a clear interpretation of all steps of the evaluation process and brings understanding of methodology to the user. The FuzzME software is user-friendly. The positive features of the software product proved themselves by solving the mentioned soft-fact-rating problem.

References

[1] R.E. Bellman and L.A Zadeh. Decision-making in fuzzy enviroment. *Management Sci.*, 17 (4):141–164, 2007.

[2] R.R. Yager. On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Trans.Systems Man Cybernet*, 3 (1):183–190, 1988.

[3] Y. J. Lai and C. L. Hwang. *Multiple Objective Decision Making.* Springer - Verlag, Berlin, Heidelberg, 1994.

[4] H. Rommelfanger. *Fuzzy Decision Support Systeme.* Springer - Verlag, Berlin, Heidelberg, 1988.

[5] INFORM GmbH. FuzzyTECH home page. `http://www.fuzzytech.com/`.

[6] C. Von Altrock. *Fuzzy Logic and NeuroFuzzy Applications in Bussiness and Finance*. Prentice Hall, New Jersey, 1996.

[7] J. Talašová. *Fuzzy methods of multiple criteria evaluation and decision making (in Czech).* Publishing House of Palacký University, Olomouc, 2003.

[8] J. Talašová. NEFRIT - multicriteria decision making based on fuzzy approach. *Central European Journal of Operations Research*, 8(4):297–319, 2000.

[9] O. Pavlačka. *Fuzzy methods of decision making (in Czech).* PhD thesis, Faculty of Science, Palacký Univerzity, Olomouc, 2007.

[10] O. Pavlačka and J. Talašová. Application of the fuzzy weighted average of fuzzy numbers in decision-making models. *New Dimensions in Fuzzy Logic and related technologies, Proceedings of the 5th EUSFLAT Conference, Ostrava, Czech Republic, September 11-14 2007, (Eds. M. Štěpnička, V. Novák, U. Bodenhofer)*, II:455–462, 2007.

[11] O. Pavlačka and J. Talašová. The fuzzy weighted average operation in decision making models. *Proceedings of the 24th International Conference Mathematical Methods in Economics, 13th - 15th September 2006, Plzeň (Ed. L. Lukáš)*, pages 419–426, 2006.

[12] J. Talašová and I. Bebčáková. Fuzzification of aggregation operators based on Choquet integral. *Aplimat - Journal of Applied Mathematics*, 1(1):463–474, 2008. ISSN 1337-6365.

[13] B. Kosko. *Fuzzy thinking: The new science of fuzzy logic.* Hyperion, New York, 1993.

[14] K. Fürst. *Applying Fuzzy Models in Rating Systems.* Department of Statistics and Probability Theory, Vienna University of Technology, Vienna, 2008. Term paper.

ISBN: 978-989-95079-6-8

A Computing with Words path from fuzzy logic to natural language *

Claudio Moraga Enric Trillas

European Centre for Soft Computing, 33600 Mieres, Asturias, Spain
{claudio.moraga, enric.trillas}(at) softcomputing.es

Abstract *– Linguistic variables are reviewed and some aspects of the original definition revisited, showing that they provide a first formal frame to do computing with words in a "narrow sense". A two-levels hierarchy of languages is discussed in order to move to simple cases of computing with words in natural language, supported by linguistic variables.*

Keywords – Computing with words, linguistic variables, meaning modifiers

1 Introduction

In the late 90's, L.A. Zadeh introduced the idea of *computing with words* [1]. Much efforts have been done ever since to face the challenge and to develop the required proper concepts, models and methods to turn the bright initial idea into a sound area of research. Fuzzy logic seems predestined to be an appropriate formal frame to support the above efforts. Among other things, because fuzzy if-then rule bases may be understood as a specification of (the model of) a system in a language close to natural language and each rule, as the specification of a computation based on words, even though at the moment of doing real computations, this is done at the level of numbers. At this time, fuzzy if-then rule bases may be considered as a metaphor of what we would like to achieve by *really* computing with words. In this sense, what is currently known as fuzzy logic may be considered as a particular case of computing with words.

This paper shows, that in the context of fuzzy linguistic variables there are already processes that are representative of computing with words, albeit possibly in a "narrow sense". Furthermore it will be shown that it is possible to build a hierarchical relationship between words based on linguistic variables at the level of fuzzy logic and "words" – (actually short sentences)- at the level of natural language; to move from linguistic variables to "linguistic expressions" and from linguistic modifiers to "meaning modifiers". In analogy to the fact that linguistic modifiers, when applied to linguistic variables constitute possibly the simplest case of computing with words, it will be shown that "meaning modifiers", when applied to "linguistic expressions" may lead to non-trivial linguistic results of non-negligible complexity, in what may be thought of, as computing with "words".

The rest of the paper provides first a set of needed definitions to continue with a discussion about computing with words in fuzzy logic. An extension to computing with "words" in natural language is analyzed in the closing section.

2 Definitions

Definition 1 [1]:
Computing with words and perceptions, or CWP for short, is a mode of computing in which the objects of computation are words, propositions and perceptions described in a natural language.

Definition 2:
Let Ω be a set of (non-ambiguous) words that are used in a given context. The elements of Ω are considered to be pairs (representation, meaning). Representations may be taken from different syntactic domains, but meanings, from a single semantic domain. Moreover, let Γ be a finite set of functions $\{\gamma_1, \gamma_2, \ldots, \gamma_k\}$, with

$$\gamma_i : \Omega^n \rightarrow \Omega$$

The functions $\{\gamma_1, \gamma_2, \ldots, \gamma_k\}$ must be *designed* in such a way, that applied to words of Ω will produce reasonable words of Ω for the given context.

Definition 3:
In this paper when speaking of computing with words, the *objects* of computation are specified in definition 1. The *agents* of computation are functions taken from Γ, as stated in definition 2.

Examples: Let Ω be a subset of English.
If γ_1 is an appropriate "association" function. Then
γ_1(parent, children) = family
If γ_2 is a synonym function. Then
γ_2(pretty) = beautiful
If γ_3 is an antonym function. Then
γ_3(near) = far

These examples show that since the functions in Γ represent some "linguistic transformation", their interpretation should be taken take in account in the line of language precisiation (*à la* Zadeh).

At this stage, it should be mentioned that not under the scope of "computing with words", but simply with the goal of designing good fuzzy if-then rules, work has been done to select or build the most convenient operations connecting premises and the conclusion [2], [3]. In the context of this paper, the given references correspond to a basic contribution in the design of the most convenient functions for the corresponding Γ.

* This work has been partially supported by the Foundation for the Advancement of Soft Computing (Asturias, Spain) and CICYT (Spain) under project TIN2008-06890-C02-01

ISBN: 978-989-95079-6-8

3 Computing with words in fuzzy logic

One of the key concepts in the development of fuzzy logic is that of a *linguistic variable*, introduced by L.A. Zadeh in 1975 [4]. The value set of a linguistic variable is a set of words –"linguistic terms"– denoting predicates or linguistic labels, and the syntactic domain of those words is constituted by appropriate convex fuzzy sets. "*… in general, a linguistic variable is associated with two rules: (1) a syntactic rule, which may have the form of a grammar for generating the names of the values of the variable; and (2) a semantic rule which defines an algorithmic procedure for computing the meaning of each value"* (See section 2 of Part II of [4])

In what follows, this idea is extended to include a formal grammar and a data base. The grammar will control possible transformations on the fuzzy sets representing the linguistic terms of a linguistic variable. The effect of linguistic modifiers (originally called "hedges of type I" [5]) as well as the generation of antonyms of linguistic terms (which was studied much later [6], [7], [8]) may, for instance, be formalized in terms of productions of a grammar. The data base, on the other hand, is meant to contain the meanings associated to the linguistic terms and their considered possible modifications through the grammar.

It is easy to see that the grammar and the data base have a functionality which is subsumed by that of the functions presented in Definition 2.

For the representation of linguistic terms as fuzzy sets, if no further information is available, a trapezoidal shape is appropriate since an agreement among users about the subset of the universe where the predicate is indeed valid, defines the core of the fuzzy set and the agreement about the subset of the universe where the predicate is not at all valid, defines its co-support. Our human experience with predicates indicates that the transitions between co-support and core are continuous and monotone. Therefore, if no further information is available, linear transitions leading to a piecewise linear convex normalized fuzzy set is an appropriate first choice. If the core contains only one point, then the trapezium reduces to a triangle. Notice that the above analysis does not require the trapezium (triangle) to be symmetric.

Trapezoidal fuzzy sets are simple to represent as an ordered quadruple of the ordinates of their four corners (t_1, t_2, t_3, t_4). The interval between the first and the fourth corner represents the support of the fuzzy set. (If non-normalized trapezoidal fuzzy sets are considered, then a preceding scaling factor is needed.)

Another basic assumption, normally associated to linguistic variables, is that all linguistic terms of the variable constitute a partition of unity, *i.e.*, at every point of the universe of discourse, the membership degrees of the linguistic terms add up to 1.

Definition 4:

A linguistic modifier is a unary function in the set of fuzzy sets representing linguistic terms. A linguistic modifier LM is *compressing* if for any fuzzy set A representing a linguistic term, $LM(A) \subset A$, and is *expanding* if $A \subset LM(A)$. There are however linguistic modifiers that are neither compressing nor expanding.

In what follows, to keep the notation as simple as possible, the label of a linguistic term will also be used to denote the membership function of the fuzzy set representing it.

Let $T := (t_1, t_2, t_3, t_4)$

$LM(T) := ((t_1+t_2)/2, t_2, t_3, (t_3+t_4)/2)$ is a compressing modifier and in what follows will be given the meaning *more_strictly*(T).

$LM(T) := (t_1, (t_1+t_2)/2, (t_3+t_4)/2, t_4)$ is an expanding modifier and in what follows will be given the meaning *roughly*(T).

Lemma 1:

Let $T := (t_1, t_2, t_3, t_4)$

Define $T_{left} := ½(t_1, (t_1+ t_2)/2, (t_1+ t_2)/2, t_2)$

$T_{right} := ½(t_3, (t_3+ t_4)/2, (t_3+ t_4)/2, t_4)$

Then:

$T = T_{left} + more_strictly(T) + T_{right}$

Proof: By adding point-to-point the piecewise linear segments.

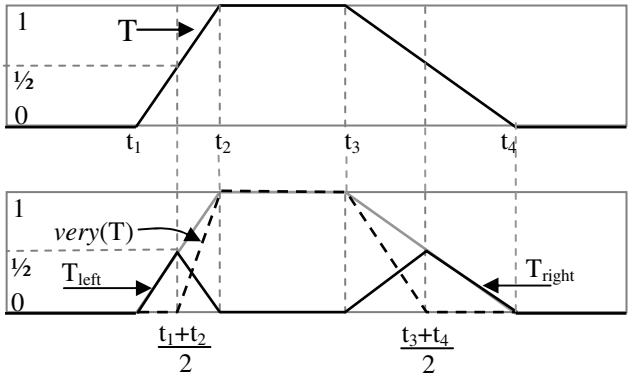


Figure 1**:** The relationship among the fuzzy sets T, *more_strictly*(T), T_{left} and T_{right}

It is easy to see that for all $x < t_3$, T_{right} takes the value 0. Moreover, since $(t_1 + t_2)/2$ is equidistant of t_1 and t_2, then at this point, T takes the value ½. (See Figure 1). Moreover between t_2 and t_3, both T_{left} and T_{right} have membership degree 0, meanwhile T and *more_strictly*(T) have membership degree 1, therefore in this interval the claim trivially holds. Table 1 summarizes the relevant membership values at t_1, $(t_1 + t_2)/2$ and t_2.

Table 1: relevant membership values

	t_1	$(t_1 + t_2)/2$	t_2
T	0	½	1
more_strictly(T)	0	0	1
T_{left}	0	½	0
more_strictly(T) + T_{left}	0	½	1

In the interval $[t_1, t_2]$, both *more_strictly*(T) and T_{left} are piecewise linear; therefore their sum will also be piecewise linear. In the subinterval $[t_1, (t_1+t_2)/2]$, since the membership value of *more_strictly*(T) is 0, the sum will be given by T_{left}, which has a representation as a straight line segment with slope $1/(t_2-t_1)$. In the subinterval $[(t_1+t_2)/2, t_2]$, both *more_strictly*(T) and T_{left} have a representation as straight line segments, therefore also their sum. It is simple to see

ISBN: 978-989-95079-6-8

from Figure 1, that the sum segment runs from the point $\langle(t_1+t_2)/2, 1/2\rangle$ to the point $\langle t_2, 1\rangle$, with slope $1/(t_2-t_1)$. Since both sum segments share a common point and have the same slope, they constitute a single straight line segment with slope $1/(t_2-t_1)$ in the whole interval $[t_1, t_2]$. This corresponds with the left part of T. A similar analysis will prove that in $[t_3, t_4]$ the sum of *more_strictly*(T) plus T_{right} returns the corresponding part of T. For reasons that will become clear below, no attempt will be done here to associate a meaning to T_{right} and T_{left}.

Corollary 1.1:
Assume that T is a rectangular trapezium. Without loss of generality let T := (t_1, t_2, t_3, t_3). Then :

$T = T_{left} + more_strictly(T)$

since T_{right} has zero-support.

Example 1:
Consider a linguistic variable with three linguistic terms building a partition of unity, as shown in Figure 2(a). Furthermore, let it be assumed that some experiments done using this representation –(model)- of the variable, made advisable a refinement of the middle linguistic term, and that this is done by applying Lemma 1 (with $t_2 = t_3$). This is shown in Figure 2(b). Solving the superposition of the small triangles with the respective trapeziums leads to a new model of the linguistic variable, as shown in Figure 2(c).

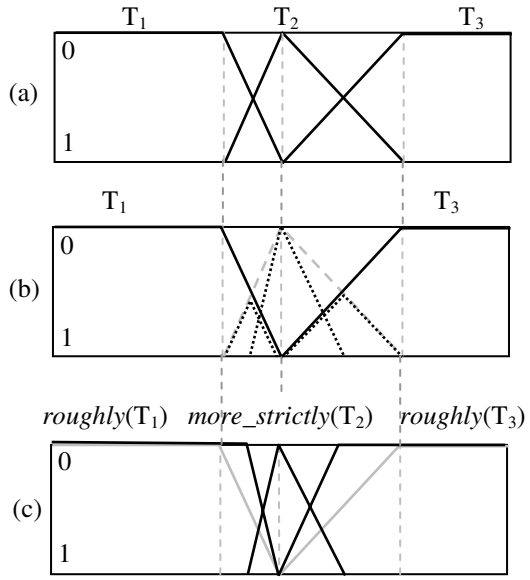


Figure 2**:** (a) Initial model of the linguistic variable
(b) Effect of applying Lemma 1 to T_2
(c) The resulting new model for the linguistic variable

An analysis of the process shows that T_2 changed into *more_strictly*(T_2). Furthermore, the core of T_1 was expanded by one half of the support corresponding to the right wing of the trapezoidal representation. From Definition 4 follows that it represents *roughly*(T_1). Similarly, T_3 turns into *roughly*(T_3). In summary, a new representation of the linguistic variable, which preserves the partition of unity, was obtained through a refinement process by using appropriate linguistic modifiers. This represents (possibly the simplest) elementary operations in the context of computing with words.

Case study 1:
In the context of linguistic variables with structure of a partition of unity, *two* consecutive linguistic terms T_j and T_{j+1} will be considered, which will be refined under Lemma 1.

Let T_j := (t_1, t_2, t_3, t_4)
and T_{j+1} := (t_3, t_4, t_5, t_6)
Lemma 1 generates:

$more_strictly(T_j)$:= $((t_1+ t_2)/2, t_2, t_3, (t_3+ t_4)/2)$
$(T_j)_{right}$:= $½(t_3, (t_3+ t_4)/2, (t_3+ t_4)/2, t_4)$
$more_strictly(T_{j+1})$:= $((t_3+ t_4)/2, t_4, t_5, (t_5+ t_6)/2)$
$(T_{j+1})_{left}$:= $½(t_3, (t_3+ t_4)/2, (t_3+ t_4)/2, t_4)$

(Lemma 1 also generates $(T_j)_{left}$ and $(T_{j+1})_{right}$, which are however not in the focus of this case study).
Notice that $(T_j)_{right} = (T_{j+1})_{left}$ therefore
if $T' = (T_j)_{right} + (T_{j+1})_{left}$ then
T' := $(t_3, (t_3+ t_4)/2, (t_3+ t_4)/2, t_4)$
It is easy to see that T' is an isosceles triangle. To T' the following *specification of meaning* can be associated:

$T' = between(more_strictly(T_j), more_strictly(T_{j+1}))$

and, the other way around, the operation *between* is defined as the structure of T'. See Figures 3(a) and 3(b).

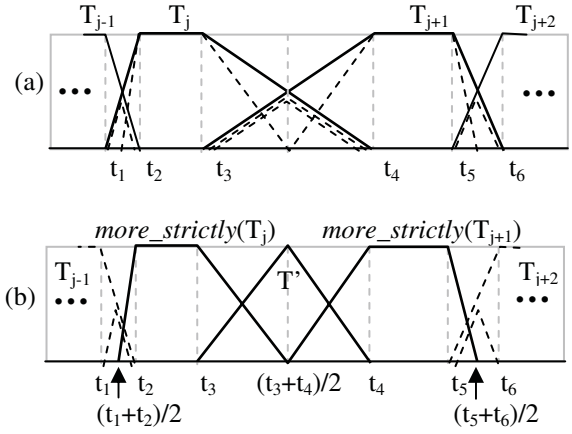


Figure 3**:** (a) Two neighbor linguistic terms and their refinement with Lemma 1 (dotted).
(b) Effect of the superposition of the overlapping side triangles generated by the refinement.

From $T' = between(more_strictly(T_j), more_strictly\ (T_{j+1}))$ it may be concluded that there exists a $\gamma \in \Gamma$ such that

$\gamma = between \circ (more_strictly \times more_strictly)$

and $T' = \gamma(T_j, T_{j+1})$

Notice that if all linguistic terms of a linguistic variable are refined with Lemma 1, then all "new" linguistic terms may be obtained with the composition of *between* and the Cartesian product *more_strictly* × *more_strictly*. A refinement of all linguistic terms after Lemma 1, may be related to wavelets [9], but almost duplicates the number of linguistic terms. Partial refinements, however, were studied in [10] under the scope of a metasemantics.

ISBN: 978-989-95079-6-8

Before studying a next case, two elementary operations will be introduced.

Definition 5:
Let $T_j := (t_1, t_2, t_3, t_4)$
Then, the operation *less than or equal to*, abbreviated "LE" is defined for all x in the universe of discourse, as:

$$(LE(T_j))(x) = \begin{cases} 1 & \text{if } x \le t_3 \\ T_j & \text{if } x > t_3 \end{cases}$$

Similarly, the operation *more than or equal to*, abbreviated "ME" is defined as:

$$(ME(T_j))(x) = \begin{cases} 1 & \text{if } x \ge t_2 \\ T_j & \text{if } x < t_2 \end{cases}$$

Case study 2:
Consider a linguistic variable with more than 3 linguistic terms and refine some linguistic term in the middle. What will be the effect upon its non-refined neighbors?

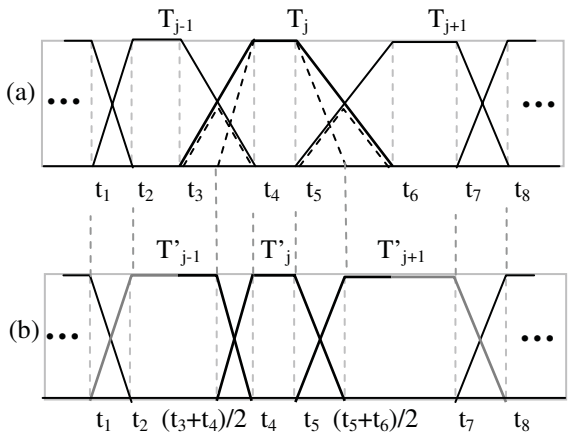


Figure 4 **:** (a) A linguistic variable with the prepared refinement of the j-th linguistic term.
(b) Effect of refining T_j upon its neighbors.

Figure 4 shows the linguistic terms before and after refining the j-th term. From Lemma 1 follows that:

$$T'_j = more_strictly(T_j)$$

An interpretation of T'_{j+1} is not straight forward. Fig. 4(b) shows however that at the left hand side, the core of T'_{j+1} increased by one half of the support of the left wing, *i.e.*, by $\frac{1}{2}(t_6-t_5)$, with respect to the original core of T_{j+1}. Notice that this is the same (partial) effect caused by the operation $roughly(T_{j+1})$; however, the right hand side of T'_{j+1} does not change as compared to T_{j+1}. This may be given a formal expression as follows, for all x in the universe of discourse:

$$(T'_{j+1})(x) = minimum(roughly(T_{j+1}), LE(T_{j+1}))$$

or "in words",

$$T_{j+1} = roughly(T_{j+1})\ and\ LE(T_{j+1})$$

where *and* is realized with the operation *minimum*.

Similarly,

$$(T'_{j-1})(x) = minimum(roughly(T_{j-1}), ME(T_{j-1}))$$

or "in words",

$$T_{j-1} = roughly(T_{j-1})\ and\ ME(T_{j-1})$$

Example 1 and the two above cases show that at the lowest level, that of terms of linguistic variables, transformations may be formalized, which represent the possibly simplest cases of computing with words in narrow sense.

4 First steps towards natural language

When working with formal languages, "words" are build with "symbols" of an alphabet. Quotations are being used, since the "symbols" may be words of a reference language and, consequently the "words" will (normally) be (meaningful) sentences of the reference language. This model of a hierarchy of languages will be used in an attempt to transfer or imitate the elementary operations discussed above to short statements in natural language. These "words" will be written in square brackets. Furthermore the notation "linguistic statement" instead of linguistic variable, and "meaning modifier" instead of linguistic modifier will be used. Finally it becomes apparent that "words" will no longer be just pairs (representation, meaning), but at least triples (representation, structure, meaning). Representation is concerned with the rules of writing, meanwhile the structure refers to the grammatical ordering of the reference words to support a meaning.

Example 3:
Consider the "word" [*The weather is improving*], where the linguistic statement has the value set {hardly a bit, slightly, steadily, strongly}. How could the meaning of this "word" be emphasized?

i) *Use* the "word" consistently within the value subset {steadily, strongly}

ii) Design a meaning modifier MM_1 which integrates explicitly in the "word" the desired predicate from the value set, preserving the structure:
MM_1 (***steadily***, [*The weather is improving*]) =
[*The weather is* ***steadily*** *improving*]

It may be observed that this modifier changes the representation of the "word", preserves the grammar of the "word" and emphasizes its meaning.

iii) Design a meaning modifier MM_2 which integrates an additional reference word, taken from a pre-defined "subset of emphasizers" of the corresponding natural language, preserving the structure, to produce an enhancement:

MM_2 (***indeed***, [*The weather is improving*]) =
[*The weather is* ***indeed*** *improving*]

Notice that it is only claimed that both [*The weather is* ***steadily*** *improving*] and [*The weather is* ***indeed*** *improving*]

ISBN: 978-989-95079-6-8

emphasize the meaning of [*The weather is improving*], but not that they are synonyms.

Example 4:
Consider the "word" [*Yesterday it rained*]. A value set for the linguistic statement may be thought of as $D_1 \times D_2 \times D_3$, where D_1 = {in the morning, at midday, in the afternoon, in the evening, at night}, D_2 = {a couple of minutes, several hours, the whole day} and D_3 = {a few drops, moderately, cats-and-dogs}. Notice that D_1 and D_2 may be associated both to *yesterday* and to *rain*, meanwhile D_3 is clearly related only to *rain*.

i) It is requested that [*Yesterday it rained*] be emphasized. The *rained*-component may be emphasized in a way similar to that discussed in example 3. Can *yesterday* be emphasized? At the level of natural language, it is possible to design a 2-place meaning modifier MM_1 giving:

MM_1([*December the 1st*], [*Yesterday it rained*]) =
[*Yesterday December the 1st it rained*],

which may be considered as a possible way of emphasizing *yesterday*.

ii) In the context of [*Yesterday it rained*] D_1, D_2 and D_3 may be seen as value sets of the linguistic variables "when", "how_long" and "how_much" (at the reference level), respectively. As linguistic variables, they have grammars that, among other things, may change number, shape and distribution of the fuzzy sets representing the corresponding linguistic terms, as well as a database keeping track of their meanings. Assume that the corresponding grammars are G_1, G_2 and G_3. Let $r_{2,i}$ and $r_{3,j}$ represent the i-th and j-th rules of G_2 and G_3, respectively, such that:

$r_{2,i}(D_2)$ = {a couple of minutes, *half an hour*, several hours, the whole day}

$r_{3,j}(D_3)$ = {a few drops, moderately, *a lot*, cats-and-dogs}

Some elements of the new domains may have the same labels as in the original domains, but will possibly be different since at least some of them are shifted and compressed, because in both cases, a new linguistic term was introduced. (Notice here the importance of the data base with meanings).
Then it is possible to use linguistic modifiers at the low level to build meaning modifiers MM (at the upper level) to generate new, richer "words" as, for instance,

MM[*Yesterday it rained*] =
[*Yesterday it rained quite a lot about half an hour almost at midday*]

It may be observed that the meaning modifier reordered the domains as $r_{3,j}(D_3) \times r_{2,i}(D_2) \times D_1$, made use of the new generated linguistic terms "half an hour" and "a lot", applied the linguistic modifiers "quite" (to "a lot") "about" (to "half an hour") and "almost" (to "at midday"). Fuzzy logic offers sound methods to perform all these reference level operations, which put together, formally specify the meaning modifier.

Analysis of the meaning modification
[*quite a lot about half an hour almost at midday*]:

Recall that the value set for [*Yesterday in rained*] was specified as the Cartesian product of three domains. Therefore, its "value" is a triple, whose components are possibly modified linguistic terms within the prevailing versions of the corresponding domains. Therefore, each one of the linguistic terms may be analyzed separatedly and the required linguistic modifiers, *designed* to satisfy the requirements of the use of the language.

i) Design of *quite*, to produce the modified linguistic term *quite a lot*.
It should be noticed that *quite* is a rather complicated linguistic modifier, since it is very context dependent: in American English it seems to be used mostly to emphasize, in a way similar to *more_strictly*, meanwhile in British English it may also be ironically used to mean exactly the contrary. Furthermore, *quite* is also used as an idiom, like in "*quite* a few", possibly meaning the same as "*quite* a lot". In the case under analysis however, it seems that *quite* is intended to be used to suggest rather more precisely, what is expressed by "a lot" alone. If this is the case, then a compressing linguistic modifier should be used, but with possibly a smaller amount of compression than in the case of *more_strictly*.
A general analysis leading to a proposal for a prototype of *quite* is shown in Figure 5.

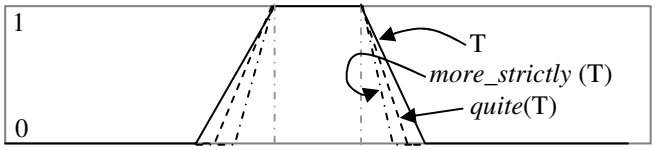


Figure 5**:** A possible prototype for *quite* as a linguistic modifier, which is less compressing than *more_strictly*

Observations on the *use* of *quite* in the context where it will be applied and in similar contexts will be needed to properly tune this modifier. Notice that it may well be the case that the contexts at the reference level and at the natural language level pose different constraints for the design of the modifier.

ii) Design of *about*, to produce the modified linguistic term *about half an hour*.
In the use of the language, *about* half an hour, appears to have a meaning close to that of *around* half an hour or *roughly* half an hour. Experiences with numerical approximation with fuzzy numbers indicate that *around* is an expanding modifier that preserves the core –(which in this case contains only one point, since the fuzzy set is triangular)– and increases the support. It seems reasonable to extend this modifier to trapezoidal fuzzy sets in the same way. On the other hand, *roughly*, as defined earlier, expands the core and preserves the support. If the meaning of *about* half an hour is indeed close to that of *around* half an hour and *roughly* half an hour, then the corresponding modified fuzzy sets should exhibit a high degree of "family resemblance" [11] and this can be used as an additional guiding constraint to design *about*.
A proposal for a prototype of *about* is illustrated in Figure 8 with respect to some abstract T (to more clearly show the effect of the (known) linguistic modifiers and of the proposed prototype).

ISBN: 978-989-95079-6-8

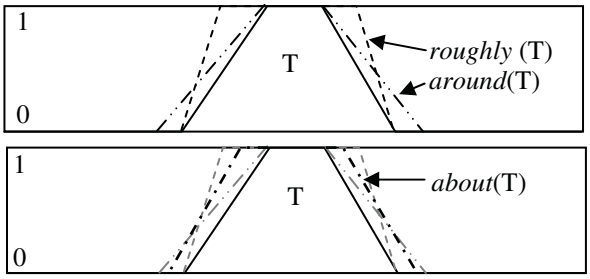


Figure 6: A possible prototype for *about*, for some abstract T, as related to *roughly* and *around*

As in the former case, observations on the *use* of the modifier in the target context will be needed to properly tune it.

iii) Design of *almost*, to produce the modified linguistic term *almost at midday*.
The use of the language, *e.g.* "the glass is *almost* empty" or "Peter was *almost* crazy with so many problems", indicates that this is a one-sided shifting and possibly expanding modifier, which applied to a fuzzy set representation of a linguistic term, displaces and possibly expands the fuzzy set towards the "less than" side. (See Figure 7). It is simple to see that *almost at midday* and *around midday* will have some "family resemblance", but to a clear weaker degree than the one discussed in the former paragraph.

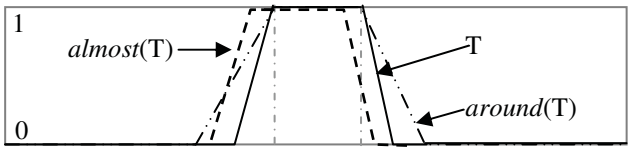


Figure 7: A possible prototype for *almost*, as compared to *around*, applied to some abstract T

Observations on the *use* of all above modifiers in the actual target context, including experimentation at the level of the language, will be needed to ensure their empirical correctness. See, for instance [12], [13], [14].

5 Conclusions

Within the formalism of linguistic variables in fuzzy logic, linguistic modifiers represent a very simple case of computing with words. Some new elementary operations have been defined, that allow new non trivial instances of low level computing with words. The paper shows that it is possible to extend these ideas to "linguistic statements" (short sentences) and "meaning modifiers" in natural language. The "linguistic statements" become the new "words" (as is the case of programming languages when considered as context free formal languages). This is done however keeping a solid base on linguistic terms and linguistic modifiers. The hierarchy of languages suggests that the more about computing with words is learnt in fuzzy logic, the more can be projected to natural languages to do computing with "words".

Acknowledgement

The authors thank the reviewers for their constructive criticism and the suggestions for the improvement of the paper.

References

[1] L.A. Zadeh, From computing with numbers to computing with words. From manipulation of measurements to manipulation of perceptions. *IEEE Trans. Circuits and Systems* **45** (1): 105-119, 1999

[2] C. Moraga, A. Pradera and E. Trillas, Evolutionary tuning of fuzzy if-then rules at the level of operations: A proposal. *Proc. III Congr. Español sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados MAEB'2003*, 530-537. Press University of Oviedo, 2003

[3] E. Trillas, On some features of fuzzy logic. *Proc. 27th Annual Meeting of the North American Fuzzy Information Processing Society NAFIPS.* 2008

[4] L.A. Zadeh, The concept of linguistic variable and its application to approximate reasoning. *Information Sciences* **8**: 199-249; 301-357; **9**: 43-80, 1975

[5] L.A. Zadeh, A fuzzy-set-theoretic interpretation of hedges. Memorandum M-335, Electronics Research Laboratory, University of California, Berkeley, 1972

[6] E. Trillas and T. Riera, Towards a representation of synonyms and antonyms by fuzzy sets. *Busefal* V:30-45, 1981

[7] E. Trillas, S. Cubillo and E. Castiñeira, On antonymy from fuzzy logic. *Proc. 9th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge Based Systems (IPMU'00)*, 79-84, 2000

[8] E. Trillas , C. Moraga, S. Guadarrama, S. Cubillo and E. Castiñeira, Computing with Antonyms. *Studies in Fuzziness and Soft Computing* **217**: 133-153, Springer, 2007

[9] A. Rodriguez de Soto, Building a hierarchical set of fuzzy rules using Wavelet Theory. *Proc. 9th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge Based Systems (IPMU'00)*, 1764-1769, Madrid, 2000.

[10] C. Moraga, A metasemantics to refine fuzzy if-then rules. *Proc. IEEE Int. Symp. on Multiple-valued Logic*, 148-153. IEEE-CS-Press, 2004

[11] E. Trillas, C. Moraga and A. Sobrino, On "family resemblances" with fuzzy sets. *Proc. IFSA'2009*, IFSA-Press, 2009

[12] J. Hellendorn, Reasoning with Fuzzy Logic. Dissertation, Delft University of Technology, The Netherlands, 1990

[13] V. Novak, A horizon shifting model of linguistic hedges for approximate reasoning, *Proceedings of the 5th IEEE International Conference on Fuzzy Systems,* vol. 1, 423–427, IEEE-CS-Press, 1996

[14] P.J. Macvicar-Whelen, Fuzzy sets, the concept of height, and the hedge 'very'. *IEEE Trans. On Systems, Man, and Cybernetics*, SMC-8, 507-511, 1978

ISBN: 978-989-95079-6-8

Approximation properties of inverse fuzzy transforms over a residuated lattice

Martina Daňková Martin Štěpnička
Institute for Research and Applications of Fuzzy Modeling, University of Ostrava
Ostrava, Czech republic
Email: {martina.dankova,martin.stepnicka}@osu.cz

Abstract— *This contribution focuses on inverse fuzzy transforms (shortly inverse F-transforms) over residuated lattices introduced by I. Perfilieva and their approximation properties. We will try to reduce some requirements used in the original work to prove Approximation theorem. Moreover, we show in which sense F-transforms are the best approximations.*

Keywords— Fuzzy transform, Approximation, Extensional relation.

1 Introduction

In [1], F-transforms were introduced basically in three forms: the first (ordinary) was constructed over the ordinary algebra of reals[1], while the other two were constructed over the residuated lattice. Note that up to now, most of the applications of F-transforms, e.g. ordinary/partial differential equations [2]/[3], image compression/fusion (see [4]/[5, 6]), data analysis [7] etc., employ only the ordinary one. The main reason follows from the character of the inverse F-transforms, where the ordinary F-transform average the function values. Whereas the other two approximate them from above and from below, respectively. And since these F-transforms (over the residuated lattice) can be viewed as formalizations of a collection of *graded fuzzy rules* [8], they are primarily suitable for applications, which use fuzzy "IF-THEN" rules.

In this contribution, we will focus on the inverse F-transforms over the residuated lattice and show their approximation properties. Moreover, we will provide a simpler way to get the approximation theorem than the one in [1] (Theorem 10 and 11). Thus, we would like to brush up this topic and make it more attractive for a broader community.

Note that we present mainly an excerpt of a larger manuscript [9] that has been submitted for publication. For background information and proof details, readers are referred to this upcoming article.

2 Basic definitions and overview of the known results.

2.1 Basic definitions

Definition 1 *A residuated lattice on L is an algebra*

$$\mathcal{L} = \langle L, \vee, \wedge, *, \to, \mathbf{0}, \mathbf{1} \rangle \tag{1}$$

with four binary operations and two constants such that

- $\mathcal{L} = \langle L, \vee, \wedge, \mathbf{0}, \mathbf{1} \rangle$ *is a lattice with the largest element* $\mathbf{1}$ *and the least element* $\mathbf{0}$ *w.r.t. the lattice ordering* $\leq$,
- $\mathcal{L} = \langle L, *, \mathbf{1} \rangle$ *is a commutative semigroup with the unit element* $\mathbf{1}$*, i.e.* $*$ *is commutative, associative, and* $\mathbf{1} * x = x$ *for all* $x \in L$,
- $*$ *and* $\to$ *form an adjoint pair, i.e.*

$$z \leq (x \to y) \quad \textit{iff} \quad x * z \leq y \quad \textit{for all } x, y, z \in L.$$

In the sequel, let us assume $\mathcal{L}$ be a residuated lattice of the form (1). Moreover, we define the biresidual operation (biresiduum)

$$x \leftrightarrow y = (x \to y) \wedge (y \to x).$$

Throughout the whole text, we will deal with fuzzy relations whose membership functions take values from the support of $\mathcal{L}$ and we denote this fact by $\subseteq$. Let $M_1, \ldots, M_n$ be some nonempty sets of objects, $\mathcal{R}_I = \{R_i \subseteq M_i^2\}_{i \in I}$ be a system of binary fuzzy relations ordered by $I = \{1, \ldots, n\}$, $f \subseteq M_1 \times \ldots \times M_n$ be an n-ary fuzzy relation.

Convention 2 *For the sake of brevity, we will denote* $M_1 \times \cdots \times M_n$ *by* $M^{(n)}$*. Moreover, we will write* $R(\bar{x}, \bar{y}) = R_1(x_1, y_1) * \cdots * R_n(x_n, y_n)$.

Definition 3 *We say that f is extensional w.r.t. $\mathcal{R}_I$ if*

$$R(\bar{x}, \bar{y}) * f(\bar{x}) \leq f(\bar{y}),$$

for each $\bar{x}, \bar{y} \in M^{(n)}$.

Let us recall that a fuzzy $*$-equivalence (also known as similarity or indistinguishability) $R \subseteq M^2$ (M be some nonempty set of objects) is reflexive, symmetric and transitive w.r.t. $*$. Observe that $\leftrightarrow$ is $*$-equivalence on L. In the following we will omit specification $*$ whenever it will be clear from the context.

Additionally, we recall notions used in the original approximation theorems in [1].

Definition 4 *A system of fuzzy sets* $\{A_1, \ldots, A_k\}$ *on a nonempty set M establishes a semi-partition if*

$$\bigvee_{x \in M} A_i(x) * A_j(x) \leq \bigwedge_{x \in M} A_i(x) \leftrightarrow A_j(x),$$

for each $i, j \in \{1, \ldots, k\}$.

[1] Due to a nature of the used operations, it can be reinterpreted over ŁΠ-algebra.

ISBN: 978-989-95079-6-8

2.2 $F^{\uparrow(\downarrow)}$-transforms

Let us recall inverse F-transforms from [1] generalized for the n-dimensional case[2]:

$$\begin{aligned}
&\text{Assume } f \subsetsim M^{(n)}, C \subseteq M^{(n)}, A_{j_i} \subsetsim M_i,\\
&\qquad \forall j \in J = \{1,\dots,k\}, \forall i \in I = \{1,\dots,n\},\\
&\qquad \text{and } \mathbf{j} \text{ be abbreviation for } [j_1,\dots,j_n],\\
&f^{\downarrow}_{F,k}(\bar{x}) = \bigvee_{\mathbf{j}\in J^n} A_{\mathbf{j}}(\bar{x}) * [\bigwedge_{\bar{c}\in C} A_{\mathbf{j}}(\bar{c}) \to f(\bar{c})],\\
&f^{\uparrow}_{F,k}(\bar{x}) = \bigwedge_{\mathbf{j}\in J^n} A_{\mathbf{j}}(\bar{x}) \to [\bigvee_{\bar{c}\in C} A_{\mathbf{j}}(\bar{c}) * f(\bar{c})], \text{ where}\\
&A_{\mathbf{j}}(\bar{x}) = A_{j_1}(x_1) * \ldots * A_{j_n}(x_n), \forall \mathbf{j} \in J^n.
\end{aligned}$$

In [1], $f^{\downarrow}_{F,k}$ is called inverse F$^{\downarrow}$-transform and $f^{\uparrow}_{F,k}$ inverse F$^{\uparrow}$-transform. Moreover, the expression in the square brackets above is denoted by $F^{\uparrow(\downarrow)}_{\mathbf{j}}$ in the case of $f^{\uparrow(\downarrow)}_{F,k}$ and known as $\mathbf{j}$-th component of F$^{\uparrow(\downarrow)}$-transform.

Below, we overview known results from [1] that are in the center of our interest:

1. From the inequality $a * (a \to b) \le b \le a \to (a * b)$, it follows that for an arbitrary f:
$$f^{\downarrow}_{F,k}(\bar{x}) \le f(\bar{x}) \le f^{\uparrow}_{F,k}(\bar{x}),\ \forall \bar{x} \in C. \tag{2}$$
2. The $\mathbf{j}$-th components of F$^{\uparrow}$-transform and F$^{\downarrow}$-transform are the least elements of the following respective sets:
$$S_{\mathbf{j}} = \{a \in M^{(n)} |\ A_{\mathbf{j}}(\bar{x}) \le f(\bar{x}) \to a,\ \forall \bar{x} \in C\}, \tag{3}$$
$$T_{\mathbf{j}} = \{a \in M^{(n)} |\ A_{\mathbf{j}}(\bar{x}) \le a \to f(\bar{x}),\ \forall \bar{x} \in C\}. \tag{4}$$
3. A core of the approximation theorems inheres in the following inequality:
$$\bigvee_{\mathbf{j}\in J^n} A_{\mathbf{j}}(\bar{x}) * A_{\mathbf{j}}(\bar{x}) \le f^{\uparrow(\downarrow)}_{F,k}(\bar{x}) \leftrightarrow f(\bar{x}),\ \forall \bar{x} \in C. \tag{5}$$
Remark that a closeness of elements is expressed using $\leftrightarrow$, which is the dual concept to pseudo-metric as shown in [10], and naturally, the estimation goes there from below.

Below, we summarize the requirements that were used in [1] to prove (5), i.e., the estimation of the equivalence of $f^{\downarrow(\uparrow)}_{F,k}$ to the original function f:

	Requirements in [1] (Theorems 10 and 11)
1.	$\mathcal{L}$ be a BL-algebra,
2.	each $A \in \{A_{j_i}\}_{j\in J, i\in I}$ be normal,
3.	$\{A_{1_i},\dots,A_{k_i}\}$ forms a semi-partition $\forall i \in I$,
4.	f be extensional w.r.t. $\mathcal{P}_I = \{P_i\}_{i\in I}$ defined by $P_i(x,y) = \bigwedge_{j\in J} A_{j_i}(x) \leftrightarrow A_{j_i}(y)$, for all $i \in I$,
5.	$E = \{\bar{e}_j \in M^{(n)} |\ P_i(x, e_{j_i}) = A_{j_i}(x), i \in I, j \in J\}$ be subset of C (see p. 1013 in [1]).

[2] Note that in the original work, f was a mapping from a finite subset of L into L (support of the fixed residuated lattice $\mathcal{L}$), where $L = [0,1]$. Since we measure a precision of the approximation using values of L and we do not use specific properties of $[0,1]$, this determination is not needed in our work.

3 Approximation properties

3.0.1 Formulation of the problem and basic properties

Let $\mathcal{L}$ and $\mathcal{R}_I = \{R_i \subsetsim M_i^2\}_{i\in I}$ be as specified in Subsection 2.1, and $\varepsilon \in L$. Moreover, we assume to have only a partial information about $f \subsetsim M^{(n)}$ in the form of a set of samples $\{f(\bar{c}) |\ \bar{c} \in C \subseteq M^{(n)}\}$. The problem (ApproxF) that we are going to solve is formulated as follows:

$$\begin{aligned}
&\text{Find } D \subseteq M^{(n)} \text{ and } \tilde{f} \in \mathbb{M}^D_{\vee(\wedge)} \text{ such that}\\
&\qquad \varepsilon \le f(\bar{x}) \leftrightarrow \tilde{f}(\bar{x}),\ \forall \bar{x} \in C,
\end{aligned}$$

where

$$\mathbb{M}^D_{\vee} = \Big\{ \bigvee_{\bar{d}\in D} (R(\bar{d},\bar{x}) * g(\bar{d})) |\ g \subsetsim M^{(n)} \Big\} \quad \text{and}$$
$$\mathbb{M}^D_{\wedge} = \Big\{ \bigwedge_{\bar{d}\in D} (R(\bar{x},\bar{d}) \to g(\bar{d})) |\ g \subsetsim M^{(n)} \Big\}.$$

Due to terminology introduced in [11], we approach this problem from the view point of fuzzy relations of a special structure, the so-called *normal forms*:

$$\mathrm{DNF}^P_h(\bar{x}) = \bigvee_{\bar{p}\in P} (R(\bar{p},\bar{x}) * h(\bar{p})) \quad \text{and} \tag{6}$$
$$\mathrm{CNF}^P_h(\bar{x}) = \bigwedge_{\bar{p}\in P} (R(\bar{x},\bar{p}) \to h(\bar{p})), \tag{7}$$
$$\text{where } h \subsetsim M^{(n)}, P \subseteq M^{(n)}, \bar{x} \in M^{(n)}, \tag{8}$$

are called the disjunctive and conjunctive normal forms for h w.r.t. P, respectively. Hence,

$$\mathbb{M}^D_{\vee} = \Big\{ \mathrm{DNF}^D_g |\ g \subsetsim M^{(n)} \Big\} \quad \text{and}$$
$$\mathbb{M}^D_{\wedge} = \Big\{ \mathrm{CNF}^D_g |\ g \subsetsim M^{(n)} \Big\}.$$

It is possible to rewrite inverse F-transforms using normal forms if we take $\{R_i\}_I$ and D:

- Assume $M^{(n)}$ contains sufficiently many elements to define set D.
- Take $D = \{\bar{d}_j \in M^{(n)} |\ d_{j_i} \ne d_{j'_i},\ \forall j,j' \in J, \forall i \in I\}$.
- Define $R(\bar{x},\bar{d}_j) = R(\bar{d}_j,\bar{x}) = A_{\mathbf{j}}(\bar{x}), \forall j \in J, \mathbf{j} \in J^n$ and $R(\bar{x},\bar{y}) = \mathbf{0},\ \forall \bar{y} \in M^{(n)} \setminus D$.

In the case of lacking elements to construct set D, which can hypothetically arise because there is no restriction to the number of sets A_{j_i}, we can artificially extend $M^{(n)}$ to reach the required cardinality. Then we obtain that $\mathrm{DNF}^D_{\mathrm{CNF}^C_f}$ corresponds to an inverse $F^{\downarrow}$-transform $f^{\downarrow}_{F,k}$ and $\mathrm{CNF}^D_{\mathrm{DNF}^C_f}$ to an inverse $F^{\uparrow}$-transform $f^{\uparrow}_{F,k}$.

For an arbitrary $f \subsetsim M^{(n)}$, we can certainly show[3]

$$\bigvee_{\bar{d}\in D} (R(\bar{d},\bar{x}) * f(\bar{x})) \le f(\bar{x}) \le \bigwedge_{\bar{d}\in D} (R(\bar{x},\bar{d}) \to f(\bar{x})), \forall \bar{x} \in C.$$

[3] Considering $\bigvee_{\bar{d}\in D} R(\bar{d},\bar{x}) = \bigvee_{\bar{d}\in D} R(\bar{x},\bar{d}) = \mathbf{1}$ gives the following equality $f(\bar{x}) = \bigvee_{\bar{d}\in D} (R(\bar{d},\bar{x}) * f(\bar{x})) = \bigwedge_{\bar{d}\in D} (R(\bar{x},\bar{d}) \to f(\bar{x}))$, for all $\bar{x} \in C$, because in an arbitrary residuated lattice $\mathbf{1} * a = \mathbf{1} \to a = a$, for all $a \in L$. Hence, it is reasonable to apply this requirement for the partitioning of the domain of f in the practise.

ISBN: 978-989-95079-6-8

The task stands in the replacement of $f(\bar{x})$ by suitable values so that the approximation is the best possible w.r.t. the available knowledge. In the following proposition, we are going to show that the combinations of $\{\mathrm{DNF}_f^D, \mathrm{CNF}_f^C\}$ and $\{\mathrm{CNF}_f^D, \mathrm{DNF}_f^C\}$ have this property.

Theorem 5 – DNF_f^C *is the least element of the following set:*

$$S = \{g \subsetsim M^{(n)} |\ f(\bar{x}) \leq \bigwedge_{d\in D} (R(\bar{x},\bar{d}) \to g(\bar{d})),\ \forall \bar{x} \in C\}. \tag{9}$$

– CNF_f^C *is the greatest element of the following set:*

$$S' = \{g \subsetsim M^{(n)} |\ \bigvee_{d\in D} (R(\bar{d},\bar{x}) * g(\bar{d})) \leq f(\bar{x}),\ \forall \bar{x} \in C\}. \tag{10}$$

As it has been shown above, there cannot be better lower approximation of f from the set $\mathbb{M}_\vee^D$ than $\mathrm{DNF}^D_{\mathrm{CNF}_f^C} \in \mathbb{M}_\vee^D$. And analogously, $\mathrm{CNF}^D_{\mathrm{DNF}_f^C} \in \mathbb{M}_\wedge^D$ is the best upper approximation of f. In other words, the best choice for coefficients in the inverse formula of $\mathrm{F}^{\uparrow(\downarrow)}$-transform is to take just the coefficients of the direct $\mathrm{F}^{\uparrow(\downarrow)}$-transform. Both results do not consider any boundary for an $\varepsilon-$precision of the approximations. It is dependent on the suitable choice of $\mathcal{R}_I$ and the set of nodes D arising from the following estimation in terms of a $*$-equivalence:

Theorem 6 *Let $f \subsetsim M^{(n)}$ be extensional w.r.t. $\mathcal{S}_I = \{S_i \subsetsim M_i^2\}_{i\in I}$ as follows:*

$$S_i(x,y) = \bigvee_{\bar{d}\in D} R_i(x,d_i) * R_i(d_i,y). \tag{11}$$

Then we can prove

$$\bigvee_{\bar{d}\in D} R(\bar{x},\bar{d}) * R(\bar{d},\bar{x}) \leq \begin{array}{l}\mathrm{DNF}^D_{\mathrm{CNF}_f^C}(\bar{x}) \leftrightarrow f(\bar{x}),\\ \mathrm{CNF}^D_{\mathrm{DNF}_f^C}(\bar{x}) \leftrightarrow f(\bar{x}),\end{array} \quad \forall \bar{x} \in C. \tag{12}$$

Corollary 7 *Let $\overline{\mathcal{R}}_I = \{\bar{R}_i\}_{i\in I}$ be defined by*

$$\bar{R}_i(x,y) = \bigvee_{j\in J} A_{j_i}(x) * A_{j_i}(y), \tag{13}$$

for all $i \in I$. If $f \subsetsim M^{(n)}$ is extensional w.r.t. $\overline{\mathcal{R}}_I$ then (5) is valid.

Here, we provide an alternative proof of this assertion independent of the above theorem.

Proof: Due to extensionality, we have

$$A_{\mathbf{j}}(\bar{x}) * A_{\mathbf{j}}(\bar{y}) \leq R(\bar{x},\bar{y}) = \bigvee_{\mathbf{j}\in J^n} A_{\mathbf{j}}(\bar{x}) * A_{\mathbf{j}}(\bar{y}) \leq$$
$$\leq f(\bar{x}) \leftrightarrow f(\bar{y}), \text{ and hence}$$
$$A_{\mathbf{j}}(\bar{x}) * A_{\mathbf{j}}(\bar{x}) * f(\bar{x}) \leq A_{\mathbf{j}}(\bar{x}) * (A_{\mathbf{j}}(\bar{y}) \to f(\bar{y})),$$
$$A_{\mathbf{j}}(\bar{x}) * A_{\mathbf{j}}(\bar{x}) * f(\bar{x}) \leq A_{\mathbf{j}}(\bar{x}) * \bigwedge_{\bar{y}\in C} (A_{\mathbf{j}}(\bar{y}) \to f(\bar{y})),$$
$$f(\bar{x}) * \bigvee_{\mathbf{j}\in J^n} A_{\mathbf{j}}(\bar{x}) * A_{\mathbf{j}}(\bar{x}) \leq$$
$$\leq \bigvee_{\mathbf{j}\in J^n} (A_{\mathbf{j}}(\bar{x}) * \bigwedge_{\bar{y}\in C} (A_{\mathbf{j}}(\bar{y}) \to f(\bar{y}))) = f^{\downarrow}_{F,k}(\bar{x}),$$

for all $\bar{x} \in M^{(n)}$. Adding the left side of inequality (2), we receive the first required estimation (5). Analogously we proceed to prove the inequality for $f^{\downarrow}_{F,k}$. QED

Hence, to obtain the desired ε-approximation[4], it only remains to choose D w.r.t. $\mathcal{R}_I$ so that

$$(\forall \bar{x} \in C)(\varepsilon \leq \bigvee_{\bar{d}\in D} R(\bar{x},\bar{d}) * R(\bar{d},\bar{x})),$$

which completes the task.

Corollary 8 *Let $\varepsilon \in L, D \subseteq M^{(n)}$ and $f \subsetsim M^{(n)}$ be extensional w.r.t. $\mathcal{S}_I$ defined by (11). If for each $\bar{x} \in C$ there exists $d \in D:\ \varepsilon \leq R(\bar{x},d) * R(d,\bar{x})$ then*

$$\varepsilon \leq \begin{array}{l}\mathrm{DNF}^D_{\mathrm{CNF}_f^C}(\bar{x}) \leftrightarrow f(\bar{x}),\\ \mathrm{CNF}^D_{\mathrm{DNF}_f^C}(\bar{x}) \leftrightarrow f(\bar{x}),\end{array} \quad \forall \bar{x} \in C. \tag{14}$$

Corollary 9 *Let $\varepsilon \in L$ and $f \subsetsim M^{(n)}$ be extensional w.r.t. $\bar{\mathcal{R}}_I$ defined by (13). If for each $\bar{x} \in C$ there exists $\mathbf{j} \in J^n$: $\varepsilon \leq A_{\mathbf{j}}(x) * A_{\mathbf{j}}(x)$ then*

$$\varepsilon \leq f^{\downarrow(\uparrow)}_{F,k}(\bar{x}) \leftrightarrow f(\bar{x}),\ \forall \bar{x} \in C. \tag{15}$$

Corollary 10 *Let $L = [0,1]$, $*$ be a continuous t-norm, $\varepsilon \in L$, M be a pre-compact (totally bounded) subset over a standard metric space, $R \subsetsim M$ be reflexive and uniformly continuous. If $f \subsetsim M$ is extensional w.r.t. R then for every $\varepsilon < 1$ there exists a finite $D \subseteq M$ such that (14) is valid.*

Whenever we assume each $R \in \mathcal{R}$ to be a $*$-equivalence, we obtain an estimation by means of the following pseudometrics:

$$d_n(\bar{x},\bar{y}) = \varphi^{-1}R_1(x_1,y_1) \oplus_\varphi \ldots \oplus_\varphi \varphi^{-1}R_n(x_n,y_n)$$
$$= \bigoplus_{i\in I} \varphi^{-1}R_i(x_i,y_i),\ \forall \bar{x},\bar{y} \in M^{(n)},$$
$$d(x,y) = \varphi^{-1}(x \leftrightarrow y),\ \forall x,y \in L$$

where φ is an order-reversing bijection on L, $\oplus_\varphi$ is an operation defined by $x \oplus_\varphi y = \varphi^{-1}(\varphi(x) * \varphi(y))$.

Corollary 11 *Let each $R \in \mathcal{R}_I$ be a $*$-equivalence and $d_n : M^{(n)} \times M^{(n)} \to L, d : L \times L \to L$ be defined as above. If $f \subsetsim M^{(n)}$ is extensional w.r.t. $\mathcal{S}_I$ defined by (11) then f is 1-Lipschitz continuous w.r.t. (d,d_n), i.e.,*

$$d(f(\bar{x}),f(\bar{y})) \leq d_n(\bar{x},\bar{y}),\ \textit{for all } \bar{x},\bar{y} \in M^{(n)},$$

and

$$\begin{array}{l} d(\mathrm{DNF}^D_{\mathrm{CNF}_f^C}(\bar{x}), f(\bar{x}))\\ d(\mathrm{CNF}^D_{\mathrm{DNF}_f^C}(\bar{x}), f(\bar{x}))\end{array} \leq \bigwedge_{\bar{d}\in D} 2d_n(\bar{x},\bar{d})$$
$$= \bigwedge_{\bar{d}\in D} d_n(\bar{x},\bar{d}) \oplus d_n(\bar{d},\bar{x}),\ \forall \bar{x} \in C. \tag{16}$$

[4]Notice that the choice of $\mathcal{R}_I$ is the limitation for ε as well as for D, e.g. D might become very huge if ε is close to $\mathbf{1}$.

ISBN: 978-989-95079-6-8

Let us consider a simple problem to illustrate the way of approximation by F-transforms.

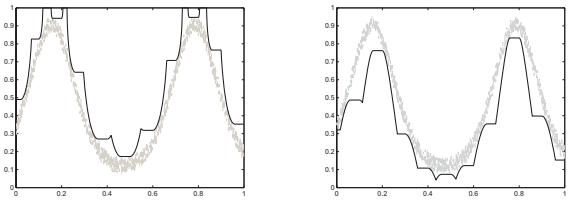
Example 12 *Let us take the following function*

$$f(x) = \frac{1}{3} \cdot e^{\sin(10x)} + Rand(x),\ \forall x \in M = [0,1],$$

where $Rand(x)$ is some additive random noise. Moreover, assume $*$ to be the t-norm generated by $g(x) = (1-x)^2$ and $\mathcal{L}$ be the associated residuated lattice. Take

$$R(x,y) = 1 \wedge (0 \vee (1.2 - 10 \cdot |x-y|)),$$

and create the approximation using 11 and 21 nodes and equidistant discretization $C = \{k \cdot 0.001 |\ k = 1, \ldots, 999\}$. The resulting approximations are depicted on Fig. 1, where f is represented by grey line and inverse $F^{\uparrow(\downarrow)}$-transforms by black lines, set D contains 11 nodes and set D' contains 21.



(a) $\mathrm{CNF}^{D}_{\mathrm{DNF}^{C}_{f}}$ (b) $\mathrm{DNF}^{D}_{\mathrm{CNF}^{C}_{f}}$

(c) $\mathrm{CNF}^{D'}_{\mathrm{DNF}^{C}_{f}}$ (d) $\mathrm{DNF}^{D'}_{\mathrm{CNF}^{C}_{f}}$

Figure 1: Approximation of f (grey line) from Example 12 using F-transforms (black line) with 11 (set D) and 21 (set D') nodes.

4 Conclusions

We have shown that the only one requirement is needed to show (5):

New requirements
$\mathcal{L}$ be a residuated lattice
f be extensional w.r.t. $\{\bar{R}_i\}_{i \in I}$

While in [1], we were limited to the stronger algebraic structure, a special type of partition, $E \subseteq C$, additionally requiring normality, and extensionality of f (see Subsection 2.2). Notice that all the additional requirements follow from the procedure of replacement of fuzzy sets A_{i_i} by P_i based on the special set E (such relations can be found in [12]). We have found a way how to get around this construction and to reach such simplification.

Moreover, our approach enabled to express direct and inverse $F^{\uparrow(\downarrow)}$-transforms using combinations of normal forms. This allowed us to show an optimality of choice of all coefficients of the direct $F^{\uparrow(\downarrow)}$-transform in the inverse formula at once (see Theorem 5). Additionally, we gave some interesting corollaries that contributes only to the theory of normal forms and they cannot be translated into the theory of $F^{\uparrow(\downarrow)}$-transforms, which shows that even though both theories have a lot in common, it is also important to study them separately.

Acknowledgment

We gratefully acknowledge partial support of the projects MSM6198898701 and 1M0572 of the MŠMT ČR.

References

[1] Irina Perfilieva. Fuzzy transforms: Theory and applications. *Fuzzy Sets and Systems*, 157:992–1023, 2006.

[2] Irina Perfilieva and Dagmar Plšková. Solving ode with fuzzy initial condition using fuzzy transform. In *Proc. of 10'th Czech-Japan Seminar on data Analysis and Decision Making under Uncertainty*, pages 130–139, Prague, 2007. Oeconomica.

[3] Martin Štěpnička and Radek Valášek. Numerical solution of partial differential equations with help of fuzzy transform. In *Proceedings of the FUZZ-IEEE 2005*, pages 1104–1109, Reno, Nevada, 2005. Fuzz-IEEE 2005.

[4] Irina Perfilieva and Radek Valášek. Fuzzy approach to data compression. In *Proc. of 8'th Czech-Japan Seminar on data Analysis and Decision Making under Uncertainty*, pages 91–100, Prague, 2005. Oeconomica.

[5] Irina Perfilieva and Martina Daňková. Image fusion on the basis of fuzzy transforms. In *Computational Intelligence in Decision and Control*, pages 471–476, New Jersey, 2008. World Scientific.

[6] Martina Daňková and Radek Valášek. Full fuzzy transform and the problem of image fusion. *Journal of Electrical Engineering*, 12:82–84, 2006.

[7] Vilém Novák, Irina Perfilieva, Antonín Dvořák, G. Chen, Q. Wei, and P. Yan. Fuzzy transform in the analysis of data. *International Journal of Approximate Reasoning*, 48:4–22, 2008.

[8] Martina Daňková. On approximate reasoning with graded rules. *Fuzzy Sets and Systems*, 158:652–673, 2007.

[9] Martina Daňková. Approximation of extensional fuzzy relations over residuated lattices. Submitted.

[10] Llorenç Valverde. On the structure of F-indistinguishability operators. *Fuzzy Sets and Systems*, 17(3):313–328, 1985.

[11] Irina Perfilieva. Normal forms in BL-algebra of functions and their contribution to universal approximation of functions. *Fuzzy Sets and Systems*, 143:111–127, 2004.

[12] Frank Klawonn. Fuzzy points, fuzzy relations and fuzzy functions. In V. Novák and I. Perfilieva, editors, *Discovering the World with Fuzzy Logic*, pages 431–453. Physica-Verlag, Heidelberg, 2000.

ISBN: 978-989-95079-6-8

An Interval Type-2 Fuzzy Distribution Network

Simon M. Miller, Viara Popova, Robert John, Mario Gongora

Centre for Computational Intelligence, De Montfort University, Leicester, UK.
Email: {smiller,vpopova,rij,mgongora}@dmu.ac.uk

Abstract— *Planning resources for a supply chain is a major factor determining its success or failure. In this paper we introduce an* ***Interval Type-2 Fuzzy Logic model*** *of a distribution network. It is believed that the additional degree of uncertainty provided by Interval Type-2 Fuzzy Logic will allow for better representation of the uncertainty and vagueness present in resource planning models. First, the subject of* ***Supply Chain Management*** *is introduced, then some background is given on related work using Type-1 Fuzzy Logic. A description of the Interval Type-2 Fuzzy model is given, and a test scenario detailed. A* ***Genetic Algorithm*** *uses the model to search for a near-optimal plan for the scenario. A discussion of the results follows, along with conclusions and details of intended further work.*

Keywords— Distribution model, Evolutionary Computing, Interval Type-2 Fuzzy Logic, Resource Planning, Supply Chain Management

1 Introduction

There are a number of definitions of Supply Chain Management (SCM), each having minor variances but describing the same core idea. SCM is the management of material flow in and between facilities including vendors, manufacturing/assembly plants and distribution centres [1]. Planning the allocation of resources within a Supply Chain (SC) has been critical to the success of manufacturers, warehouses and retailers for many years. Mastering the flow of materials from their creation to the point of sale offers considerable advantages to those within a well managed SC. Poorly managed resources result in two main problems: stock outs and surplus stock.

The consequence of stock outs is lost sales, and potentially lost customers. Surplus stock causes additional holding cost and the possibility of stock losing value as it becomes obsolete. Holding some surplus stock is advantageous however; safety stock can be used in the event of an unexpected increase in demand or to cover lost productivity.

The problem has been addressed numerous times in the literature using different approaches. An overview of traditional and Computational Intelligence approaches to supply-chain resource planning can be found in [2]. Here a novel approach using Type-2 Fuzzy Logic (T2FL) [3] is reported together with optimisation by means of a Genetic Algorithm (GA)[4].

This research is part of a project on demand forecasting and resource planning (Data Storage, Management, Retrieval and Analysis: Improving Customer Demand and Cost Forecasting Methods funded by the Technology Strategy Board in the UK). The project aims to: (1) improve the forecasting of demand by using a variety of disparate sources and statistical and Machine Learning (ML) methods for analysis; (2) improve the allocation of resources in which the generated forecast is used as an input and the output is a (long- or short-term) plan of raw materials and resources within the supply chain in order to meet such demand. Various degrees of uncertainty are present in the different data sources used that are amplified in the generated forecast by applying methods of analysis with (again) varying degrees of inherent uncertainty. Furthermore, other data that is often used in resource planning such as transportation and other costs, customer satisfaction information, etc. is also uncertain. Therefore, FL and especially T2FL are particularly appropriate for this problem. While Type-1 FL (T1FL) has successfully been used many times for modelling SC operation (see Section 2), T2FL has been shown to offer a better representation of uncertainty on a number of problems (e.g., [5] and [6]). However, to date T2FL has not been applied to SCM in the literature. This paper presents a T2FL model of a SC problem which is optimised by a GA to find a near-optimal configuration.

The paper is organised as follows. Section 2 discusses Fuzzy Logic and its application to SCM. Section 3 introduces the model and the test case used for evaluation. The results from the experiments on the test case are presented in Section 4. Section 5 concludes the paper with a summary and future research directions.

2 Fuzzy Logic for Supply Chain Management

FL and GAs have been successfully used for supply chain modelling [2] and are particularly appropriate for this problem due to their capacity to tackle the inherent vagueness, uncertainty and incompleteness of the data used. A GA [4] is a heuristics search technique inspired by evolutionary biology. Selection, crossover and mutation are applied to a population of individuals representing solutions in order to find a near-optimal solution. FL is based on fuzzy set theory and provides methods for modelling and reasoning under uncertainty, a characteristic present in many problems, which makes FL a valuable approach. It allows data to be represented in intuitive linguistic categories instead of using precise (crisp) numbers which might not be known, necessary or in general may be too restrictive. For example, statements such as *'the cost is about n'*, *'the speed is high'* and *'the book is very old'* can be described. These categories are represented by means of a membership function which defines the degree to which a crisp number belongs to the category. In this research the aim is to allow linguistic terms to be used by Supply Chain Managers when describing their operation. For example, instead of asking for crisp numbers to describe the current stock level of a product, we may allow them to make statements like 'Warehouse 1 has *About 500* of product 1' or even more abstract 'Warehouse 1's stock level of product 1 is *low*'. By removing the need for exact information it is possible to produce a system that is much more usable when the information available is vague, uncertain or incomplete.

ISBN: 978-989-95079-6-8

T1FL has been applied to SC modelling numerous times with good results. Some of the research that is considered most relevant for this paper is discussed in the following paragraphs.

Petrovic et al. [7] use fuzzy sets to model vagueness and imprecision in customer demand, external supplier reliability and supply within a SC. The system demonstrates the effect of differing conditions and strategies on fill rate and holding costs of a SC. The results of the evaluation show that there is a slight improvement in the performance of the SC when the inventories are partially co-ordinated. In [8] Petrovic et al. employ two-level fuzzy optimisation to find ideal order-up-to quantities in a one warehouse-multiple retailer SC. The problem is decomposed for individual control of the warehouse and retailers; a co-ordination mechanism provides overall control of the SC. Customer demand, inventory levels, holding cost, and shortage cost are represented by fuzzy sets. A measure of satisfaction is derived from the cost incurred at each element. The solution produced by the system is the best compromise between members, though not necessarily the cheapest.

In [9] a fuzzy system is used with a GA to model a SC. The model uses a global policy of management with emphasis on integrating the production and distribution models. The GA searches for a near-optimal configuration; fuzzy sets are used to describe costs, returns, production capacities, storage capacities and forecasts. The proposed fuzzy method, a crisp method and a non-integrated method are compared. The crisp system is unable to produce a feasible configuration if actual demand is lower than the forecast. In contrast, the fuzzy model presented is robust and able to cope with fluctuation in demand and production capacity with little impact on profitability. The non-integrated model performed significantly worse than the fuzzy integrated model.

A similar approach is presented by Wang and Shu [10]. FL is used to represent customer demand, processing time and delivery reliability; a GA finds order-up-to levels. The system attempts to find the configuration that incurs the minimum cost. An optimism-pessimism index is set by the user and passed to the system. When optimistic, the model assumes the best case scenario for material response time. A pessimistic attitude produces the opposite effect. The results show that more pessimistic strategies increase the fill rate, reducing the sales lost through stock outs, and incur higher inventory cost as more stock is kept. More optimistic strategies result in a drop in fill rate and an increase in sales loss, though inventory cost is also reduced.

T2FL [3] provides the means to model an additional dimension of uncertainty as the membership functions are themselves modelled as fuzzy numbers, thus providing the means for a more accurate representation of uncertainty in a complex system and the potential for better performance. On a number of problems, T2FL has been shown to outperform T1FL (e.g., [5] and [6]), however it has not yet been applied to SC modelling in the literature. In this paper, Interval Type-2 FL (IT2FL) [11] is used which is computationally cheaper as it restricts the additional dimension, referred to as secondary membership function, to only take the values 0 or 1. We believe that the extra degree of freedom will allow a better representation of the uncertain and vague nature of data used in SCM. Section 3 describes the created model in detail as well as the test case used for evaluation.

3 Model

The model presented here describes a distribution network consisting of a set of manufacturing warehouses and customers. Each warehouse holds one or more products; when demand of an item is forecast for a customer, a warehouse is selected to satisfy the demand. Before the model is used it is configured to model the required network. This is done by telling the model how many warehouses, customers and products there are, the distances between the warehouses and customers, the costs associated with network operation and how many periods the model is to run for. As input, the model takes a forecast of demand over a given period and a suggested resource plan for the warehouses detailing how much of each item to hold at each warehouse in each period.

The purpose of the model is to calculate the cost of a given resource plan. This is calculated based on a number of factors:

Production Cost - Each product is assigned an individual production cost. The total production cost for each batch is calculated by multiplying the number of items by their production cost.

Holding Cost - A holding cost is charged if a product is kept in a warehouse for more than one period. The cost is calculated by taking a specified percentage of the purchase price of the goods held, for items carried over from one period to the next. The purpose of the charge is to represent the costs of storing items, depreciating value of stock and the losses incurred by tying up capital in unsold stock.

Transport Cost - The cost of transporting goods is produced using a matrix of distances between warehouses and customers, and a list of transport costs per mile for each product. The product of the relevant cost and mileage gives the overall transport cost for a batch of product.

Stock out cost - Stock out is the shortfall of a product in a particular period. In this model we make the assumption that the customer is always provided with an item. If it is not in the warehouse, it is purchased at full retail price from a competing producer. The stock out cost is the sum of the value of items that had to be purchased in this period. Cost of purchase is not the only penalty incurred when stock out occurs. The customer may cancel their order if they discover that it is not in stock, and will be ordered elsewhere. To represent this a multiplier is applied to the stock out cost, taking into account the cost of lost orders.

Batch cost - The cost of setting up an order is called the batch cost. This represents the cost of administration and setting up any machines that are required, and picking the items for dispatch. There is a flat fee for each batch which is charged once at each warehouse for the production of a particular item for a particular customer.

Using these costs a total cost is produced, allowing comparison of competing solutions.

3.1 Interval Type 2 Fuzzy Logic

For the experimental model IT2FL has been used to represent some of the values within the model. Previous examples (as

ISBN: 978-989-95079-6-8

discussed in Section 2) of research in this area have focused on the use of T1FL; we believe there exists an opportunity to exploit the extra degree of uncertainty provided by IT2FL in a model of this type.

As the model operates on fuzzy numbers, fuzzy arithmetic is used to calculate costs. This involves taking fuzzy sets, discretising them, performing the arithmetic operation, and then reconstructing the fuzzy set. In this model, fuzzy sets are represented using a series of α-cuts. Each set is an array of pairs of intervals. Each pair shows the area of values covered at a particular value of μ, the first interval is the left hand side of the set, and the second the right. Storing the sets in this way removes the need to discretise before fuzzy arithmetic is performed, and then reconstruct the result. Operations on fuzzy sets are performed at interval level, corresponding intervals (at the same μ) are taken from two sets, the operation performed and the result stored in a third fuzzy set.

The following values are represented by IT2 fuzzy numbers: forecast demand, inventory level, transportation distances, transportation cost, stock out level, stock out cost, carry over and holding cost. For each of these values we can use the linguistic term *'about n'*, e.g., forecast demand of product 1 for customer 1 in period 1 may be *'about 200'*. When the fuzzy total cost for a solution has been calculated, it is defuzzified to produce a crisp cost value. Fig. 1 shows how the set *'about 200'* may look with the α-cut representation used, where x is the scale of values being represented. The set is described using a collection of pairs of intervals. Each pairing represents the left and right hand intervals of the set for a given value of μ. As stated before, representing the set in this way considerably simplifies the arithmetic operators that are applied as simple interval arithmetic is used throughout, without discretisation or reconstruction.

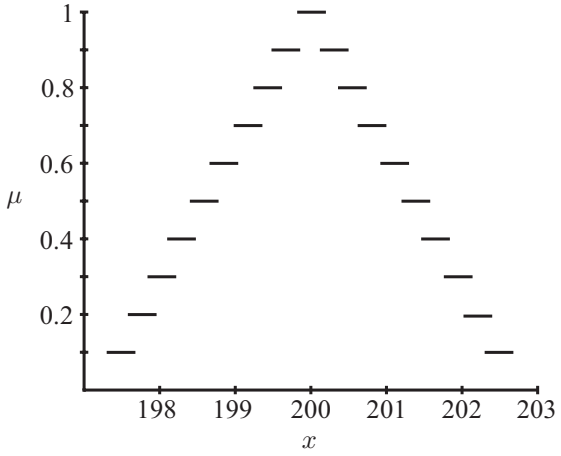


Figure 1: Interval representation of IT2 fuzzy set *'about 200'*

3.2 Defuzzification

The defuzzification process used is described in (1) where n is equal to the number of α-cuts, and l and r denote the left and right hand intervals respectively. To reference the left and right endpoints of the intervals $_1$ and $_2$ are used; so l_1 refers to the left bound of l and l_2 the right bound. M denotes the mid-point of an interval, in the case of M_i^x this means the centres of the intervals in x described by the i^{th} pair of intervals. For M^{α_i} we take the mid-point between the current and subsequent α-cut's μ values. The mid-point values are used along with the height (H) and length (L) to calculate the weighted area of each pair of intervals, giving more weight the those with higher μ values.

$$\frac{\sum_{i=0}^{n-1} M_i^x H_i L_i M^{\alpha_i}}{\sum_{i=0}^{n-1} H_i L_i M^{\alpha_i}} \tag{1}$$

where:

$$M_i^x = \begin{cases} \frac{x_{l_2}^{\alpha_i} + x_{l_1}^{\alpha_i}}{2} & \text{for left interval} \\ \frac{x_{r_2}^{\alpha_i} + x_{r_1}^{\alpha_i}}{2} & \text{for right interval} \end{cases} \tag{2}$$

$$H_i = \alpha_{i+1} - \alpha_i \tag{3}$$

$$L_i = \begin{cases} x_{l_2}^{\alpha_i} - x_{l_1}^{\alpha_i} & \text{for left interval} \\ x_{r_2}^{\alpha_i} - x_{r_1}^{\alpha_i} & \text{for right interval} \end{cases} \tag{4}$$

$$M^{\alpha_i} = \frac{\alpha_i + \alpha_{i+1}}{2} \tag{5}$$

In the test scenario, μ is discretised in a uniform manner. This removes the need to calculate the height of intervals as they are all the same, and therefore cancel out. The process used in the actual code is given in (6).

$$\frac{\sum_{i=0}^{n-1} M_i^x L_i M^{\alpha_i}}{\sum_{i=0}^{n-1} L_i M^{\alpha_i}} \tag{6}$$

It should be pointed out that while this method works well for the convex fuzzy numbers used in this system, it cannot be applied to non-convex fuzzy sets as the interval representation employed will cease to accurately represent the set. In this model, the fuzzy numbers will always be convex. In addition to this the method of representation being used ensures that the intervals are already present, no further discretisation is required. Using a method that requires re-discretisation would incur unnecessary computation. A consideration for future work is the possibility of using another approach such as the one proposed by Karnik and Mendel [12], however this method would involve rediscretising on the x axis, at present the model is discretised on the μ axis.

3.3 Optimisation

The focus of the experiment described here is the validation of an IT2FL model that has been constructed. This is to be achieved by using the model to find a good resource plan for a

ISBN: 978-989-95079-6-8

given forecast, confirming that the model can be used to evaluate potential solutions. A GA has been chosen for this purpose. GAs have been used successfully in previous work (e.g., [13] and [14]) to find good solutions with T1FL models. GAs are useful when a search area is too large to allow evaluation of every solution. In this case the GA is used to search for a resource plan that incurs the minimum cost. The GA has a population of 250 and is executed for 500 generations, in all 125000 solutions are evaluated. Even in this relatively small problem the GA only covers approximately $\frac{1}{10^{62}}$ of the total search space. New generations consist of: 1% individuals produced with elitism, 20% copied individuals, 20% individuals created with single point crossover and 59% of individuals created using mutation. A description of the chromosome, operators and processes employed follows.

Chromosome - The chromosome used to describe potential solutions is a 3-D matrix of inventory levels. Columns represent customers, and rows represent a product at a warehouse. If there are 3 warehouses and 5 products this will result in 15 rows. The first 5 rows will be products 1 to 5 for warehouse 1, and second 5 for warehouse 2 and so on. The third dimension represents the period, in each period there is a complete inventory plan for all warehouses and customers.

Initial population - The initial population is randomly generated. Each element of the resource plan matrix can be a number between 0 and 500 in steps of 100. This has been done to reflect the fact that in industry, products are usually manufactured in round quantities. If the model suggests that a warehouse should make 102 of product 1, this could lead to difficulties and extra expense. Limiting the valid inventory numbers also has the side effect of reducing the search space.

Fitness evaluation - Fitness is evaluated using the IT2FL model described. Candidate solutions are given to the model which evaluates them, and returns the cost. The cheaper a solution is, the fitter it is judged to be. In reality, cost may not be the only factor in deciding how much of a product to stock at each warehouse. Other criteria such as customer service level can also be used to prevent the system from choosing solutions that do not meet service requirements. Customer service level could be calculated by looking at the percentage of orders fulfilled completely, or orders fulfilled on time.

Selection - Selection is performed using a fitness ranking proportionate method similar to *roulette wheel selection*. First all solutions in the population are ranked by fitness. They are then given a number of elements of an array in proportion to their fitness ranking. For example, if we have a population of 250 the fittest individual would be allocated 250 elements in the array, the second fittest 249 and so on. An element of the array is then selected at random, and the identification number of the individual it contains is used to retrieve a parent. This tombola style approach ensures that it is possible for any individual to be selected, while weighting in favour of those with greater fitness.

Crossover - Crossover is achieved with a single point crossover. Two parents are selected using the method of selection described. Then, a new individual is created with the first half of the first parent and the second half of the second parent. Division is done by rows, if we have 5 nodes with 4 products we would have 20 rows; in this case the first 10 rows of the first parent and the final 10 of the second parent would be used to create a new individual. If the example were to span 6 periods, we would take the first 10 rows for all 6 periods of the first parent, and the final 10 rows for all 6 periods from the second to create our new individual.

Mutation - To create a mutated individual, a parent is selected, then one of the elements of its resource plan is randomly replaced with another valid value to create a new child.

3.4 Test Scenario

The test scenario is as follows:

The distribution network consists of 3 manufacturing warehouses located in Bristol (W1), Manchester (W2) and Edinburgh (W3) and 6 customers in Plymouth (C1), Cardiff (C2), Derby (C3), Liverpool (C4), Newcastle (C5) and Aberdeen (C6). Each manufacturing warehouse holds the same 5 products, and may supply any customer. The model will be executed over one period and holding costs are 10% of the purchase price of an item per period. Stock outs will be charged at 5 times the cost of purchasing an item, per item short. Each batch of product produced will incur a batch cost of £100. Table 1 details the costs associated with each product.

Table 1: Product costs (£s)

Product	1	2	3	4	5
Purchase price (per item)	3	4	2	3	3
Production cost (per item)	2	3	1	2	2
Transport cost (per item/per mile)	1	2	1	3	1

The forecast being used states that each customer would like 200 of each product. The location of the manufacturing warehouses and customers have been arranged so that a good solution is easily identifiable. Each warehouse has 2 customers that are best served by it. Table 2 shows how an ideal plan should look. In this case the problem is intentionally kept relatively simple, making it easy to evaluate the found solution and see if it is near the optimal solution. Clearly, simple problems do not require such a system in order to solve them. The work shown here is an initial experiment, future work will build upon this so that it can be applied to larger non-trivial problems where the benefit from the system will be greater.

The next section presents and interprets the results from the experiments with the GA.

4 Results

As stated previously, in each test the GA was applied to the model using 250 individuals over 500 generations. The test was repeated 10 times with differing random seeds, the best solution found in each test was recorded. Table 3 shows the costs incurred by the best solution found in each test. For reference, the cost of the ideal solution presented is £19178.26.

Using the IT2FL model the GA is able to find solutions that are close to the optimum. The best solution found is just £264.76 (1.38%) from the ideal solution, and the mean difference between solutions found and the ideal solution is £1344.79 (7.01%). Table 4 shows the best solution found (with random seed 1). Comparing this solution with the ideal

ISBN: 978-989-95079-6-8

Table 2: Ideal solution

Ware.	Prod.	C1	C2	C3	C4	C5	C6
W1	1	200	200	0	0	0	0
W1	2	200	200	0	0	0	0
W1	3	200	200	0	0	0	0
W1	4	200	200	0	0	0	0
W1	5	200	200	0	0	0	0
W2	1	0	0	200	200	0	0
W2	2	0	0	200	200	0	0
W2	3	0	0	200	200	0	0
W2	4	0	0	200	200	0	0
W2	5	0	0	200	200	0	0
W3	1	0	0	0	0	200	200
W3	2	0	0	0	0	200	200
W3	3	0	0	0	0	200	200
W3	4	0	0	0	0	200	200
W3	5	0	0	0	0	200	200

Table 3: Results of test - Best fitness

Seed	Best Fitness (£s)
0	19789.32
1	19443.02
2	20860.01
3	21048.27
4	20407.71
5	20854.52
6	20952.15
7	20950.17
8	20229.67
9	20695.66

solution we can see that not only are the plans found cost-effective, but also represent a sensible allocation of resources. The found solution is very close to the ideal solution, and where there are differences the found model is producing the correct amount of product, but a different warehouse has been chosen. This would not be the case if the model was inadequate for the problem and did not provide a clear distinction between good and bad solutions. All customers will receive their order, the only negative point is that some of customer 4 and 5's orders could be supplied by a closer warehouse.

When the program begins, the best individual in the first generation tends to cost around £70000. Fig. 2 shows a typical run, in this case for the test with a random seed of 1. Evolution can clearly be seen to be heading toward the ideal solution, we also know what the perfect solution looks like, and that the solutions found are close to it. This suggests that the model is working correctly. If the model were not correct, evolution may occur, but possibly in an unexpected direction. This could result in cheap solutions that are not sensible, or in an inability to find cost-effective solutions at all.

Out of interest, an extra test was run to see whether the GA would find a better solution if it were left to run over more generations. The best seed from the previous tests (random seed 1) was used as a starting point, and the GA was allowed to run for 2500 generations. In generation 2141 the ideal solution was found, proving that using only feedback from the model, the GA is able to find the ideal solution.

Table 4: Solution found with a random seed of 1

Ware.	Prod.	C1	C2	C3	C4	C5	C6
W1	1	200	200	0	200	0	0
W1	2	200	200	0	0	0	0
W1	3	200	200	0	0	0	0
W1	4	200	200	0	0	0	0
W1	5	200	200	0	0	0	0
W2	1	0	0	200	0	200	0
W2	2	0	0	200	200	0	0
W2	3	0	0	200	200	200	0
W2	4	0	0	200	200	0	0
W2	5	0	0	200	200	200	0
W3	1	0	0	0	0	0	200
W3	2	0	0	0	0	200	200
W3	3	0	0	0	0	0	200
W3	4	0	0	0	0	200	200
W3	5	0	0	0	0	0	200

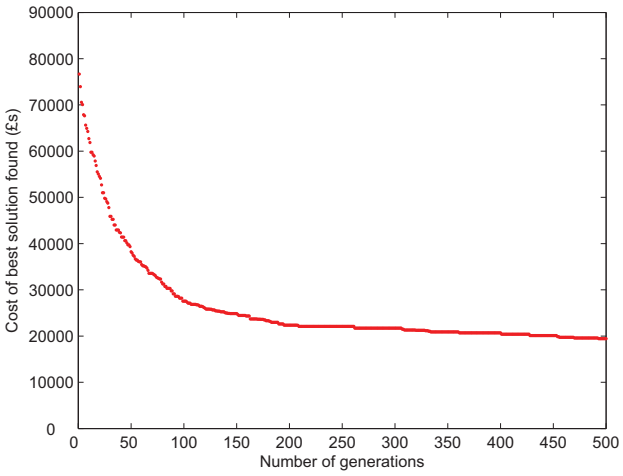


Figure 2: Evolution of best fitness

5 Conclusion

In this paper we have presented an IT2FL model that can be used to model a distribution network. The advantage of this model is that IT2FL accounts for the uncertain nature of the real world, allowing inputs to the model to be vague. IT2FL has been chosen as it is believed that the extra level of uncertainty over T1FL offered will benefit a model of this type, while avoiding the computational complexity of a Generalised Type-2 FL model.

Using the model, it was shown that a GA was able to find good resource plans that were both cost-effective and sensible. Over a longer duration, the GA was able to find the ideal solution. As a GA is solely guided by fitness and essentially 'blind' to the practicality of the solutions it finds, this was taken as indication of the model's validity.

There is much further work to be done with this model in order to make it useful in real world applications. Initially, the model will be extended for use in cases where an ideal result has been calculated by other means (e.g., a statistical model) so that performance comparison is possible. In these cases we would be looking for both accuracy of costings provided,

ISBN: 978-989-95079-6-8

and the ability of an optimisation algorithm to find good solutions. As well as being larger problems, these case studies may also call for more than just cost as a measure of fitness. As discussed in the paper, this will mean adding in a measure of customer satisfaction such as fill rate to ensure that good plans are those that are not only cost-effective, but also help to retain customers.

The current model does not take into account the temporal nature of manufacture and supply. Another dimension that is to be considered is how *lead times* affect the operation of the network. Lead times are another attribute that could benefit from an IT2FL representation, as we can never be certain when a product will arrive, or exactly how long it will take to produce. Part of this will include representing capacity, as at present only inventory is accounted for by the model. This will allow the optimising algorithm to suggest not only stock levels, but also capacity levels within a factory. Capacities can then be used to provide some indication of how long we should expect to wait for a given batch of product to be manufactured.

Acknowledgment

The research reported here has been funded by the Technology Strategy Board (Grant No. H0254E).

References

[1] D.J. Thomas and P.M. Griffin. Coordinated supply chain management. *European Journal of Operational Research*, 94(1):1–15, October 1996.

[2] S.M. Miller, V. Popova, R. John, and M. Gongora. Improving resource planning with soft computing techniques. In *Proceedings of UKCI 2008, De Montfort University, Leicester, UK.*, September 2008. available at: http://www.cci.dmu.ac.uk/preprintPDF/SimonUKCI(2).pdf.

[3] J.M. Mendel and R.I.B. John. Type-2 fuzzy sets made simple. *IEEE Transactions on Fuzzy Systems*, 10(2):117–127, April 2002.

[4] J.H. Holland. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. Ann Arbor: University of Michigan, 1975.

[5] H.A. Hagras. A hierarchical type-2 fuzzy logic control architecture for autonomous mobile robots. *IEEE Transactions on Fuzzy Systems*, 12(4):524–539, August 2004.

[6] N.N. Karnik and J.M. Mendel. Applications of type-2 fuzzy logic systems to forecasting of time-series. *Information Sciences*, 120:89–111, 1999.

[7] D. Petrovic, R. Roy, and R. Petrovic. Supply chain modelling using fuzzy sets. *International Journal of Production Economics*, 59:443–453, 1999.

[8] D. Petrovic, Y. Xie, K. Burnham, and R. Petrovic. Coordinated control of distribution supply chains in the presence of fuzzy customer demand. *European Journal of Operational Research*, 185:146–158, 2008.

[9] R.A. Aliev, B. Fazlollahi, B.G. Guirimov, and R. R. Aliev. Fuzzy-genetic approach to aggregate production-distribution planning in supply chain management. *Information Sciences*, 177:4241–4255, 2007.

[10] J. Wang and Y-F. Shu. Fuzzy decision modelling for supply chain management. *Fuzzy Sets and Systems*, 150(1):107–127, 2005.

[11] J.M. Mendel, R.I. John, and F. Liu. Interval type-2 fuzzy logic systems made simple. *IEEE Transactions on Fuzzy Systems*, 14(6):808–821, December 2006.

[12] N. Karnik and J. Mendel. Centroid of a type-2 fuzzy set. *Information Sciences*, 132:195–220, 2001.

[13] M. Sakawa and T. Mori. An efficient genetic algorithm for job-shop scheduling problems with fuzzy processing time and fuzzy duedate. *Computers & Industrial Engineering*, 36:325–341, 1999.

[14] C. Fayad and S. Petrovic. A fuzzy genetic algorithm for real-world job shop scheduling. In *Proceedings of the 18th international conference on Innovations in Applied Artificial Intelligence. Bari, Italy.*, pages 524–533, 2005.

ISBN: 978-989-95079-6-8

Fuzzy differential equation with π-derivative

Y. Chalco-Cano[1] H. Román-Flores[2] M.D. Jiménez-Gamero[3]

1.Departamento de Matemática, Universidad de Tarapacá
Arica, Chile
2.Instituto de Alta Investigación, Universidad de Tarapacá
Arica, Chile
3.Departamento de Estadística e Investigación Operativa Facultad de Matemáticas, Universidad de Sevilla
Sevilla, Spain
Email: {ychalco,hroman}@uta.cl, dolores@us.es

***Abstract*— *The π-derivative for fuzzy function is considered. Consequently, using this π-derivative we study fuzzy differentuial equations. In particular, we build a solution for a fuzzy differential equation with the help of a system of ordinary differential equations which is generates of the π-derivative.*

***Keywords*—** π-derivative for set valued functions, derivative for fuzzy functions, fuzzy differential equations.

1 Introduction

The spaces $\mathcal{K}$, of all closed and bounded intervals of $\mathbb{R}$ is not linear spaces since it do not contain inverse elements for the addition and therefore subtraction is not well defined. As a consequence, alternative formulations for subtraction have been suggested. One of those alternatives are the H-diference [3, 13] and by embedding the space $\mathcal{K}$ in a linear space [22].

The Radström's embedding theorem [22] tell us that there is a real normed linear space $\mathcal{B}$ and an isometric mapping $\pi : \mathcal{K} \to \mathcal{B}$. Then, taking advantage of this embedding theorem, a set-valued mappings F is π-differentiable at t_0 if $\pi \circ F$ is differentiable at t_0. In [3], we can find various properties of this derivative and its connection with other definitions of derivatives for set-valued mappings. On the other hand, in [20] each element of $\mathcal{B}$ can be represented in the form (x, δ). Thus, the π-derivative of $\pi \circ F$ is associated with the derivative of an ordered pair. In [9] the authors obtained a set $F'_{\pi}(t) \in \mathcal{K}$, which is said π-derivative of F at t. Some properties and the conection with generalized derivative are studied.

The aim of this paper is introduce the π-derivative for fuzzy interval valued functions based in the generalization of π-derivative for set-valued mappings given in [9]. Consequently, using the π-derivative for fuzzy functions, we study fuzzy differential equations. In particular, we build a solution for a fuzzy differential equation with the help of a system of ordinary differential equations which is generates of the π-derivative.

2 Preliminaries

Let $\mathbb{R}$ be the 1-dimensional Euclidian space. We denote by $\mathcal{K}$ the family of all bounded intervals

$$\mathcal{K} = \{A = [a, b] \subset \mathbb{R} \mid A \neq \emptyset \text{ is bounded } \}$$

The Hausdorff metric H is defined by

$$H(A, B) = \max \{|a - c|, |b - d|\}.$$

where $A = [a, b]$ and $B = [c, d]$.

Also, by using the Minkowski sum between two sets, we defined the following operations on $\mathcal{K}$

$$A + B = \{a + b \mid a \in A,\ b \in B\} \text{ and } \lambda A = \{\lambda a \mid a \in A\}. \tag{1}$$

The spaces $\mathcal{K}$ is not linear spaces since it do not contain inverse elements and therefore subtraction is not well defined [1, 2].

We have, from (1), that $A - B = A + (-1)B$. Also, if $A = B + C$, then the Hukuhara difference of A and B is denoted by $A -_H B$ and it is equal to C. The Hukuhara difference of A and B is also called the geometrical difference between the sets A and B [27].

In [3] the authors introduce subtraction in $\mathcal{K}$ by using the Radström's embedding theorem [22] which tell us that there is a real normed linear space $\mathcal{B}$ and an isometry $\pi : \mathcal{K} \to \mathcal{B}$ such that $\pi(\mathcal{K})$ is a convex cone in $\mathcal{B}$.

To construct $\mathcal{B}$, we consider the following equivalence relation in the space $\mathcal{K} \times \mathcal{K}$

$$(A, B) \sim (C, D) \quad \text{if and only if } \ A + D = B + C.$$

If $\langle A, B\rangle$ is the equivalence class of the pair (A, B), then $\mathcal{B}$ is the quotient space $\mathcal{K} \times \mathcal{K}/\sim$. Now, in $\mathcal{B}$ we introduce the operations of addition and scalar multiplication by means

$$\langle A, B\rangle + \langle C, D\rangle = \langle A + C, B + D\rangle,$$

$$\lambda \langle A, B\rangle = \begin{cases} \langle \lambda A, \lambda B\rangle & \lambda \geq 0, \\ \langle |\lambda| B, |\lambda| A\rangle & \lambda < 0 \end{cases}$$

Then $\mathcal{B}$ is a linear space.

One defines the embedding $\pi : \mathcal{K} \to \mathcal{B}$ as follows

$$\pi(A) = \langle A, 0\rangle, \ \ A \in \mathcal{K},$$

so that $\langle A, 0\rangle$ is the equivalence class
$\{(A + D, D) \mid A, D \in \mathcal{K}\}$.

The metric and the norm in $\mathcal{B}$ are defined by

$$\rho(\langle A, B\rangle \ , \ \langle C, D\rangle) = H(A + D \ , \ C + B)$$

$$\|\langle A, B\rangle\| = \rho(\langle A, B\rangle \ , \ \langle 0, 0\rangle)$$

If $A, B \in \mathcal{K}$, then the difference of the sets A and B is an alement of $\mathcal{B}$ equal to $\langle A, B\rangle$. Since $\mathcal{B}$ is a linear space, difference has all the properties of difference in linear spaces.

ISBN: 978-989-95079-6-8

In general, the difference of two sets in $\mathcal{K}$ is not necessarily an element of $\mathcal{K}$. On the other hand if the Hukuhara diference is well defined for two sets A, B in $\mathcal{K}$, then

$$\langle A, B\rangle = \langle A -_H B, 0\rangle .$$

3 π-derivative for set-valued mappings

In this Section we present the π-derivative for set-valued mappings. This concept has been studied by the authors in [9].

Elements of the space $\mathcal{K}$ are bounded and closed intervals, therefore each equivalence class can be represented in the form

$$\langle [x, x+\delta]\ ,\ 0\rangle\ ,\ \delta \geq 0 \quad \text{or} \quad \langle 0\ , [-x, -x-\delta]\rangle\ ,\ \delta < 0.$$

We will call this the canocical representation [20].

In fact, an arbitrary element $([a,b]\ ,\ [c,d])$ belongs to a class of the first kind if $b-a \geq d-c$ (then $x = a-c,\ \delta = (b-a)-(d-c)$) and to a class of the second kind if $b-a < d-c$ (then $x = a-c,\ \delta = (b-a)-(d-c)$).

For the canonical representation of an equivalence class we will use the notation (x, δ), which implies that

$$(x,\delta) = \begin{cases} \langle [x,\ x+\delta]\ ,\ 0\rangle & \text{if} \quad \delta \geq 0 \\ \langle 0,\ [-x,\ -x-\delta]\rangle & \text{if} \quad \delta < 0. \end{cases} \tag{2}$$

Given $A = [a,b] \in \mathcal{K}$ we have

$$\pi(A) = \pi([a,b]) = \langle [a,\ b]\ ,\ 0\rangle = \langle [x, x+\delta], 0\rangle ,$$

where $x = a$ and $\delta = b-a$.

Theorem 1 *([20]) The following results hold*

(a) $(x,\delta) = (y,\beta) \Longleftrightarrow x = y,\ \delta = \beta$;

(b) $(x,\delta) + (y,\beta) = (x+y, \delta+\beta)$;

(c) $a(x,\delta) = (ax, a\delta)$ *for each* $a \in \mathbb{R}$.

Theorem 2 *([20]) The equality*

$$D_\pi(x(t), \delta(t)) = (x^{'}(t), \delta^{'}(t))$$

holds, where $D_\pi(x(t),\delta(t))$ *is the* π*-derivative ([3]) of the pair* $(x(t),\delta(t))$ *and* $x^{'}$ *is the derivative of the real-valued function* $x(t)$.

Let $F : T \to \mathcal{K}$ be a set-valued mapping. If we denote by $F(t) = [f(t), g(t)]$, then

$$\pi(F(t)) = \langle [x(t), x(t)+\delta(t)], 0\rangle ,$$

where $x(t) = f(t)$ and $\delta(t) = g(t) - f(t)$.

From Theorem 2 $D_\pi(x(t),\delta(t)) = (x^{'}(t), \delta^{'}(t))$ and from canonical representation (2) we obtain the set $F_\pi^{'}(t) \in \mathcal{K}$ such that

$$F_\pi^{'}(t) = \begin{cases} \left[x^{'}(t), x^{'}(t) + \delta^{'}(t)\right] & \text{if} \quad \delta^{'}(t) \geq 0 \\ \left[x^{'}(t) + \delta^{'}(t), x^{'}(t)\right] & \text{if} \quad \delta^{'}(t) < 0. \end{cases} \tag{3}$$

We will say that $F_\pi^{'}(t_0)$ is the π-derivative of the set-valued mapping F at t_0.

Corollary 1 *([9]) Let* $F : T \to \mathcal{K}$ *be a set-valued mapping and we denote* $F(t) = [f(t), g(t)]$. *Then* F *is* π*-differentiable at* t_0 *if and only if* f *and* g *are differentiable functions at* t_0.

Example 1 *([9]) Consider the set-valued mapping* $F : (-1,1) \to \mathcal{K}$ *defined by* $F(t) = \left[-t^2, e^{-t}\right]$. *If* $f(t) = -t^2$ *and* $g(t) = e^{-t}$, *we have that* f *and* g *are differentiable functions and* $f^{'}(t) = -2t$ *and* $g^{'}(t) = -e^{-t}$. *In this case* $x^{'}(t) = -t^2$ *and* $\delta^{'}(t) = -e^{-t} + 2t$. *Therefore, from (3) we have*

$$F_\pi^{'}(t) = \begin{cases} [-e^{-t}\ ,\ -2t] & \text{if} \quad t \in (-1, 0.35173] \\ [-2t\ ,\ -e^{-t}] & \text{if} \quad t \in (0.35173, 1). \end{cases}$$

Note that the π-derivative (3) is coincident with the generalized derivative introduced in [26].

4 π-derivative for fuzzy functions

Let $\mathcal{F}$ be the set of all fuzzy intervals with bounded α-level intervals. This means that if $u \in \mathcal{F}$ then the α-level set is a closed bounded interval which we denote by $[u]^\alpha$, for all $\alpha \in [0,1]$.

Let T be a real interval. A mapping $X : T \to \mathcal{F}$ is called a fuzzy function. We denote

$$[X(t)]^\alpha = X_\alpha(t) = [f_\alpha(t), g_\alpha(t)],\ \ t \in T,\ \ 0 \leq \alpha \leq 1.$$

Note that X_α is a set-valued mapping for each $\alpha \in [0,1]$. The π-derivative $X_\pi^{'}(t)$ of a fuzzy function X is defined by

$$[X_\pi^{'}(t)]^\alpha = X_{\alpha\pi}^{'}(t),\ \ 0 \leq \alpha \leq 1, \tag{4}$$

provided that is equation defines a fuzzy interval $X_\pi^{'}(t) \in \mathcal{F}$. Note that, if the family $\{X_{\alpha\pi}^{'}(t)\}$ satisfies the conditions of the Representation Theorem [10], then there exists the π-derivative $X_\pi(t)^{'}$ of the fuzzy function X.

Example 2 *Consider the fuzzy function* $X : (0, +\infty) \to \mathcal{F}_C$ *defined by*

$$X(t)(s) = \begin{cases} \frac{s}{t} + 1 & \text{if} \quad -t \leq s \leq 0 \\ -\frac{s}{t^2} + 1 & \text{if} \quad 0 \leq s \leq t^2 \\ 0 & \text{if} \quad s \notin [-t, t^2]. \end{cases}$$

Then, for all $\alpha \in [0,1]$ *we have*

$$[X(t)]^\alpha = [f_\alpha(t), g_\alpha(t)] = [(\alpha-1)t, (1-\alpha)t^2],$$

and from (3)

$$X_{\alpha\pi}^{'}(t) = [(\alpha-1), 2(1-\alpha)t].$$

Now, the family $\{X_{\alpha\pi}^{'}(t)\}_{\alpha\in[0,1]}$ *satisfies the conditions from Representation Theorem [10] for each* $t > 0$. *Therefore, there exists the* π*-derivative* $X_\pi^{'}(t)$ *of the fuzzy function* X *for each* $t > 0$ *and*

$$\left[X_\pi^{'}(t)\right]^\alpha = [(\alpha-1), 2(1-\alpha)t].$$

ISBN: 978-989-95079-6-8

5 Fuzzy differential equation with π-derivative

We consider the initial value problem

$$X^{'}(t) = F(t, X(t)), \quad X(0) = X_0, \tag{5}$$

where $F : [0, T] \times \mathcal{F} \to \mathcal{F}$ is a continuous function and X_0 is a fuzzy interval.

Note that there are different interpretations for the problem (5). For example, in the problem (5), we can consider the H-derivative [15, 16, 17, 19, 23, 24, 25]; the Generalized derivative [4, 5, 6, 7, 8]. Also, the problem (5) can be rewrite as a family of differential inclusions [11, 12, 14]. In this section we will consider the π-derivative in the problem (5), i.e. we will study the following problem

$$X^{'}_{\pi}(t) = F(t, X(t)), \quad X(0) = X_0, \tag{6}$$

where $F : [0, T] \times \mathcal{F} \to \mathcal{F}$ is a continuous function and X_0 is a fuzzy interval. A solution of (6) is a continuous fuzzy function $X : T \to \mathcal{F}$ which verifies the equation (6) for each $t \in T$.

Taking in account the Theorem 1 and Theorem 2, we obtain a useful procedure to solve the fuzzy differential equation (6).

In fact, denote

$$[X(t)]^{\alpha} = [f_{\alpha}(t), g_{\alpha}(t)] \ , \quad [X_0]^{\alpha} = [u^0_{\alpha}, v^0_{\alpha}]$$

and

$$[F(t, X(t))]^{\alpha} = [U_{\alpha}(t, f_{\alpha}(t), g_{\alpha}(t)) \ , \ V_{\alpha}(t, f_{\alpha}(t), g_{\alpha}(t))].$$

Then, using the canocical representation (Section 3), we obtain

$$\pi\left([X(t)]^{\alpha}\right) = (x_{\alpha}(t) \ , \ \delta_{\alpha}(t)) \ , \ \pi\left([X_0]^{\alpha}\right) = \left(x^0_{\alpha} \ , \ \delta^0_{\alpha}\right)$$

where $x_{\alpha}(t) = f_{\alpha}(t)$, $\delta_{\alpha}(t) = g_{\alpha}(t) - f_{\alpha}(t)$, $x^0_{\alpha} = u^0_{\alpha}$ and $\delta^0_{\alpha} = v^0_{\alpha} - u^0_{\alpha}$. Also

$$\pi\left([F(t, X(t))]^{\alpha}\right) = \left(U^1_{\alpha}(t, x_{\alpha}(t), \delta_{\alpha}(t)), V^1_{\alpha}(t, x_{\alpha}(t), \delta_{\alpha}(t))\right)$$

Thus, we consider the following family of equations corresponding to the problem (6)

$$\begin{array}{l} D_{\pi}(x_{\alpha}(t), \delta_{\alpha}(t)) = \left(U^1_{\alpha}(t, x_{\alpha}(t), \delta_{\alpha}(t)), V^1_{\alpha}(t, x_{\alpha}(t), \delta_{\alpha}(t))\right) \\ (x_{\alpha}(0), \delta_{\alpha}(0)) = (x^0_{\alpha}, \delta^0_{\alpha}) \end{array} \tag{7}$$

with $\alpha \in [0, 1]$.

Now, using the Theorem 1 and 2 and the problem (7), we obtain a solution from problem (6). For this, we proceed as follows:

(i) Solve the differential system

$$\left\{ \begin{array}{ll} x^{'}_{\alpha}(t) = U^1_{\alpha}(t, x_{\alpha}(t), \delta_{\alpha}(t)), & x_{\alpha}(0) = x^0_{\alpha} \\ \delta^{'}_{\alpha}(t) = V^1_{\alpha}(t, x_{\alpha}(t), \delta_{\alpha}(t)), & \delta_{\alpha}(0) = \delta^0_{\alpha} \end{array} \right.$$

for x_{α} and δ_{α};

(ii) from class $(x_{\alpha}(t), \delta_{\alpha}(t))$ we obtain the interval

$$X_{\alpha}(t) = \left\{ \begin{array}{ll} [x_{\alpha}(t), x_{\alpha}(t) + \delta_{\alpha}(t)] & \text{if} \quad \delta_{\alpha}(t) \geq 0 \\ [x_{\alpha}(t) + \delta_{\alpha}(t), x_{\alpha}(t)] & \text{if} \quad \delta_{\alpha}(t) < 0 \end{array} \right.$$

(iii) By using the Representation Theorem [10], we build a fuzzy solution $X(t)$ such that $[X(t)]^{\alpha} = X_{\alpha}(t)$, for all $\alpha \in [0, 1]$.

Example 3 *Let us consider the fuzzy malthusian problem*

$$\left\{ \begin{array}{l} X'_{\pi}(t) = -\lambda X(t) \\ X(0) = \ X_0, \end{array} \right. \tag{8}$$

where $\lambda > 0$ and, as in [12], the initial condition X_0 is a symmetric triangular fuzzy number with support $[-a, a]$. That is,

$$[X_0]^{\alpha} = [-a(1 - \alpha), a(1 - \alpha)].$$

For obtain a solution of (8), we have solve the following differential system

$$\left\{ \begin{array}{ll} x^{'}_{\alpha}(t) = -\lambda x_{\alpha}(t), & x_{\alpha}(0) = -a(1 - \alpha) \\ \delta^{'}_{\alpha}(t) = -\lambda \delta_{\alpha}(t), & \delta_{\alpha}(0) = 2a(1 - \alpha). \end{array} \right.$$

The solutions of this system are

$$x_{\alpha}(t) = -a(1 - \alpha)e^{-\lambda t} \quad \textit{and} \quad \delta_{\alpha}(t) = 2a(1 - \alpha)e^{-\lambda t},$$

where $\delta_{\alpha}(t) \geq 0$ for all $t > 0$. Thus there exists a fuzzy solution from (8) $X(t)$ for all $t > 0$, such that

$$\begin{array}{rcl} [X(t)]^{\alpha} & = & [x_{\alpha}(t), x_{\alpha}(t) + \delta_{\alpha}(t)] \\ & = & [-a(1 - \alpha)e^{-\lambda t}, a(1 - \alpha)e^{-\lambda t}]. \end{array}$$

6 Conclusions

In this paper we present the concept of π-derivative for fuzzy function using the canonical representation of a class equivalence of two intervals. Those classes of equivalence are obtained by embedding the space of closed and bounded intervals in a Banach space using the embedding theorem [22].

In a forthcoming paper we will study the properties of the π-derivative fuzzy and the connection between this approach and the π-derifferentiabiliy in the fuzzy context introduced in [21]. Also, we will study the connection with others concept of fuzzy derivatives. For example, the connection with the G-differentiability introduced in [4]. Note that this concept of π-derivative is equivalent to generalized derivative in the case of set-valued function [9].

In Section 5 we study fuzzy differential equations considering π-derivative. In particular we present a form for obtain a fuzzy solution from FDEs. This approach have some advantages in relation to others interpretation, for example, an advantage that have the fuzzy solutions obtained considering π-derivative seem to be more intuitive than other solution using H-derivative (Example 3). It is worthy to stress also that the interpretation using extension principle [7, 18] and fuzzy differential inclusion [11, 12, 14] has a disadvantage because we cannot speak about the fuzzy derivative. In this issue, in a forthcoming paper, we will study the existence of solution of a fuzzy differential equation considering π-derivative.

Acknowledgment

The research in this paper has been partially supported by Ministerio de Ciencia e Innovación, Spain, through grant MTM2008-00018 and by Conicyt-Chile by project Fondecyt 1080438.

ISBN: 978-989-95079-6-8

References

[1] J.P. Aubin and A. Cellina *Differential Inclusions* Springer-Verleg, New york, 1984.

[2] S.M. Assev Quasilinear operators and their application in the theory of multivalued mappings *Proc. Steklov Inst. Math.* 2, 23–52, 1986.

[3] H.T. Banks and M.Q. Jacobs A differential calculus for multifunctions *J. Math.Anl. Appl.* 29, 246–272, 1970

[4] B. Bede and S. G. Gal Generalizations of the differentiability of fuzzy number valued functions with applications to fuzzy differential equation *Fuzzy Sets and Systems* 151, 581–599, 2005.

[5] B. Bede, I. J. Rudas and A. L. Bencsik First order linear fuzzy differential equations under generalized differentiability *Information Sciences* 177, 1648–1662, 2007.

[6] Y. Chalco-Cano and H. Román-Flores On the new solution of fuzzy differential equations *Chaos, Solitons & Fractals* 38, 112–119, 2008.

[7] Y. Chalco-Cano and H. Román-Flores Comparation between some approaches to solve fuzzy differential equations *Fuzzy Sets and Systems* 160, 1517–1527, 2009.

[8] Y. Chalco-Cano, M.A. Rojas-Medar and H. Román-Flores Sobre ecuaciones diferenciales difusas *Boletín de la Sociedad Española de Matemática Aplicada* 41, 91–99, 2007.

[9] Y. Chalco-Cano and H. Román-Flores Generalized derivative and π-derivative for set-valued mappings, Submitted to publication.

[10] P. Diamond and P. Kloeden *Metric Space of Fuzzy Sets: Theory and Application* World Scientific, Singapure, 1994.

[11] P. Diamond Brief note on the variation of constants formula for fuzzy differential equations *Fuzzy Sets and Systems* 129, 65–71, 2002.

[12] P. Diamond Time-dependent differential inclusions, cocycle attractors and fuzzy differential equations *IEEE Trans. Fuzzy System* 7, 734–740, 1999.

[13] M. Hukuhara Integration des applications mesurables dont la valeur est un compact convexe *Funkcial Ekvac.* 10, 205–223, 1967.

[14] E. Hüllermeier An approach to modeling and simulation of uncertain dynamical systems *International Journal Uncertainty, Fuzziness Knowledge-Bases Systems* 5, 117–137, 1997.

[15] O. Kaleva Fuzzy differential equations *Fuzzy Sets and Systems* 24, 301–317, 1987.

[16] O. Kaleva A note on fuzzy differential equations *Nonlinear Analysis* 64, 895–900, 2006.

[17] V. Lakshmikantham and R.N. Mohapatra *Theory of Fuzzy Differential Equation and Inclusions* Taylor & Francis, London, 2003.

[18] M. Misukoshi, Y. Chalco-Cano, H. Román-Flores and R.C. Bassanezi Fuzzy differential equations and the extension principle *Information Sciences* 177, 3627–3635, 2007.

[19] J.J. Nieto and R. Rodríguez-López Euler polygonal method for metric dynamical systems *Information Sciences* 177, 587–600, 2007.

[20] N.V. Plotnikova Systems of linear differential equations with π-derivative and linear differential inclusions, *Sbornik: Mathematics* 196, 1677–1691, 2000.

[21] M. Puri and D. Ralescu Differential and fuzzy functions *J. Math. Anal. Appl.* 91, 552–558, 1983.

[22] H. Radström An embedding theorem for spaces of convex sets *Proc. Amer. Math. Soc.* 3, 165–169, 1952.

[23] R. Rodriguez-López Comparison results for fuzzy differential equations *Information Sciences* 178, 1756–1779, 2008.

[24] L.J. Rodríguez-Muñis, M. López-Díaz, M.A. Gil The s-differentiability of a fuzzy-valued mapping *Information Sciences* 151, 283–299, 2003.

[25] H. Román-Flores and M. Rojas-Medar Embedding of level-continuous fuzzy sets on Banach spaces *Information Sciences* 144, 227–247, 2002.

[26] L. Stefanini and B. Bede Generalized Hukuhara differentiability of interval-valued functions and interval differential eqautions, *Nonlinear Analysis*, In Press.

[27] A. Tolstonogov. *Differential Inclusions in a Banach Space* Kluwer Academic Publishers, The Netherlands, 2000.

ISBN: 978-989-95079-6-8

Fuzzy logic approach to represent and propagate imprecision in agri-environmental indicator assessment

Assaghir Zainab[1,2] Girardin Philippe[1] Napoli Amedeo[2]

1.INRA: French National Institute for Agronomic Research
Colmar, France
2.LORIA: Lorraine Laboratory of Informatics Research and its Applications
Nancy, France
Email: {assaghiz,napoli}@loria.fr, {zassaghir,girardin}@colmar.inra.fr

Abstract— *The indicator of groundwater contamination developed and used in agriculture is calculated from data available in the field or data estimated by an expert. The modeling of this indicator generally requires a large number of parameters whose measure is imprecise. Several information sources provide information about the same imprecise quantities which have to be combined for defining what is called an "indicator of groundwater contamination" (I_{gro}). This indicator estimates the impact of cultivation practices on the groundwater contamination.*

In this paper, we explore a possibilistic information fusion method by using the notion of maximal coherent subsets to represent the imprecisions of multisource variables of the indicator. We also calculate the bounds of this indicator, and we propagate imprecision by using an interval analysis. Finally, we present the indicator's results for pesticides applied on different crops.

Keywords— Possibility theory, fuzzy set, fusion, maximal coherent subset, interval analysis, indicator

1 Introduction

Data used in application domains are often imperfect: imprecise, uncertain, incoherent... Some data come from information sources (database, expert) and others are approximative measurements. When multiple sources –such as a database, an expert, a website, a book– deliver information about some unknown quantity, aggregating this information can be a tedious task, especially when information are incoherent. Actually, there are many approaches to specify the impact of input parameter imprecision of a model, such as statistical sampling methods, the Monte-Carlo method [1], and bayesian method. These methods require either a significant number of data and a lot of computation time, or an exact definition of the statistical properties of the input parameters. There exists also alternative approaches based on fuzzy set and possibility theory [2, 3] to express, in a non-probabilistic sense, imprecision of parameters. Each approach represents parameters imprecision in a specific way. When, we need to aggregate data from multiple sources, fuzzy approaches show more flexibility in the treatment of incoherent information. Recently, Destercke et al [4] explored a possibilistic information fusion using the notion of maximal coherent subsets to synthesize information from several incoherent data. In this paper, we will use this method to synthesize an information from several information sources to calculate variables used in indicator assessment.

An indicator measures a certain aspect of the agrosystem. It is proposed to help farmers to improve the environmental sustainability of their agricultural practices and to reduce the pollution of environment such as indicator of groundwater contamination I_{gro}. It estimates the possibility for a pesticide to reach groundwater through leaching. The modeling and the assessment of an indicator generally require a large number of parameters whose measure is imprecise such as the variable "soil depth", that indicates how thick the soil cover is. Multiple sources provide information about the imprecise quantities which compose the indicator I_{gro}, for example, pesticide characteristics such as pesticide half-life $DT50$. The calculation of I_{gro} is based on a system of decision rules using fuzzy sets [5]. Authors in [5] do not take into account the imprecision in parameters of the indicator.

In this paper, we present an original work that, at our knowledge, has not been carried out until now. We propose here an analysis of imprecision of the indicator, especially variables used and their nature. We are interested in imprecision of parameters provided by several sources. Then, we use imprecision for evaluating the amplitude and the effect of imprecision on the value of the indicator I_{gro}. Firstly, we change the input values of variables in the process of calculation. Then, we use intervals of values in place of single values of input variables in the calculation process of the indicator I_{gro}. We also explore a fusion method, using the notion of maximal coherent subset, for summarizing information about the quantity given by several sources [4]. Finally, we calculate a variation interval for the indicator I_{gro} given by the input intervals of the variables.

The paper is divided as follows: the first section presents the indicator of groundwater contamination I_{gro}, its definition, structure and data used. Theoretical preliminaries are introduced in the third section. Then, the fourth section describes the fusion method used and the way how maximal coherent subsets can be used to obtain intervals of the imprecise variables; we also explore a way to compute the lower and upper bounds of the indicator. Then, the fifth section presents the indicator of groundwater contamination, its definition and its structure. Finally, the sixth section shows the results of the indicator for four pesticides applied in different crops.

2 Application description

In this section, we define the indicator I_{gro}, its structure and data used in the process of calculation.

ISBN: 978-989-95079-6-8

2.1 The risk of groundwater contamination I_{gro}

2.1.1 Definition and data used

The indicator module I_{gro} reflects the potential of a pesticide to reach groundwater through leaching and to affect its potential use as a source of drinking water. I_{gro} as proposed by Van der Werf and Zimmer [5] depends on four input variables : (1) pesticide leaching potential (GUS); (2) position ; (3) leaching risk, and (4) toxicity of the pesticide for humans (based on Acceptable Daily Intake). Table 1 shows variables used to calculate I_{gro}.

Table 1: All variables for the I_{gro} and their description.

Variables	Description
Pesticide characteristics (quantitative variables)	
GUS	pesticide leaching potential
DT50	pesticide half-life
koc	organic-carbon partition
ADI	Acceptable Daily Intake
Environmental characteristics (qualitative variables)	
Leaching risk	Quantity leaching
Application characteristics (quantitative variables)	
Position	Position of application

Variable "pesticide leaching potential" (GUS) provides an estimation of the risk of leaching of the compound. This potentiel is calculated by the formula:

$$GUS = \log(DT50) * (4 - \log(koc)) \quad (1)$$

where variable DT50 is the pesticide half-life and "koc" is the organic-carbone partition coefficient.
Variable DT50 is the time required for the pesticide concentration under defined conditions to decrease to 50% of the amount in application. Variable "koc" is the organic-carbon constant that describes the tendency of a pesticide to bind to soil particles. Variable "position" is the position of application of the pesticide (on the crop, on the soil, in the soil). This position is the interception rate of active ingredient per leaf area of studied culture. It is obtained from variable "soil cover". Variable "soil cover" is a value thats depends on the leaf surface of the culture. The variation of "soil cover" over time is fixed. Then, the position depends on the date of application by the farmer. The variable "leaching risk" depends on characteristics of the soil. In I_{gro}, the estimation of soil "leaching risk" is given by experts on a scale between 0 (minor leaching risk) and 1 (major leaching risk). The "position" and "leaching risk" take values between 0 and 1. The variable "Acceptable Daily Intake" (ADI) reflects chronic toxicity from humans. The pesticide properties (DT50, koc, and ADI) are given by several heterogeneous information sources such as Agritox (database, France), RIVM (database, Netherlands), Pesticide Manual (book, england), Dabene (expert, France), etc .... For each pesticide, these infomation sources give fuzzy intervals for each characteristic. Let us consider the following example: Three sources (Pesticide manuel: a book, Agritox: a database, Dabene: an expert) provide information, about the variable "DT50" (pesticide half-life) of "I_{gro}", in term of two intervals Int_1 and Int_2. The information is summarized in Table 2 and represented on Figure 1. The interval Int_2 represents the most plausible values for the variable DT50.

Table 2: Example information from sources

Source	Int_1	Int_2
Dabene	[10,42]	[20,35]
Agritox	[46,53]	50
Pesticide Manual	[21,56]	[35,40]

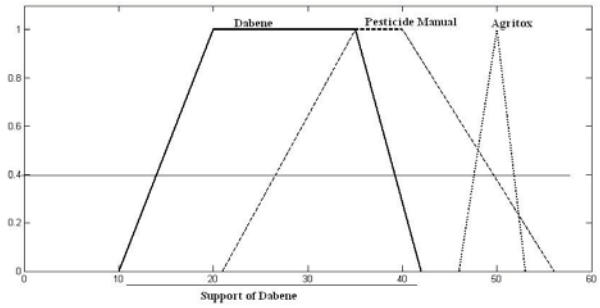


Figure 1: An example of values of DT50 given by sources, α-cut with $\alpha = 0.4$ and the support of data given by Dabene

2.1.2 The structure of the indicator of groundwater contamination I_{gro}

The indicator of groundwater contamination I_{gro} is calculated on a scale between 0 (maximal risk) and 10 (minimal risk). These values are calculated according to fuzzy if-then rules and to the degree of membership of the input variables to fuzzy subsets. The mechanism is explained below.

To calculate the indicator I_{gro}, we need to aggregate expertise knowledge and variables heterogeneous (qualitatives and quantitatives), authors in [5] are used fuzzy control [6] for achieving this two goals. For all input variables, two fuzzy subsets F (Favourable, i.e. the sets of values giving rise to acceptable environmental effect) and U (Unfavourable, i.e. the set of values giving rise to unacceptable environmental effect) are defined. Table 3 shows favourable and unfavourable limits for input variables of I_{gro}, which are extracted from literature or based on human expert knowledge.

Table 3: Favourable and unfavourable limits of input variables

Variable	Favourable limit	Unfavourable limit
DT50	1	30
GUS	1.8	2.8
ADI	1	10^{-4}
Position	1	0
Leaching risk	0	1

For each class (favourable, unfavourable, and transition interval which corresponds to values between favourable and unfavourable limits), we have defined a membership function on knowledge of experts.
The membership of values of input variables takes any value in the interval [0,1]. The value 0 represents complete non-membership and the value 1 represents complete membership. Values between 0 and 1 are used to represent partial membership. The membership function is defined in such a way that the value of an input variable either belongs fully to one of the

ISBN: 978-989-95079-6-8

two fuzzy subsets (favourable, unfavourable), or partially to both. In the latter case, the value is within a transition interval. We used membership functions that are sinus shaped in the transition interval, as they provide smoother variations of the output values than membership functions that are linear in the transition interval [5].
For example, experts classify the leaching potential GUS into three classes. They affect pesticides classified as "leacher" (i.e. GUS $>$ 2.8 or called unfavourable subset) a membership value of 1 for the fuzzy subset U and membership value of 0 for the fuzzy subset F. Pesticides classified as "no-leacher" (GUS $<$ 1.8 or called favourable subset) are given a membership value of 0 for the subset U and a membership value of 1 for the fuzzy subset F. The class of borderline compounds (1.8 $<$ GUS $<$ 2.8) falls within a transition interval where the membership value for F decreases from 1 (GUS = 1.8) to 0 (GUS= 2.8), and the membership value of U increases from 0 to 1 (see Fig. 2). The favourable (respectively unfavourable) membership function "F-Function" (respectively "U-Function") of GUS is the degree for GUS to be in the favourable subset (respectively unfavourable subset). We have also "U-Function" = 1 - "F-Function". The function of subset favourable of GUS is $0.5 + 0.5\cos(\pi(\text{GUS} - 1.8))$. If variable GUS has a value of 2.5, then "F-Function" = 0.16 and "U-Function" = 0.84.

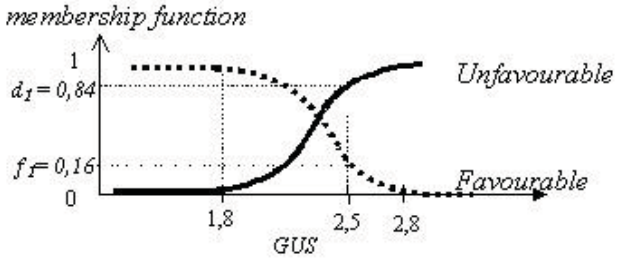


Figure 2: F-Function and U-Function function of variable GUS

The calculations are carried out according to a set of if-then rules. The experts attribute values between 0 and 10 for each rule.They consider firstly –what determines the leaching– the properties of organic matter (molecule), so they use variable GUS. If the variable GUS is favourable (GUS $<$ 1.8), then there is no problem. The molecule does not present the characteristics of leaching. If variable GUS is unfavourable (GUS $>$ 2.8), the experts are interested in the other variables of indicator. When a molecule is able to leach, it is able to reach the soil. Then, the experts use variable "position" of application. If the variable "position" is favourable, experts set a value to 9 for indicating the risk of leaching, but it is not too important. Then the environment is taken into account through the variable "leaching risk" (is the molecule leaching or not?, is the soil sensitive to the leaching?) and, finally they set on the variable "ADI".
Fig. 3 shows the if-then rules of I_{gro}. For example, if all input variables of I_{gro} are F then conclusion is 10. If all input variables of I_{gro} are U then conclusion is 0.

For each rule, we obtain the truth value by applying the minimum operator on the set of membership of the rule. For the first rule, if all input variables are F, the memberships of

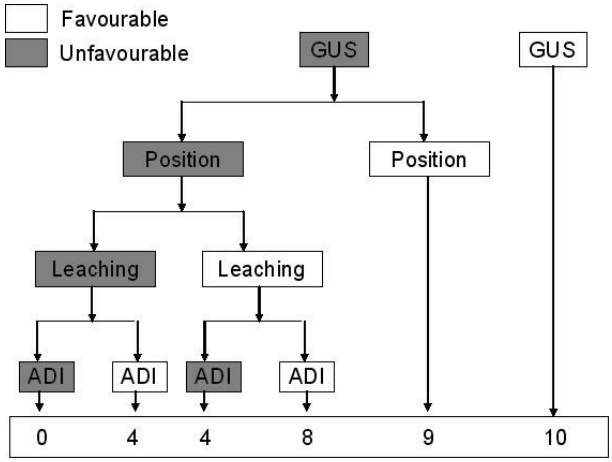


Figure 3: If-then rules of the indicator I_{gro} (I_{gro} has $2^4 = 16$ rules)

four input variables are f_i for each variable, then the truth value is $\omega_1 = \min_{i=1,\ldots,4}(f_i)$. The final score of the indicator is the average of rule conclusions weighted by their values of truth

$$I_{gro} = \frac{\sum_{i=1}^{16} \omega_i s_i}{\sum_{i=1}^{16} \omega_i}$$

3 Preliminaries

3.1 Problem statement

In this paper, we consider the indicator of groundwater contamination I_{gro}, which depends on parameters: GUS, DT50, "koc", ADI, "leaching risk" and "position". It is calculated by Van der Werf and others in [5] as $I_{gro} = \frac{\sum_{i=1}^{16} \omega_i s_i}{\sum_{i=1}^{16} \omega_i}$, where ω_i depends of variables of I_{gro}, and $s_i \in [0, 10]$.
We are interested here in multisource variable DT50, koc, ADI. Assume that the set of n information sources provide n fuzzy intervals about these variables (see Table 2). We have to search the most plausible values to calculate the indicator. We use here the support of fuzzy interval (α-cut with $\alpha = 0$ i.e. the interval Int_1 in Table 2). We use the notion of maximal coherent subset into a possibilistic fusion to synthesize a final result for multisource variables. These results combined with values of others variables will be used to compute the bounds (lower and upper) of the indicator I_{gro}.
In this paper, $\forall x \in \mathbb{R}$, x^- and x^+ represent the lower and upper bounds of interval where x can vary.

3.2 Possibility theory, fuzzy sets

In this section, we briefly summarize basic concepts of possibility theory, which are needed for understanding this paper. Fuzzy sets were firstly introduced by Zadeh [2] as a possible way to handle uncertainty in processing incertain or imprecise data, and to control expert knowledge. This theory allows the notion of graduation to express whether an element belongs to a set (For more details, see[2, 3, 7]).

3.2.1 Possibility theory

Possibility theory was introduced in 1978, in connection with fuzzy set theory, to allow reasoning to be carried out on imprecise or vague knowledge, making it possible to deal with uncertainties on knowledge [3, 8]. The basic tool of possibility

ISBN: 978-989-95079-6-8

theory is possibility distribution. A possibility distribution is equivalent to the definition of a normalized fuzzy membership function. We are interested in this theory, in this paper, because we need to synthesize information of a variables which come from various sources. When we want to synthetize the value of a variable (i.e. approximate the value of the variable depending a set of a different sources (This process is called "information fusion" (see here after). In such a situation, possibility theory can be used to handle this kind of problems. This theory presents many choices of fusion operators in different contexts.

3.3 Information fusion

Definition 1 *Information fusion consists of merging, or exploiting conjointly, several sources of information for answering questions of interest and make proper decisions.*

A large set of information fusion operators has been designed in the possibility theory framework [3]. They can be split into three subsets, according to the behavior of operators:

- Conjunctive type operators: it is the equivalent of a set intersection. It makes the assumption that all sources are reliable (i.e. all sources give information with high confidence), and usually results in very precise information. If there is an incoherence between variable values (i.e. detected by domain expert), then the result of the conjunction becomes poorly reliable, or even empty.
- Disjunctive type operators: it is the equivalent to a set union. It makes the assumption that at least one source is reliable. The result of a disjunctive operator can be considered as very reliable.
- Trade-off type operators: they are compromise operators between conjunctive and disjunctive operators. They are typically used when sources are partially conflicting (i.e. in contradiction). As its name indicates, such an operator tries to make a trade-off between disjunction and conjunction for achieving a good balance between informativeness and reliability.

4 Methodology

4.1 Fusion based on maximal coherent subsets to assess imprecision of multi-source variables

Information sources which supply the information about imprecise variables (DT50, koc, ADI) are heterogeneous (book, expert, database). They come from different places (France, USA, Netherlands), then the type of soil, weather and laboratory conditions are not similar. Furthermore, we ignore the reliability of sources (they do not supply a degree of confidence with the value). They provide data in term of a fuzzy interval (or fuzzy number). Also, no information about conflict and relations between sources is available. Then, we can not use a conjunctive rule (respectively disjunctive and compromise rule). We need a fusion method which take into account all information given by sources (i.e. without discarding any source). The notion of maximal coherent subset is a natural way for achieving this goal. The fusion method using the notion of maximal coherent subset consists firstly inn applying a conjunctive operator into each non-conflicting subset of sources, then a disjunctive operator between those partial results. With such a method, as much precision as possible is gained when we do not ignore any sources [9, 10]. We will explain in details how this approach applies to support of fuzzy intervals of variables of I_{gro} (see the section 4.3).

4.2 Computing maximal coherent subsets of intervals

In this section, we describe how we can obtain values of multisource variable using the fusion method based on maximal coherent subset of intervals, which are give by information sources. Let us consider n intervals, $I_i = [a_i, b_i], (i = 1, \ldots, n)$. This method allows to find every maximal subsets of sources. A maximal subset is obtained when $\bigcap_{i\in[n]} I_i \neq \emptyset$, where $[n]$ represents the set of 2^n subsets of the set $\{1, \ldots, n\}$. Then, we apply the union of these partial results (i.e. $\bigcup_j \bigcap_{i\in[n]} I_i$).

Algorithm 1: Maximal coherent subset of intervals

Input: n intervals
Output: List of m maximal coherent subsets I_j
List=$\emptyset$;$j = 1$;$I_j = \emptyset$;
Order in an increasing order
$\{a_i, i = 1, \ldots, n\} \cup \{b_i, i = 1, \ldots, n\}$;
Rename them $\{c_i, i = 1, \ldots, 2n\}$with $type(i) = a$
if $c_i = a_k$ and $type(i) = b$ if $c_i = b_k$;
for $i = 1$ to $2n - 1$ **do**
  if $type(i) = a$ **then**
    Add source k to I_j t.q. $c_i = a_k$;
    if $type(i + 1) = b$**then**
      Add k to List (I_j);
      $j = j + 1$;
  else
    Remove source k from I_j such as $c_i = b_k$;

Algorithm 1, that finds maximal coherent subsets, was introduced by Dubois and others in [11]. The algorithm 1 is linear in the number of intervals, and thus computationally efficient. The algorithm is based on an increasing sorting of the interval end-points into a sequence $(c_i)_{i=1,\ldots,2n}$. Each time and only then, an element c_i of type upper bound which is an upper bound of an interval, followed by an element c_{i+1} of type lower bound wich is a lower bound of an interval meet, a maximally coherent set is obtained.

4.3 Example of application

We consider the variable DT50 (pesticide half-life) provided by three information sources (see Table 2): a French expert (Dabene), a French database (Agritox), and an England book (Pesticide Manual).
Using algorithm 1 on the level $\alpha = 0$ (i.e. the support Int_1 of fuzzy intervals in Table 2), we find two maximal coherent subsets $K_1 = \{\text{Dabene, Pesticide Manual}\}$ and $K_2 = \{\text{Agritox, Pesticide Manual}\}$. After applying the maximal coherent subset method, the result of the variable DT50 is: (Dabene $\cap$ Pesticide Manual) $\cup$ (Agritox $\cap$ Pesticide Manual) = [21,42] $\cup$ [46,53].

4.4 Computation method to the fuzzy weighted average

The objective of this section is to compute the bounds of the indicator I_{gro}. The function $f(s_1, s_2, \ldots, s_n, \omega_1, \ldots, \omega_n) = \frac{\sum_{i=1}^{n} \omega_i s_i}{\sum_{i=1}^{n} \omega_i}$, where ω_i is a fuzzy interval and $s_i \in \mathbb{R}$, is called

ISBN: 978-989-95079-6-8

fuzzy weighted average by Dong and Wong [12]. Authors in [12] proposed an algorithm to compute the fuzzy weighted average. Their algorithm is based on the α-cut representation of fuzzy sets and combinatorial interval analysis. Subsequently, Liou and Wang [13] suggested an improved fuzzy weighted average algorithm to simplify the computational process. Afterwards, Lee and Park [14] have proposed an even more efficient algorithm for fuzzy weighted average by reducing the number of arithmetic operations to $O(n \log n)$. The main idea of Lee and Park in [14] is to sort the s_i variables such as for all $i < j$, we have $s_i \leq s_j$. In this order, we can find a rank k with $1 < k < n$ such as for all $i \leq k$, the function f is decreasing with respect to ω_i, and for all $i > k$, the function f is increasing with respect to ω_i. Then, we can compute the lower and the upper bounds of the fuzzy weighted average in accordance with [15],which are bounds of I_{gro}, as:

$$f_k^- = \frac{\sum_{j=1}^{k}(\omega_j)^+.s_j + \sum_{j=k+1}^{n}(\omega_j)^-.s_j}{\sum_{j=1}^{k}(\omega_j)^+ + \sum_{j=k+1}^{n}(\omega_j)^-} \quad (2)$$

$$f_k^+ = \frac{\sum_{j=1}^{k}(\omega_j)^-.s_j + \sum_{j=k+1}^{n}(\omega_j)^+.s_j}{\sum_{j=1}^{k}(\omega_j)^- + \sum_{j=i+1}^{n}(\omega_k)^+} \quad (3)$$

Where ω_i^- and ω_i^+ are respectively the lower and upper bounds of ω_i. The equation (2) (respectively (3)) represent the lower (upper) bound of the level k and $k = 1, \ldots, n$. Then, we determinate the fuzzy interval of indicator such as introduced by Fortin and others in [15]:

$$[\min_{k=1,\ldots,n}(f_k^-), \max_{k=1,\ldots,n}(f_k^+)] \quad (4)$$

5 Results

In this application when computing I_{gro}, about fifteen information sources are used such as ARS (USA), Agritox (France), Pesticide manual (England)... These sources provide information based on a fuzzy interval or on a fuzzy number. In this paper, we use the support of fuzzy intervals given by the sources. The decision rules do not change. For each variable, we replace the input value by an input interval. We propagate the imprecision into computation indicator to calculate a fuzzy interval of indicator. For a pesticide application, an indicator of groundwater contamination score is calculated.
The application considers the indicator of groundwater contamination I_{gro} in several sites with different crops. The positions of application of pesticides is shown in Table 4. Its value is calculated as: "position $= \frac{\text{soil coverd by the crop}}{100}$". The "soil covered" by the crop is estimated by the user of the systems [16]. The variable "leaching risk" set the value of 0.9 [16].

Table 4: Values of the variables of application condition

Pesticide name	Position	soil covered
Chloridazone	0.9	90
Glyphosate	0.75	75
Nicosulfuron	0.1	10
Isoproturon	0.1	10

Pesticide properties (DT50, koc, and ADI) given by several information source are calculated by fusion based on maximal coherent subsets. Table 5 shows the results of pesticide characteristics. The values shown in Table 5 are convexified to simplify the calculation. By using the interval analysis [17] and the equation 1, we compute the fuzzy interval of variable GUS, such as:

$$\text{GUS}^- = \log(\text{DT50}^-) \times (4 - \log(\text{koc}^+)) \quad (5)$$

$$\text{GUS}^+ = \log(\text{DT50}^+) \times (4 - \log(\text{koc}^-)) \quad (6)$$

where ([DT50$^-$,DT50$^+$]) and ([koc$^-$,koc$^+$]) are fuzzy intervals of variables DT50 and Koc.

Table 5: Results of pesticide characteristics

Pesticide name	GUS	DT50	koc	ADI
Chloridazone	[2.53,3.19]	[21,46]	120	0.025
Glyphosate	[0.73,2.97]	[18,47]	[167,2640]	0.3
Nicosulfuron	[2.78,5.39]	[15,43]	[5,43]	0.4
Isoproturon	[2.68,3.03]	[22,28]	[80,99.99]	0.4

Afterwards, we compute the membership intervals of input variables of indicator. For each input variable that varies in $[a, b]$ and has a membership function f, we compute the membership interval by using the function:

$$f([a,b]) = \begin{cases} [f(a), f(b)] & \text{if } f \text{ is increasing} \\ [f(b), f(a)] & \text{if } f \text{ is decreasing} \end{cases} \quad (7)$$

Then, we calculate the truth intervals of the sixteen decision rules of I_{gro}. These truth intervals can be computed by minimum extension applied on intervals such as $\min([a, b], [c, d]) = [\min(a, c), \min(b, d)]$. Then, we use the equations (2) and (3) to compute respectively the lower and upper bounds of indicator for the sixteen α-cuts established by the fuzzy input membership intervals. Finally, we obtain the final interval of indicator by the equation (4).
The Table 6 shows the results of I_{gro} for different pesticides. The fuzzy interval in Table 6 represent the exact interval when the indicator is varying. It represents also the support, of the fuzzy interval of the indicator I_{gro}, that can obtain from all fuzzy intervals of input variables.

Table 6: Results of indicator

Pesticide name	Interval of I_{gro}
Chloridazone	[4.3,9.82]
Glyphosate	[0,10]
Nicosulfuron	[4.42, 4.45]
Isoproturon	[4.42,5.27]

The results presented in Table 6 show a lot of imprecision in some cases. We take example of the organic matter "glyphosate": the indicator obtained by application of glyphosate on the soil varies between 0 and 10. This imprecision is due to the variable GUS –which varies in [0.73,2.97]– that reach limits fixed by experts in Table 3 (i.e (lower limit) $<$ (favourable limit), and (upper limit) $>$ (unfavourable limit)). We note that the indicator varies in [0,10], when we have one of the input variables has a lower (respectively upper) limit exceeds favourable (respectively unfavourable) limit. In the

ISBN: 978-989-95079-6-8

calculation of only one value of indicator I_{gro} [5], the role of the variable GUS is more important than the role of others variables. In this paper, all variables have the same influence in the final result.

6 Discussion and conclusions

In this paper, we introduced the calculus of the indicator of groundwater contamination I_{gro}. We try to control the imprecisions of input variables by using a possibilistic fusion method based on the maximal coherent subsets. Then, we use the notion of interval analysis to propagate the imprecision in the process of calculus of indicator. A fusion method for merging fuzzy subsets, based on the notion of maximal coherent subsets, is proposed to represent the imprecision of input parameters of the indicator of groundwater contamination. This notion appears as a very natural way to conciliate two objectives, gaining information and considering all the conflicts between sources. The method of fusion is simple: it can be applied without any additional information, and its computational complexity remains affordable. The way is summarizes information is conceptually attractive: maximal coherent subsets are the best we can do in the presence of conflict.
We have also described a way to calculate the bounds of indicator. The method of fuzzy interval arithmetic used in this paper is also simple. The support of the fuzzy number of intervals is calculated. Also, we can calculate the membership function of indicators by computing several values of α-cut of input variables. This method allows to obtain exact fuzzy profiles of indicator with reducing time complexity as in the sampling case. It still remains to validate the fusion method using the notion of maximal coherent subset in contrast with other fusion rules. We have to calculate the fuzzy interval for all α-cut to obtain the distribution of I_{gro}. We plane to use the methodology described in this paper (representing by fuzzy intervals the imprecision and propagating this imprecision by interval analysis) to calculate other indicators such as indicator of pesticide in environment which has a process of calculation similar to process of I_{gro}.

References

[1] N. Metropolis and S. Ulam. The monte carlo method. *Journal of the American statistical association*, 44(247):335–341, 1949.

[2] L.A. Zadeh. Fuzzy sets. *Information and Control*, 3(8):338–353, 1965.

[3] D. Dubois and H. Prade. *Théorie des possibilités*. Masson, Paris, 1985.

[4] S. Destercke, D. Dubois, and E. Chojnacki. Fusion using maximal coherent subset. In *IEEE, International Conference on Fuzzy Systems*, pages 445–450, 2007.

[5] H. Van der Werf and C. Zimmer. An indicator of pesticide environmental impact based on a fuzzy expert system. *Chemosphere*, 36(10):2225–2249, 1998.

[6] D. Driankov, H. Hellendoorn, and M. Reinfrank. *An introduction to fuzzy control*. Springer-Verlag New York, Inc., New York, NY, USA, 1993.

[7] B. Bouchon-Meunier. *La logique floue*. Presses Universitaires de France, Etats-Unis, 1999.

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

[9] N. Rescher and R. Manor. On inference from inconsistent premises. *Theory and Decision*, 1:179–217, 1970.

[10] S. Benferhat, D. Dubois, and H. Prade. Reasoning in inconsistent startified knowledge bases. *Proceedings of the 26th inter. Symp. in Multi-valued logic(ISMVL)*, pages 184–189, 1996.

[11] D. Dubois, P. Hajek, and H. Prade. Knowledge-driven versus data-driven logics. *J. of Logic, Language and Information*, 9(1):65–89, 2000.

[12] W. M. Dong and F. S. Wong. Fuzzy weighted averages and implementation of the extension principle. *Fuzzy Sets and Systems*, 21(2):183 – 199, 1987.

[13] T.S. Liou and M. J. Wang. Fuzzy weighted average: An improved algorithm. *Fuzzy Sets and Systems*, 49(3):307 – 315, 1992.

[14] D. H. Lee and D. Park. An efficient algorithm for fuzzy weighted average. *Fuzzy Sets and Systems*, 87(1):39 – 45, 1997.

[15] J. Fortin, D. Dubois, and H. Fargier. Gradual numbers and their application to fuzzy interval analysis. *IEEE Trans. on fuzzy systems*, 16(2):388–402, 2008.

[16] C. Bockstaller and P. Girardin. Mode de calcul des indicateurs agri-environnementaux de la méthode indigo. Rapprt technique, INRA, 2003.

[17] Ramon Edgar Moore. *Interval Analysis*. Prentice-Hall, New York, 1966.

ISBN: 978-989-95079-6-8

Representation of States on MV-algebras by Probabilities on R-generated Boolean Algebras

Brunella Gerla[1] Tomáš Kroupa[2,3]
1. Department of Informatics and Communication, University of Insubria,
Via Mazzini 5, 21100 Varese, Italy
2. Institute of Information Theory and Automation of the ASCR
Pod Vodárenskou věží 4, 182 08 Prague, Czech Republic
3. Department of Mathematics, Faculty of Electrical Engineering
Czech Technical University in Prague, 166 27 Prague
Email: brunella.gerla@uninsubria.it, kroupa@utia.cas.cz

***Abstract*— *Any MV-algebra M can be embedded as a lattice in the Boolean algebra $B(M)$ that is R-generated by M. We relate the study of states on an MV-algebra M to the study of finitely additive probabilities on $B(M)$. In particular, we show that each state on M can be uniquely extended to a finitely additive probability on $B(M)$. In case that M is a PMV-algebra, the conditional state $s(a|b)$ defined for $a, b \in M$ with $s(b) \neq 0$ is extended to the classical conditional probability $p(a \cdot b|b)$ on $B(M)$ of the a-proportion of the event b, given the event b.*

***Keywords*—** MV-algebra, R-generated Boolean algebra, state, probability, conditional probability

1 Introduction

States on MV-algebras were investigated by Mundici in [1] as $[0, 1]$-valued additive functionals on (equivalence classes of) formulas in Łukasiewicz propositional logic with the intention to capture the notion of "average" truth degree of a formula. In [2] a probability theory on MV-algebras is systematically developed. The study of states is frequently enhanced by looking directly at certain probabilities induced by the states: there is a one-to-one correspondence between the states on an MV-algebra and the Borel probability measures on the maximal spectrum of the MV-algebra [3, 4], where the bijection is the Lebesgue integral of a continuous function on the maximal spectrum with respect to a uniquely determined Borel probability measure. In this paper we introduce another way of representing a state within a Boolean probability theory. By Jenča's result [5], each MV-algebra M is embedded as a lattice in a uniquely determined Boolean algebra $B(M)$ that is said to be R-generated by M (see [6], [5]). We will show how to relate the study of states over M to the study of certain finitely additive probabilities on $B(M)$. In particular, our approach sheds a new light on the definition of conditional state appearing in [7].

In Section 2 we repeat the basic notion concerning MV-algebras and R-generated Boolean algebras. Section 3 contains the main results: Proposition 3, in which we show that there is only one probability on $B(M)$ extending a state on M, and Proposition 4, which gives a geometrical and topological description of such extensions in case M is a semisimple MV-algebra. In Section 4 we relate conditional states on M to conditional probabilities on $B(M)$ (Proposition 5).

2 Preliminary Notions

2.1 MV-algebras

MV-algebras [8] were introduced by Chang in 1958 as the algebraic counterpart of propositional Łukasiewicz logic.

Definition 1. An MV-algebra is a structure $\langle A, \oplus, \neg, 0\rangle$ with a binary operation $\oplus$, a unary operation $\neg$ and a constant 0 such that $\langle A, \oplus, 0\rangle$ is an abelian monoid and the following equations hold for every $x, y \in A$:

$$\neg\neg x = x$$
$$x \oplus \neg 0 = \neg 0$$
$$\neg\left(\neg x \oplus y\right) \oplus y = \neg\left(\neg y \oplus x\right) \oplus x$$

The most important example of an MV-algebra is the real unit interval $[0, 1]$ equipped with operations $x \oplus y = \min(1, x + y)$ and $\neg x = 1 - x$. Indeed, the class of MV-algebras form a variety that is generated by $[0, 1]$.
On each MV-algebra A we define $1 = \neg 0$, $x \odot y = \neg\left(\neg x \oplus \neg y\right)$, $x \ominus y = x \odot \neg y$. Note that in the MV-algebra $[0, 1]$ we have $x \odot y = \max\left(0, x + y - 1\right)$ and $x \ominus y = \max(0, x - y)$.

Let A be an MV-algebra. For any two elements x and y of A we write $x \leq y$ if and only if $\neg x \oplus y = 1$. It follows that $\leq$ is a partial order that in the case of the algebra $[0, 1]$ coincides with the natural order. Further, defined connectives

$$x \vee y = \neg\left(\neg x \oplus y\right) \oplus y \quad (1)$$
$$x \wedge y = \neg\left(\neg x \vee \neg y\right) \quad (2)$$

are such that the structure $\langle A, \wedge, \vee, 0, 1\rangle$ is a distributive lattice with bottom element 0 and top element 1. An *MV-chain* is an MV-algebra in which the order relation $\leq$ is total.

For any MV-algebra M, the set $\{x \in M \mid x \oplus x = x\}$ is the largest Boolean algebra contained in M.

Given an MV-algebra A and a set X, the set A^X of all functions $f : X \to A$ becomes an MV-algebra if the operations $\oplus$ and $\neg$ and the element 0 are defined pointwise. MV-algebras of functions taking values in $[0, 1]$ can be characterized by means of their ideals.

An *ideal* of an MV-algebra A is a subset I of A such that $0 \in I$, if $x, y \in I$ then $x \oplus y \in I$, and if $x \in I$, $y \in A$ and $y \leq x$ then $y \in I$. An ideal is *maximal* if it is not contained in any proper ideal. An MV-algebra A is said to be *semisimple* if

ISBN: 978-989-95079-6-8

and only if A is non trivial and the intersection of all maximal ideals of A is $\{0\}$. By the *maximal spectrum* we mean the (nonempty) set of all maximal ideals of A. This set can be made into a compact Hausdorff space.

It can be shown that an MV-algebra A is semisimple if and only if A is isomorphic to a separating MV-algebra of $[0,1]$-valued continuous functions on some nonempty compact Hausdorff space (actually the maximal spectrum of A), with pointwise operations. For other notions related to MV-algebras we refer the reader to [9].

A *product MV-algebra* (or PMV-algebra, for short, see [10], [11]) is a structure $\langle A, \oplus, \neg, \cdot, 0\rangle$, where $\langle A, \oplus, \neg, 0\rangle$ is an MV-algebra and $\cdot$ is a binary associative and commutative operation on A such that for any $x, y, z \in A$, $x \cdot 1 = x$ and $x \cdot (y \ominus z) = (x \cdot y) \ominus (x \cdot z)$.

Note that $[0,1]$ is a PMV-algebra where the operation $\cdot$ is the usual multiplication of real numbers. Further, a semisimple MV-algebra A is a PMV-algebra if and only if A is closed with respect to the pointwise real product of functions and, in this case, the unique product on A is the pointwise real product of functions.

2.2 Boolean algebra R-generated by an MV-algebra

Let M be an MV-algebra. There exists a unique (up to a Boolean isomorphism) Boolean algebra $B(M)$ such that the lattice reduct of M

(i) is a sublattice of $B(M)$ containing both the elements 0 and 1 of M,

(ii) generates $B(M)$ as a Boolean algebra.

The Boolean algebra $B(M)$ is called an *R-generated Boolean algebra*. See [6, Chapter II.4] for details. For every $a \in B(M)$ there exists $n \in \mathbb{N}$ and a finite chain $a_1 \le \cdots \le a_{2n}$ in M such that

$$a = \bigvee_{i=1}^{n} (a_{2i} \setminus a_{2i-1}) \tag{3}$$

holds true in $B(M)$ and $(a_{2i} \setminus a_{2i-1}) \wedge (a_{2j} \setminus a_{2j-1}) = 0$ for each $i, j \in \{1, \ldots, n\}$ with $i \neq j$, where $\setminus$ denote the symmetric difference in the Boolean algebra $B(M)$.

Theorem 1 (Jenča [5], Theorem 2)**.** *Let M be an MV-algebra. Then there exists a surjective mapping $\varphi_M : B(M) \to M$ such that*

$$\varphi_M(a) = \bigoplus_{i=1}^{n} (a_{2i} \ominus a_{2i-1}), \quad a \in B(M), \tag{4}$$

where $a_1, \ldots, a_{2n} \in M$ are as in (3) *and the value of $\varphi_M(a)$ is independent on the choice of the representation* (3)*. Moreover, the mapping φ_M satisfies*

(i) $\varphi_M(1) = 1$,

(ii) $\varphi(a) = a$, *for every* $a \in M$,

(iii) if $a, b \in B(M)$ *are such that* $a \wedge b = 0$, *then* $\varphi_M(a) \odot \varphi_M(b) = 0$ *and* $\varphi_M(a \vee b) = \varphi(a) \oplus \varphi(b)$.

The following two examples appear in [5].

Example 1. Let M be an MV-chain. Given two elements $a, b \in M$, put $[a, b) = \{x \in M \mid a \le x < b\}$. By [6], $B(M)$ is isomorphic to the Boolean algebra of all subsets of M having the form

$$[a_1, b_1) \cup \cdots \cup [a_n, b_n)$$

where $a_i, b_i \in M$ and $[a_i, b_i) \cap [a_j, b_j) = \emptyset$, for every $i, j \in \{1, \ldots, n\}$ with $i \neq j$. Note that each $a \in M$ can be identified with $[0, a) \in B(M)$. Then

$$\varphi_M([a_1, b_1) \cup \cdots \cup [a_n, b_n)) = (b_1 \ominus a_1) \oplus \cdots \oplus (b_n \ominus a_n).$$

Example 2. Suppose now M is a semisimple MV-algebra, then M can be viewed as an MV-algebra of continuous functions from the maximal spectrum X of M to the unit interval $[0,1]$. Let $\mathcal{B} = B([0,1])$ be the Boolean algebra of all subsets of $[0,1]$ of the form $[a_1, b_1) \cup \cdots \cup [a_n, b_n)$ for some $n \in \mathbb{N}$, where $[a_i, b_i) \cap [a_j, b_j) = \emptyset$ with $i, j \in \{1, \ldots, n\}$ and $i \neq j$. Just as in Example 1, each element a of the MV-algebra $[0,1]$ can be identified with the interval $[0, a)$ of $\mathcal{B}$.

Since lattice operations in $[0,1]^X$ (and hence in M) are componentwise, then $B(M)$ is a Boolean subalgebra of $\mathcal{B}^X$. Each element $f \in M \subseteq [0,1]^X$ can be identified with the function $f^* \in B(M)$ such that $f^* : x \in X \mapsto [0, f(x)) \in \mathcal{B}$. Then, according to (3), for every $g \in B(M)$ there is $n \in \mathbb{N}$ and $a_1 \le b_1 \le \ldots \le a_n \le b_n \in M$ such that

$$g = \bigvee_{i=1}^{n} b_i^* \setminus a_i^*. \tag{5}$$

The mapping φ_M is defined as follows. For every $g \in B(M)$ and any $x \in X$, we have, by (5):

$$g(x) = \bigvee_{i=1}^{n} [0, b_i(x)) \setminus [0, a_i(x)).$$

It suffices to let $\varphi_M(g) : x \in X \mapsto (b_1(x) \ominus a_1(x)) \oplus \cdots \oplus (b_n(x) \ominus a_n(x)) \in [0,1]$.

Observe that the length n of the chain representing g does not depend on the coordinate (maximal ideal) x.

3 States and Probabilities

A *state* on an MV-algebra M is a mapping $s : M \to [0,1]$ such that $s(1) = 1$ and $s(a \oplus b) = s(a) + s(b)$, for every $a, b \in M$ with $a \odot b = 0$. In case that M is a Boolean algebra, then we denote a state on it by p and call p a *(finitely additive) probability.* Let $\mathscr{S}(M)$ be the *state space* of M, that is, the (nonempty) set of all states on M. Analogously, by $\mathscr{S}(B(M))$ we denote the set of all probabilities on the R-generated Boolean algebra $B(M)$. In the sequel we will introduce and study the relation between the two state spaces.

Proposition 1. *If s is a state on an MV-algebra M, then there is a probability p on $B(M)$ such that $p(a) = s(a)$ for every $a \in M$.*

Proof. Put

$$p(a) = s(\varphi_M(a)), \quad a \in B(M), \tag{6}$$

where φ_M is the mapping from Theorem 1, and observe that this definition is correct since, by the same Theorem 1, it does not depend on the representation of a given in (3). Then $p(1) = s(\varphi_M(1)) = s(1) = 1$. Let $a, b \in B(M)$ be such that $a \wedge b = 0$. Then Theorem 1(iii) yields

$$\begin{aligned} p(a \vee b) &= s(\varphi_M(a \vee b)) = s(\varphi_M(a) \oplus \varphi_M(b)) \\ &= s(\varphi_M(a)) + s(\varphi_M(b)) = p(a) + p(b). \end{aligned}$$

Since every $a \in M$ is a fixed point of φ_M due to Theorem 1(ii), we get that p coincides with s over M. □

ISBN: 978-989-95079-6-8

The next proposition shows that every probability on $B(M)$ is uniquely determined already on the embedded MV-algebra M.

Proposition 2. *If p and p' are two probabilities on $B(M)$ and $p(b) = p'(b)$ for every $b \in M$, then $p = p'$.*

Proof. Let $a \in B(M)$. By equation (3), there exists a finite chain $a_1 \leq \cdots \leq a_{2n}$ in M such that

$$a = \bigvee_{i=1}^{n} (a_{2i} \setminus a_{2i-1}) \tag{7}$$

in $B(M)$ and $(a_{2i} \setminus a_{2i-1}) \wedge (a_{2j} \setminus a_{2j-1}) = 0$, for each $i, j \in \{1, \ldots, n\}$ with $i \neq j$. Hence

$$\begin{aligned} p(a) &= p\Big(\bigvee_{i=1}^{n} (a_{2i} \setminus a_{2i-1})\Big) = \sum_{i=1}^{n} p(a_{2i} \setminus a_{2i-1}) \\ &= \sum_{i=1}^{n} (p(a_{2i}) - p(a_{2i-1})) = \sum_{i=1}^{n} (p'(a_{2i}) - p'(a_{2i-1})) \\ &= \sum_{i=1}^{n} p'(a_{2i} \setminus a_{2i-1}) = p'\Big(\bigvee_{i=1}^{n} (a_{2i} \setminus a_{2i-1})\Big) \\ &= p'(a). \end{aligned}$$

□

Putting together Proposition 1 with (6) and Proposition 2, we get the following uniqueness result.

Proposition 3. *If s is a state on an MV-algebra M, then the probability $p = s \circ \varphi_M$ on $B(M)$ is the unique probability on $B(M)$ such that $p(a) = s(a)$ for every $a \in M$.*

The above introduced correspondence between the states in $\mathscr{S}(M)$ and the probabilities in $\mathscr{S}(B(M))$ works in general only in one direction. Define a mapping $\Phi : \mathscr{S}(M) \to \mathscr{S}(B(M))$ by

$$\Phi(s)(a) = s(\varphi_M(a)), \tag{8}$$

for every $s \in \mathscr{S}(M)$ and every $a \in B(M)$. Since every MV-algebra M is embedded into $B(M)$, it is natural to expect that the state space of the Boolean algebra $B(M)$ is much larger than that of M. Indeed, already in case that M is the standard MV-algebra $[0, 1]$, the set $\mathscr{S}([0, 1])$ contains only one element (the state s defined by $s(x) = x$, for every $x \in [0, 1]$). So the image of $\Phi(\mathscr{S}([0, 1]))$ is a singleton, but $\mathscr{S}(B([0, 1]))$ contains infinitely-many probabilities: for every $x \in [0, 1]$, the mapping

$$a \in B([0, 1]) \mapsto \begin{cases} 1, & x \in a, \\ 0, & \text{otherwise}, \end{cases}$$

is a two-valued probability and hence it belongs to $\mathscr{S}(B([0, 1]))$.

Properties of the operator Φ are summarized below. In particular, we will show that Φ preserves the geometrical-topological structure of $\mathscr{S}(M)$. In order to show this, the convex sets $\mathscr{S}(M)$ and $\mathscr{S}(B(M))$ are endowed with the subspace topology of the product spaces $[0, 1]^M$ and $[0, 1]^{B(M)}$, respectively, so that both $\mathscr{S}(M)$ and $\mathscr{S}(B(M))$ became compact spaces (see [1]).

Proposition 4. *Let M be a semisimple MV-algebra. Then:*

(i) $\Phi(s)(a) = s(a)$, for every $s \in \mathscr{S}(M)$ and every $a \in M$;

(ii) the mapping Φ is an affine homeomorphism of $\mathscr{S}(M)$ onto the compact convex set $\Phi(\mathscr{S}(M))$;

(iii) the set $\Phi(\mathscr{S}(M))$ is affinely homeomorphic to the set of all Borel probability measures on the maximal spectrum X of M;

(iv) if M is not a Boolean algebra, then $\Phi(\mathscr{S}(M)) \subsetneq \mathscr{S}(B(M))$.

Proof. (i) This follows directly from the definition (8) and Proposition 3.

(ii) First, we will show that the mapping Φ is *affine,* that is, for every $s, s' \in \mathscr{S}(M)$ and $\alpha \in [0, 1]$, we have

$$\Phi(\alpha s + (1 - \alpha)s') = \alpha\Phi(s) + (1 - \alpha)\Phi(s').$$

For every $a \in B(M)$, the definition (8) yields:

$$\begin{aligned} \Phi(\alpha s + (1 - \alpha)s')(a) &= (\alpha s + (1 - \alpha)s')(\varphi_M(a)) \\ &= \alpha s(\varphi_M(a)) + (1 - \alpha)s'(\varphi_M(a)) \\ &= \alpha\Phi(s)(a) + (1 - \alpha)\Phi(s')(a). \end{aligned}$$

The mapping Φ is injective. Given $s, s' \in \mathscr{S}(M)$ with $s \neq s'$, find $a \in M$ such that $s(a) \neq s'(a)$. Hence it follows from (i) that $\Phi(s) \neq \Phi(s')$. The mapping Φ is continuous: take a net s_γ of elements of $\mathscr{S}(M)$ such that $s_\gamma \to s$ for some $s \in \mathscr{S}(M)$. This means that $s_\gamma(b) \to s(b)$ for every $b \in M$. Thus for every $a \in B(M)$, we get

$$\Phi(s_\gamma)(a) = s_\gamma(\varphi_M(a)) \to s(\varphi_M(a)) = \Phi(s)(a).$$

The set $\Phi(\mathscr{S}(M))$ is compact convex as an affine continuous image of $\mathscr{S}(M)$. Since Φ is a continuous bijection $\mathscr{S}(M) \to \Phi(\mathscr{S}(M))$, it is actually a homeomorphism by compactness.

(iii) The set $\mathscr{B}(X)$ of all Borel (σ-additive) probability measures on the maximal spectrum X of M is compact and convex [12, Proposition 5.22]. In the topology of $\mathscr{B}(X)$ a net μ_γ converges to μ in $\mathscr{B}(X)$ if for every continuous function $f : X \to \mathbb{R}$ we have $\int f \, \mathrm{d}\mu_\gamma \to \int f \, \mathrm{d}\mu$. By Corollary 29 in [3] or [4, Proposition 1.1], there exists a unique mapping $\Psi : \mathscr{S}(M) \to \mathscr{B}(X)$ such that

$$s(a) = \int_X \hat{a} \, \mathrm{d}\Psi(s) \tag{9}$$

for every $s \in \mathscr{S}(M)$ and every $a \in M$, where $\hat{a}$ is the image of a via the isomorphism identifying the elements of M with the separating MV-algebra of continuous functions $X \to [0, 1]$. We claim that the mapping Ψ is an affine homeomorphism. It is onto since every integral (9) with respect to some Borel probability measure from $\mathscr{B}(X)$ determines a state. It is also injective as every two states $s, s' \in \mathscr{S}(M)$ coincide whenever $\Psi(s) = \Psi(s')$ by the representation (9). The mapping Ψ is affine since the functional $\mu \in \mathscr{B}(X) \mapsto \int_X f \, \mathrm{d}\mu$ is linear for every continuous function $f : X \to [0, 1]$. Finally, we will show that Ψ^{-1} is continuous. Take a net μ_γ and μ with $\mu_\gamma \to \mu$ in $\mathscr{B}(X)$. It results directly from the definition of convergence in $\mathscr{B}(X)$ and in $\mathscr{S}(M)$ that $\Psi^{-1}(\mu_\gamma) \to \Psi^{-1}(\mu)$ in $\mathscr{S}(M)$, so Ψ^{-1} is continuous. Since $\mathscr{S}(M)$ is compact, the mapping Ψ is a homeomorphism. As an inverse of an affine homeomorphism is an

ISBN: 978-989-95079-6-8

affine homeomorphism and a composition of affine homeomorphisms is again an affine homeomorphism, take the mapping $\Psi \circ \Phi^{-1} : \Phi(\mathscr{S}(M)) \to \mathscr{B}(X)$ to finish the proof.
(iv) The MV-algebra M is not a Boolean algebra, so there exists an element $f \in M$ and $x \in X$ with $0 < f(x) < 1$. Put $\beta = \frac{\min(f(x),1-f(x))}{2}$ and define a mapping $p : B(M) \to [0,1]$ by

$$p(g) = \begin{cases} 1, & \beta \in g(x), \\ 0, & \text{otherwise}, \end{cases} \qquad g \in B(M).$$

A routine check shows that p is a probability on $B(M)$. We will show that there is no state $s \in \mathscr{S}(M)$ such that $\Phi(s) = p$. By way of contradiction, assume that such a state s exists. Then Proposition 3 together with (8) imply that $s(g) = p(g)$ for every $g \in M$. Since $s(f) = p(f) = 1$ and $s(\neg f) = p(\neg f) = 1$, we get

$$s(f \oplus \neg f) = s(1) = 1 \neq 2 = s(f) + s(\neg f),$$

which is a contradiction and thus $p \in \mathscr{S}(B(M)) \setminus \Phi(\mathscr{S}(M))$. □

On the one hand, every state on a semisimple MV-algebra can be viewed as the integral (9) with respect to a uniquely determined Borel probability measure that is defined on the maximal spectrum X of M. On the other hand, the operator Φ maps a state to a finitely additive probability on a Boolean subalgebra of the direct sum $\mathcal{B}^X$ (see Example 2). While the transformation Φ enables only to embed $\mathscr{S}(M)$ into a (huge) state space $\mathscr{S}(B(M))$, we will show below that and how probabilities on direct sums of Boolean algebras appear naturally already in classical Kolmogorov model of probability.

Let $B_1, \dots, B_k$ be Boolean algebras. Let p_i be a probability on B_i for $i = 1, \dots, k$. By B we denote a Boolean algebra that is the direct sum of $B_1, \dots, B_k$ (see [13, §16]). Every element of $b \in B$ can be identified with an n-tuple $(b_1, \dots, b_k)$ such that $b_i \in B_i$ for each $i = 1, \dots, k$. The elements b of B are random events whose meaning is "precisely one of all the events $b_1, \dots, b_k$ occurs". These classes of random events capture *two-stage random experiments*, such as the random selection of a ball from a randomly selected box. In this interpretation, each number $i = 1, \dots, k$ denotes a box and the corresponding Boolean algebra B_i models all outcomes of a random selection of a ball from the box i. Precisely, for some nonnegative $\alpha_1, \dots, \alpha_k$ with $\sum_{i=1}^k \alpha_i = 1$, the procedure can be described as follows:

(i) select B_i with a probability α_i;
(ii) if B_j was selected in the previous stage, then perform the random experiment with outcomes in B_j and probabilities described by p_j.

A probability p on B should then express the above introduced meaning of the two-stage experiment. This means that for every $b \in B$, we get

$$p(b) = \sum_{i=1}^{k} \alpha_i p(b_i). \tag{10}$$

The interrelationship between states on MV-algebras and probabilities on direct sums as defined above is made possible by the mappings Φ and φ_M.

Example 3. Let M be the direct product of two standard MV-algebras, that is, $M = [0,1]^2$. According to (9) every state s on M is of the form

$$s(a) = s((a_1, a_2)) = \alpha a_1 + (1-\alpha)a_2, \quad a = (a_1, a_2) \in M, \tag{11}$$

for some $\alpha \in [0,1]$. The Boolean algebra $B(M)$ is precisely the direct sum of the two Boolean algebras $\mathcal{B}$, where $\mathcal{B}$ is as in Example 2. The MV-algebra M embeds into $B(M)$ by sending each $a = (a_1, a_2) \in M$ to the function $a^*(i) = [0, a_i)$, $i = 1, 2$. Let λ denotes the restriction of Lebesgue measure to $\mathcal{B}$. Then it follows from (11) that

$$\Phi(s)(g) = \alpha\lambda(g(1)) + (1-\alpha)\lambda(g(2)), \quad g \in B(M). \tag{12}$$

Indeed, as in Example 2, for every $g \in B(M)$ and $i = 1, 2$ find the representation $a_1^i, b_1^i, \dots, a_n^i, b_n^i \in [0,1]$ such that $g(i) = [a_1^i, b_1^i) \cup \dots \cup [a_n^i, b_n^i)$ with all the intervals disjoint. Then

$$\varphi_M(g)(i) = (b_1^i \ominus a_1^i) \oplus \dots \oplus (b_n^i \ominus a_n^i)$$

and $(\varphi_M(g)(1), \varphi_M(g)(2)) \in M$. As a consequence,

$$\begin{aligned} \Phi(s)(g) &= s(\varphi_M(g)) = \alpha\varphi_M(g)(1) + (1-\alpha)\varphi_M(g)(2) \\ &= \alpha(b_1^1 \ominus a_1^1) \oplus \dots \oplus (b_n^1 \ominus a_n^1) \\ &\quad + (1-\alpha)(b_1^2 \ominus a_1^2) \oplus \dots \oplus (b_n^2 \ominus a_n^2) \\ &= \alpha\lambda(g(1)) + (1-\alpha)\lambda(g(2)), \end{aligned}$$

which proves (12).

The model from Example 3 can be given a straightforward probabilistic interpretation. Suppose that some underground platform is accessible from a street either via a stairway or via an escalator. Mr. Smiley chooses the stairway with a probability $\alpha \in [0,1]$ and the escalator with a probability $1-\alpha$. Train arrivals are uniformly distributed over the interval $[0,1]$. What is a probability that Mr. Smiley's waiting time for the train will be smaller or equal to $a_1 \in [0,1]$ (in case he uses the stairway) or smaller or equal to $a_2 \in [0,1]$ (in case he uses the escalator)? The investigated event $a = ([0, a_1), [0, a_2))$ belongs to $B(M)$, while the many-valued event $a = (a_1, a_2)$ is an element of M. If $p(a)$ denotes the probability of a, then (10) together with uniformity of the waiting times yield

$$\begin{aligned} p(a) &= \alpha\lambda([0, a_1]) + (1-\alpha)\lambda([0, a_2]) = \alpha a_1 + (1-\alpha)a_2 \\ &= s(a). \end{aligned}$$

4 Conditional Probability

In this section we use the probabilistic operator Φ to study conditioning in the framework of MV-algebras and states. Conditional probability on MV-algebras was studied from a variety of perspectives. In [7] the conditional probability on MV-algebras is defined in a way that mimicks the classical (Boolean) approach of conditioning "an event by an event" or "an event by a subalgebra". This approach makes use of an additional operation (product) on an MV-algebra and it was employed by Montagna in order to prove de Finetti-style coherence theorem for conditional bets in many-valued logic [14]. A completely different definition of conditional probability (so-called "conditional") was proposed by Mundici in [15].

ISBN: 978-989-95079-6-8

The following approach to conditioning is taken from [7, Definition 5]. Let s be a state on a PMV-algebra M. Given a many-valued event $b \in M$ with $s(b) > 0$, put

$$s(a|b) = \frac{s(a \cdot b)}{s(b)}, \quad a \in M. \tag{13}$$

The number $s(a|b)$ is called a *conditional state of a given b*. If $s(b)$ vanishes, then we leave $s(a|b)$ undetermined. This makes the function $a \in M \mapsto s(a|b)$ defined for "almost all" $b \in M$. Since $(a_1 \oplus a_2) \cdot b = (a_1 \cdot b) \oplus (a_2 \cdot b)$ for every $a_1, a_2 \in M$ such that $a_1 \odot a_2 = 0$, it can be easily verified that the mapping $s(.|b) : M \to [0,1]$ is a state (so-called *conditional state*) on M whenever $b \in M$ is such that $s(b)$ is nonzero. If p is a probability on a Boolean algebra B, then a product operation $\cdot$ on B becomes $\wedge$ and the formula (13) for $p(b) > 0$ reduces to

$$p(a|b) = \frac{p(a \wedge b)}{p(b)}, \quad a \in B. \tag{14}$$

Analogously, the function $p(.|b) : B \to [0,1]$ is a probability on B whenever $p(b) > 0$. In this case we call $p(.|b)$ a *conditional probability*. The major difference between (13) and (14) is the consequence of non-idempotence of the product operation on M. For every $b \in M$ with $s(b) \neq 0$, we have in general only

$$s(b|b) \leq 1 \quad \text{and} \quad s(\neg b|b) \geq 0. \tag{15}$$

In contrast to this, in classical probability $p(b|b) = 1$ and $p(\neg b|b) = 0$ holds true for any $b \in B$ with $p(b) \neq 0$. Surprisingly, this concept of conditioning found a natural application in Montagna's framework [14]: conditional bets are updated in proportion to the truth value of the event in the condition. In that follows we make an effort to justify the definition (13) with the probabilistic interpretation of states via Φ and $B(M)$.

Proposition 5. *Let M be an MV-algebra with product and $s(.|b)$ be a conditional state on it for $b \in M$ with $s(b) \neq 0$. Then there exists a probability p on $B(M)$ such that*

$$\Phi(s(.|b))(a) = p(a \cdot b|b), \quad a \in M. \tag{16}$$

Proof. Put $p = \Phi(s)$. For every $a \in M$, we obtain

$$\begin{aligned} \Phi(s(.|b))(a) = s(a|b) &= \frac{s(a \cdot b)}{s(b)} = \frac{\Phi(s)(a \cdot b)}{\Phi(s)(b)} \\ &= \frac{p(a \cdot b)}{p(b)} = \frac{p((a \cdot b) \wedge b)}{p(b)} \\ &= p(a \cdot b|b), \end{aligned}$$

since $(a \cdot b) \wedge b = a \cdot b$. □

It is worth emphasizing that the probability $\Phi(s(.|b))$ is not a conditional probability on M since $\Phi(s(.|b))(b) = p(b \cdot b|b)$ fails to be equal to 1 in general. So the operator Φ does not map conditional states to conditional probabilities. The formula (16) nevertheless connects the values of conditional states of a given b on M to the values of the classical conditional probability of $a \cdot b$ given b on $B(M)$. Conditioning a by a many-valued event b means in terms of classical conditional probability this: given b, what is a probability $p(a \cdot b|b)$ that an observation of the a-proportion of the many-valued event b will be made? Thus the conditional state defined by the formula (13) leads in accordance with (16) to the conditional probability of $a \cdot b$ given b on the Boolean algebra $B(M)$, which focuses only on the event which is included in b. The use of product in place of infimum makes the difference: the equality $p(a \wedge b|b) = p(a|b)$ holds true in contrast to $p(a \cdot b|b) \neq p(a|b)$. Taking into account the proposed interpretation, the inequalities in (15) look no longer unnatural. The "probabilistic" meaning of $s(b|b)$ is the value of $p(b \cdot b|b)$ that corresponds to the occurrence of the event $b \cdot b$ (given b) rather than b alone as the notation $s(b|b)$ suggests.

Example 4. Consider the situation from Example 3. Assume now that we know the actual waiting time does not exceed $b = (b_1, b_2) \in M = [0,1]^2$. Then the value $s(a|b)$ is according to Proposition 5 the same as $p(a \cdot b|b)$. This probability is precisely the conditional probability that the waiting time will be at most the a-proportion of b given our knowledge that it is at most b.

Acknowledgment

The first author acknowledges the grant No.1M0572 of the Ministry of Education, Youth and Sports of the Czech Republic that supported her visit at the Institute of Information Theory and Automation of the ASCR in Prague. The work of Tomáš Kroupa was supported by the grant No.1M0572 of the Ministry of Education, Youth and Sports of the Czech Republic. The second author owes his gratitude to Mirko Navara for the explanation of a role of Boolean direct sums in classical probability theory.

References

[1] D. Mundici. Averaging the truth-value in Łukasiewicz logic. *Studia Logica*, 55(1):113–127, 1995.

[2] B. Riečan and D. Mundici. Probability on MV-algebras. In *Handbook of Measure Theory, Vol. I, II*, pages 869–909. North-Holland, Amsterdam, 2002.

[3] T. Kroupa. Every state on semisimple MV-algebra is integral. *Fuzzy Sets and Systems*, 157(20):2771–2782, 2006.

[4] G. Panti. Invariant measures in free MV-algebras. To appear in *Communications in Algebra,* arXiv:math/0508445v2, 2008.

[5] G. Jenča. Boolean algebras R-generated by MV-effect algebras. *Fuzzy Sets and Systems*, 145(2):279–285, 2004.

[6] G. Grätzer. *General lattice theory*. Birkhäuser Verlag, Basel, second edition, 1998.

[7] T. Kroupa. Conditional probability on MV-algebras. *Fuzzy Sets and Systems*, 149(2):369–381, 2005.

[8] C. C. Chang. Algebraic analysis of many valued logics. *Trans. Amer. Math. Soc.*, 88:467–490, 1958.

[9] R. L. O. 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.

[10] A. Di Nola and A. Dvurečenskij. Product MV-algebras. *Multiple-Valued Logics*, 6:193–215, 2001.

[11] F. Montagna. An algebraic approach to propositional fuzzy logic. *Journal of Logic, Language and Information*, 9:91–124, 2000.

[12] K. R. Goodearl. *Partially ordered abelian groups with interpolation*, volume 20 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1986.

ISBN: 978-989-95079-6-8

[13] R. Sikorski. *Boolean Algebras*. Springer-Verlag, Berlin, 1964.

[14] F. Montagna. A notion of coherence for books on conditional events in many-valued logic. Submitted to *Journal of Logic and Computation,* 2009.

[15] D. Mundici. Faithful and invariant conditional probability in Łukasiewicz logic. 2009.

ISBN: 978-989-95079-6-8

Contributions to Dynamic Multicriteria Decision Making Models

Tiago C. Pais[1] Rita A. Ribeiro[1]

1 UNINOVA, Campus FCT-Universidade Nova Lisboa,
Caparica 2829-516, Portugal
Email: {tpp, rar}@uninova.pt

Abstract— *In this work we start by presenting a possible architecture for MultiCriteria Dynamic Decision Making (MCDDM) problems. Then we focus on contributions for the rating process because it involves two important aspects of any dynamic spatial-temporal decision problem: how to deal with uncertainty in dynamically changing input data and how consider different criteria importance, depending on criteria satisfaction and respective phase (or iteration) of the decision process. To explain and clarify the contributions we use an illustrative example of site selection for spacecraft landing on planets.*

Keywords— multicriteria dynamic decision making, imprecision on input data, dynamic criteria weighting.

1 Introduction

The main objective of this work is to present some contributions for building general multicriteria dynamic decision making (MCDMM) models. We will focus on two important aspects of any dynamic spatial-temporal decision problem: a) how to deal with uncertainty in dynamically changing input data; b) and how to consider different criteria importance, depending on criteria satisfaction and on the decision process phase (iteration). These two aspects belong to the classical process of evaluating alternatives of MCDM [1] [2] [3].

In general, the aim of Multicriteria or Multi-attribute decision making problems is to find the best compromise solution from all feasible alternatives assessed by pre-defined criteria (or attributes). This type of problems is widespread in real life situations [2], [4]. There are many methods and approaches proposed in the literature to deal with this static decision process, from utility methods to scoring and ranking ones [2] [3]. However, when facing multicriteria dynamic decisions making problems (MCDDM), where feedback from step to step is required, there are few contributions in the literature (see for example [5] [6]) .

To better explain our proposed contributions to deal with uncertainty in dynamically changing input data and criteria weights changes, during the temporal decision process, we use an simplified illustrative case study of a site selection problem for spacecraft landing on planets [7] [8].

2 MultiCriteria dynamic decision making (MCDDM)

Usually, a multicriteria decision model [1] includes 2 main tasks, rating the alternatives by aggregating their classifications for each criterion and then ranking them. In this work we discuss a dynamic multicriteria decision model because we are addressing decision paradigms with several iterations, which need to take into account past/historic information, to reach a decision (or consensus).

According to Richardson and Pugh [9] system dynamics problems have two main features: "they are dynamic: they involve quantities which change over time"; "involves the notion of feedback." This is exactly the perspective in this work: dynamically changing data and feedback from historical data.

A general multicriteria dynamic decision making (MCDDM) architecture is depicted in Fig. 1 (adapted from [7] and using concepts from [9]).

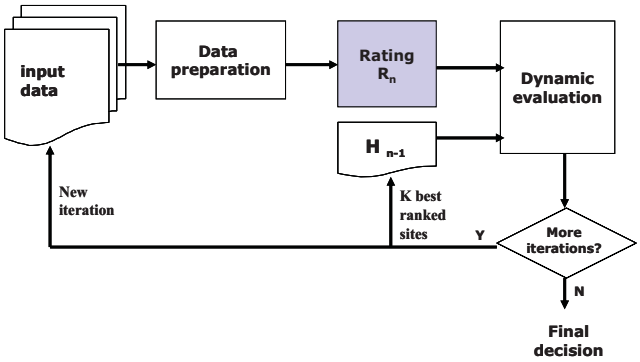


Figure 1: MCDDM architecture.

R_n represents the set of alternatives rating values from iteration n and H_{n-1} represents the historic set of rated alternatives from the previous iteration. The feedback is the set of ranked alternatives of iteration n (updated historic set of each iteration).

In summary, the main characteristics of processes/components, included in a MCDDM are:

a) Data preparation – this process deals with transformation of inputs for the decision model. It includes three main tasks: a.1) cleaning & filtering the data to extract the required criteria and the set of alternatives to be assessed (per iteration); a.2) normalization and representation of criteria for the decision model; a.3) Definition of weighting functions to express the relative importance for each criterion, within a phase or per iteration.

b) Rating process – This process refers to the classification of each alternative regarding all criteria, weighted with their relative importance. I.e. it refers to the aggregation of criteria values weighted with relative importance of criteria. Several methods exist in the literature [1] [2] [3], in our case

ISBN: 978-989-95079-6-8

we used an improved WSM method, as explained in section 3.2..

c) Dynamic evaluation process – This process refers to the aggregation of historical information about alternatives with their current rating to obtain the set of ranked alternatives [7] for the current iteration (n). When there are no more iterations the decision process stops and the best alternative is the one with the highest rate. For this evaluation we propose to use full-reinforcement operators as described in [7]

d) Historical process – This process determines a subset of good alternatives, from the ranked alternatives, to be considered as feedback information from iteration to iteration. Since this process takes places after ranking, it means historical information for a specific site "remembers" its past behaviour plus the actual rating. There are different sizes (K) for this set of historical information, depending on different phases of the decision process. For our case study we consider three phases according to expert advice.

The topic addressed in detail, in this work, will be the rating process (b), which is highlighted in grey in Figure 1.

3 Case Study: Site selection for spacecraft landing on planets

The case study used for discussing the rating process of multicriteria dynamic decision making (MCDDM) models is based on a real project [7], which objective was to determine the best target site for landing on planets for spacecrafts. The case study MCDDM model was developed for the final descent of the spacecraft (around 2 Km from surface), when hazard maps can be obtained. Some values used in the description are just indicative due to reasons of confidentiality.

3.1 Background

In this case study the criteria are hazard maps corresponding to images taken by the Spacecraft during the final descent (around 1000 m from surface) [8] [7]. We consider 7 input criteria, which correspond to the hazard maps obtained during the descent process: slope, texture, fuel, reachability, distance, shadow and scientific interest. The alternatives are the pixels of the combined search space (derived from merging 7 matrices of 512x512, corresponding to the input images/hazard maps), resulting from a data preparation process (not discussed in this work). The resulting set of alternatives is about 260.000 possible sites for each iteration (512*512= 262,144). The number of iterations is about 40 and for each one there is a evaluation process (called ranking process) including a combination of the *k* best alternatives from iteration *n-1* (historic set) with the current rated ones at the *n* iteration (for details about this process see [7]). We only consider the last iteration as historical information for two reasons: a) time constrains- we have less than 1 minute to perform 40 iterations hence an efficient process is required; b) the historic set implicitly includes all past ratings because we aggregate previous information with current one, i.e we update the alternative evaluation at each iteration and pass this information to the next iteration.

As mentioned in the abstract, in this work we focus on the rating process of MCDMM because it involves two important aspects of any dynamic spatial-temporal problem: how to deal with uncertainty in the input data of each iteration and how to consider different importance for each criteria, depending on the criteria satisfaction as well as on the phase of the decision process (iteration).

Before discussing the rating process we should explain that we assume 3 phases for this case study as follows. Phase 1 is executed until 1000 m of altitude and considers a historic of 250 set of sites, the rationale is that at higher attitudes there is a need for considering more alternatives (pixels) because the Spacecraft is still too far away from the surface and the information about the sites is still quite imprecise. Phase 2 is executed between the 1000 m and 300 m of altitude (parameters can be changed) and uses a historic set of 10 sites; since we are getting closer to the surface what we need is a small historic set (pixels) of "good" alternatives, ensuring sufficient distance between them (radius) to cover a reasonable search space. Phase 3 starts being executed at 300 m of altitude when the Spacecraft is quite close to the surface and the historic set considered was 5 sites.

Although it is out of scope to discuss the data preparation process for this case study we should mention that the results of this process are fuzzified variables (for references dealing with fuzzification see for example [4]), corresponding to each hazard map, adapted (i.e. re-calculated) at each iteration. Hence, a variable input includes a value and its corresponding membership value in the fuzzified variable.

3.2 Rating process

The rating process deals with aggregation of all classifications obtained by each alternative in each criterion. During the three phases considered, we compute these values for each alternative (per iteration). The general formulation for the weighted sum method (WSM), used as rating process in our dynamic decision process (MCDDM) is:

$$r_i = \oplus (W(ac_{i1}) \otimes ac_{i1}, ..., W(ac_{in}) \otimes ac_{in}) \quad (1)$$

Where,

$\oplus$ is the sum operator;

$\otimes$ is the product operator;

ac_{ij} is the accuracy & confidence membership value of j^{th} criterion for site *i*. (section 3.3)., and $W(ac_{ij}) = \dfrac{L(ac_{ij})}{\sum_{k=1}^{n} L(ac_{ik})}$, represents the normalized relative weight, considering a weighting function, $L(ac_{ij})$ as described in section 3.4.,

The basic principle of the aggregation method WSM (weighted sum method) is that "The overall score of an alternative is computed as the weighted sum of the attribute values" [8]. In our case study we tested other methods of aggregation besides this one (e.g. WPM, Compromise ratio [2]), however WSM combined with the proposed weighting functions, showed the best trade-of between computational efficiency versus correctness of results.

ISBN: 978-989-95079-6-8

3.3 Accuracy & confidence in input values

The accuracy and confidence parameters (a_{ij} and wc_j, respectively) will be taken into account in the decision model using the following expression:

$$ac_{ij} = wc_j * (1 - \max_{x \in [a;b]} \{ | \mu(x) - \mu(x_{ij}) | \}) * \mu(x_{ij}) \quad (2)$$

where:

wc_j is the confidence associated to criterion j;

x_{ij} is the value of jth criterion for site i;

$\mu(.)$ is the membership degree in a fuzzy set;

The interval [a,b] for x is defined by (a_{ij} is the accuracy associated to criterion j for site i):

$$a = \begin{cases} \min(D), & if\ x_{ij} - a_{ij} < \min(D) \\ x_{ij} - a_{ij}, & if\ x_{ij} - a_{ij} \geq \min(D) \end{cases}$$

$$b = \begin{cases} x_{ij} + a_{ij}, & if\ x_{ij} + a_{ij} < \max(D) \\ \max(D), & if\ x_{ij} - a_{ij} \geq \max(D) \end{cases} \quad (3)$$

D is the variable domain.

The accuracy is given as a percentage of criterion value and confidence belongs to the unit interval. For example, an accuracy of 90% for slope means that each slope value belongs to the interval $[a,b]$ where $a_{ij} = 0.9 \times x_{ij}$. On the other hand, a 0.9 confidence value means that we have a confidence on the $[a,b]$ interval of 0.9.

As can be observed, this formulation enables dealing with imprecision in the collected input hazard maps. Moreover, it allows dealing with two types of imprecision: lack of accuracy in the data and lack of confidence in the data. We believe this is a simple and easy way to tackle the problematic of how to handle imprecision in data in dynamic decision models. It is one of the contributions proposed in this work.

3.4 . Weighting functions

The second contribution is the proposal is to compute the weights using the following linear weighting function [10], [11]:

$$L(ac_{ij}) = \alpha \frac{1 + \beta ac_{ij}}{1 + \beta}, where\ 0 \leq \alpha \leq 1\ and\ 0 \leq \beta \leq 1 \quad (4)$$

Where: ac_{ij} is the accuracy and confidence membership value of j^{th} criterion for site *i*.

The logic of these weighting functions is that the satisfaction value of a criterion should influence its assigned relative importance. For example, if we are buying a car and the price is a "very important" criterion, if the car is quite expensive the final decision result should be less than the simple multiplication of weight*satisfaction value. We will not use quadratic weighting functions because it penalizes values near 1, contrarily to the linear function that presents a more smooth behaviour, which better fits our model [10], [11].

We consider the relative importance of criteria has different morphologies for each criterion, depending on each of the three phases defined. The definition of these weighting functions morphologies is given by parameters α and β. The α parameter provides the semantics for the weighting functions as follows:

- Very Important (VI =1);
- Important (I=0.8);
- Average importance (A=0.6);
- Low importance (L=0.4);
- Very Low importance (VL=0.2).

The β parameter provides the slope for the weighting functions, which will depend on the criterion at hand, with the logic that higher β means higher slope. For our case study this parameter has the following values:

- High (H =1);
- Medium (M=0.6667);
- Low (L=0.3333);
- Null (N=0);

As mentioned above the relative importance of criteria changes according to the altitude of the spacecraft, i.e. depends on the closeness to planet surface. We defined different weights for each sub-phase we are in, as described in the next two sub-sections, but these importances's can be modified by the decision maker anytime.

3.4.1 Discussion of Criteria Weighting in case study

Phase 1 (altitude: > 1000m, <2000; historic set size 250)

This phase refers to altitudes between 1000m to 2000m and a historic size of 250 candidate alternatives. In Table 1 we show the proposed values for α and β parameters for this phase, in Figure 2 we depict their respective plots, where Y-axis represents the weights and X-axis shows criteria satisfaction values within interval [0-1].

Table 1: α & β parameters (Phase 1).

	Fuel	Reach	Slope	Dist	Shad	Text	ScIn
α	1	1	1	0.6	0.4	0.8	0.2
β	0.25	0.667	0.111	0.333	0	0.333	0

For example, the rationale for "very important" is that although fuel, reachability and slope are all "very important" criteria for this phase, a lower satisfaction for criteria reachability should be quite penalized (e.g. X=0.25 corresponds to W=0.7, hence W(x)= 0.25*0.7=> 0.175). Further,, fuel and slope also consider high penalties for lower satisfaction values, visible on their plot (Figure 2).

ISBN: 978-989-95079-6-8

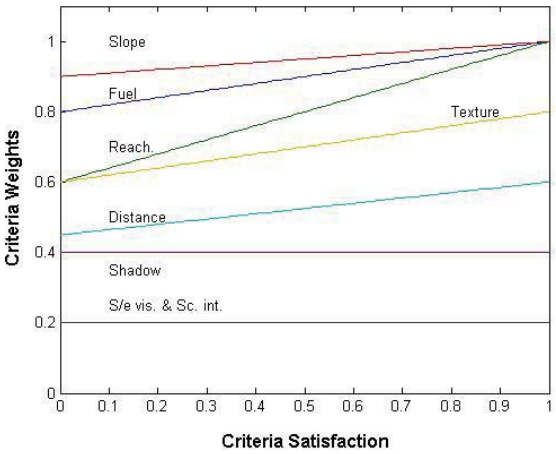


Figure 2: Weighting functions Phase I.

Phase I1 (altitude: ≤ 1000m, >300m; historic set size 10)
In Table 2 we depict the proposed values for α and β parameters for this phase, in Figure 3 their respective plots.

Table 2: α & β parameters (Phase II).

	Fuel	Reach	Slope	Dist	Shad	Text	ScIn
α	0.6	1	1	0.6	0.4	0.8	0.2
β	0.5	0.667	0.111	0.333	0	0.333	0

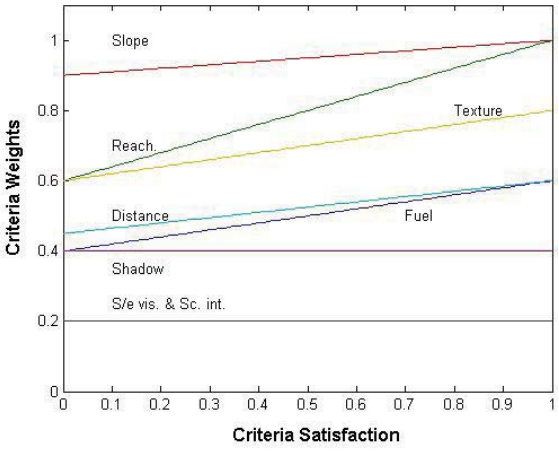


Figure 3: Weighting functions Phase II.

For example, in this phase the same rationale as in phase 1 applies for the two "very important" criteria, slope and reachability, althought slope is less penalized for lower satisfaction of criteria (e.g.0.8 weight for satisfaction 0.5 and weight 1 for satisfaction 1).

Phase III (altitude: ≤ 300m; historic set size 5)
In Table 3 we depict the proposed values for α and β parameters for this phase, in Figure 4 their respective plots.

Table 3: α & β parameters (Phase III).

	Fuel	Reach	Slope	Dist	Shad	Text	ScIn
α	0.2	1	1	0.2	0.4	0.8	0.2
β	0	0.667	0.111	0	0	0.333	0

For example, in this phase scientific interest and shadow do not play an important role because the alternatives are focused on a small area. Further, it almost does not require fuel due to the surface proximity. Hence their weights are small and constant, 0.2 for fuel & scientific interest, and 0.4 for shadow (all flat functions not dependent on satisfaction of criterion).

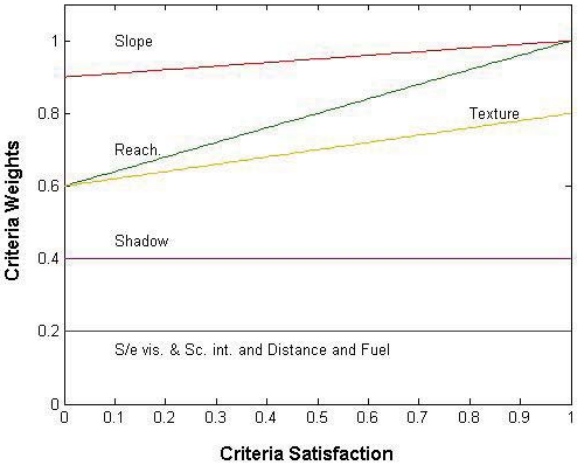


Figure 4: Weighting functions Phase III.

In summary, the relative importance proposed for each criterion followed the semantic rational:

- Slope relative importance (weight) in all phases is considered to be "very important";
- Reachability relative importance (weight) in all phases is considered to be "very important";
- Texture relative importance (weight) in all phases is considered to be "important";
- Shadow relative importance (weight) in all phases is considered to be "low";
- Distance relative importance (weight) in the first 2 phases is considered to be "average" and then it will change to "low" and "very low";
- Scientific interest (Sc int.) relative importance (weight) in all phases is considered to be "very low";
- Sun/Earth visibility (S/e vis.) relative importance (weight) in all phases is considered to be "very low";
- Fuel relative importance (weight) in phase 1 is considered to be "very important";
- Fuel relative importance (weight) in phase 2 is considered to be "medium";
- Fuel weight in phase 3 is considered to be "very low".

The rational above is expressed by the α parameter, while the β parameter provides more or less penalties for lower satisfaction values in the criteria, as depicted in the figures for each phase. A simplification of the weighting functions could be to define "a priori" a slope for each importance i.e. avoids having to define different β parameters for each weighting functions. This option simplifies the weight assignment process but it is less accurate.

ISBN: 978-989-95079-6-8

An important point to highlight is that the parameter values proposed for the phases (or iterations) can be tuned for any other dynamic decision application.
This section concludes the second proposed contribution of this work, i.e. contributions to deal with different criteria weights for steps of the dynamic decision process.

4 Discussion

In Table 4 are depicted the best three alternatives on the last iteration of each phase using default weights (described in previous section). For instance, the best alternative on the end of the 2nd phase (iteration 25) has 3D coordinates: (86 , -27 , 12) and rating value of 0.7807.

We can observe that from 16th to the 25th iteration the best site is completely different. However, the 2nd and 3rd best sites seem to be on the same region on both iterations (50 meters south). Furthermore, the best and 2nd best sites from iteration 25 changed positions with each other on the 40th iteration.

Table 4 – Rating values and 3D coordinates of the best sites on the last iteration of each phase (default weights).

Iter.	3D coord (meters)	R_n
16 (1st Phase)	(51 , 29 , 35)	0.7833
	(65 , -0.4 , 33)	0.7801
	(94 , 26 , 35)	0.7798
25 (2nd Phase)	(86 , -27 , 12)	0.7807
	(66 , -50 , 10)	0.7783
	(95 , -26 , 12)	0.7772
40 (3rd Phase)	(79 , -53 , -6)	0.9955
	(86 , -53 , -6)	0.5119
	-	-

In table 5 we present the results regarding the best sites on the last iteration of the 2nd and 3rd phase, only this time changing the Distance criterion weight to Very High on both phases. For illustrative purposes we considered the distance towards the centre of image for computing the Distance Criterion.
Comparing the results of Table 5 with Table 4, we can observe that at iteration 25th the first result is exactly the same and the 2nd and 3rd too, but on a different order. The best site maintained is position very close to the centre of the image and the distance criterion is now considered to be very high. By the same reason, site (95 , -26 , 12) went to the 2nd position because it is must closer to the image centre than site (66 , -50 , 10)..

On the other hand, the 2 best sites on the 40th iteration are completely different on both Tables. The sites on Table 4 are not close to the centre of the image and when the Distance criterion is set to very high the model tends to rate higher sites closer to the centre of the image. Hence, the best sites on Table 5 are closer to centre of the image (see Figure 5).
It should be noted that the experiences discussed here were adaptations of tests done within a research project [7] [8]. Furthermore, here we do not discuss the results rating values, because we are not addressing the aggregation methods used to evaluate the alternatives.

Table 5 – Rating and 3D coordinates of 3 best sites on last iterations of 2nd and 3rd phases (weight for Distance Criterion =Very high).

Iter.	3D coord (meters)	R_n
25 (2nd Phase)	(86 , -27 , 12)	0.7991
	(95 , -26 , 12)	0.7938
	(66 , -50 , 10)	0.7932
40 (3rd Phase)	(88 , -31 , -3)	0.9979
	(78 , -30 , -3)	0.9977
	-	-

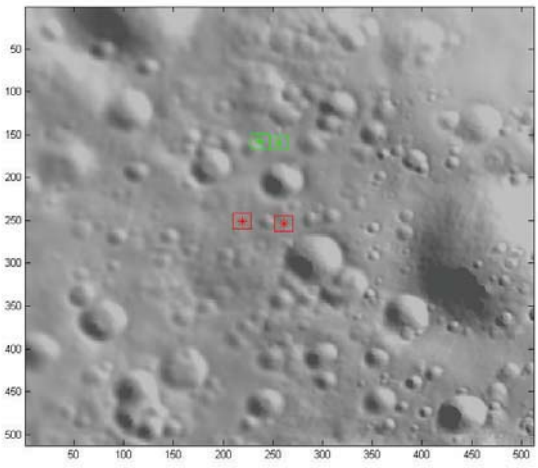


Figure 5 – Image of the surface on iteration 40. Green stars best sites from Table 4 and red stars from Table 5.

In Figure 5 we show the results, just discussed for the last iteration (40). It is quite notorious the influence of changing criterion weights for Distance, from average to very-high, on the last phase. Raising the importance of that criterion "forces" the final decision choice to be closer (shortest distance) to the initial target (center of image), even though the surface is more "bumpy" than the choices from Phase II.

5 Conclusions

In this work we presented a general architecture to cope with dynamic multi-attribute decision problems. We focused our attention on two spatial-temporal aspects of the rating process, .specifically proposing contributions for dealing with uncertainty in dynamically changing input data and also for dealing with dynamic changes in criteria importance.
Potential applications of this work could be on medical diagnose decision support and/ or fault, detection and isolation (FDI) problems.

Acknowledgment

The authors wish to thank Yannick Devassoux from Astrium EADS for his assistance with understanding the case study domain. This work was partially funded by EADS Astrium contract 4572019617.

ISBN: 978-989-95079-6-8

References

[1] Ribeiro, R.A., Fuzzy multiple attribute decision making: A review and new preference elicitation techniques. *Fuzzy Sets and Systems* 78(2): 155-181, 1996

[2] E. Triantaphyllou, *Multi criteria decision making method: a comparative study*, Kluwer academic publishers, Applied optimization series, Vol. 44, 2002.

[3] Chen, S.-J. and C.-L. Hwang. Fuzzy Multiple Attribute Decision Making, *Lecture Notes in Economics and Mathematical Systems*, Springer-Verlag, 1992.

[4] Ross, T. (2004). *Fuzzy Logic with Engineering Applications*, John Wiley & Sons.

[5] Townsend, J. T. & Busmeyer, J. Dynamic representation of decision making. In: R. Port and T. Van Gelder (Eds.), *Mind in Motion: Dynamics, Behavior and Cognition*. 4th chapter, MIT Press, 1995.

[6] S. Khanmohammadi, R. A. Ribeiro, J. Jassbi. Multi criteria decision making using dynamics of criteria. Proceedings of the 11th Mediterranean Conference on Control and Automation (MED03), June 2003, med.ee.nd.edu/MED11/pdf/papers/t3-013.pdf

[7] T. C. Pais, R.A. Ribeiro, Y. Devouassoux, S. Reynaud Dynamic ranking algorithm for landing site selection. In: Proceedings of the Int. Conference on Information Processing and Management of Uncertainty (IPMU´08), Malaga, June 2008.

[8] Y. Devouassoux, S. Reynaud, G. Jonniaux, R. A. Ribeiro, T. C. Pais. Hazard avoidance developments for planetary exploration. In: Proceedings of the 7th International ESA Conference on Guidance, Navigation & Control Systems, Tralee, Ireland,2-5 June 2008

[9] Richardson, G. P. & Pugh III, A. L. *Introduction to System Dynamics Modeling with DYNAMO*. Portland, Oregon: Productivity Press, 1981.

[10] R. Marques-Pereira, Rita A. Ribeiro. Aggregation with generalized mixture operators using weighting functions. *Fuzzy Sets and Systems*, 137 (2003) 43-58.

[11] Rita A. Ribeiro, R. Marques-Pereira. Generalized mixture operators using weighting functions: a comparative study with WA and OWA. *European Journal of Operational Research*, 145 (2), March 1 (2003) 329-342.

ISBN: 978-989-95079-6-8

A Fuzzy Set Approach to Ecological Knowledge Discovery

Arkadiusz Salski

Institute of Computer Science, University of Kiel
Kiel, Germany
E-mail: asa@email.uni-kiel.de

Abstract— *Besides the problem of searching for effective methods for extracting knowledge from large databases (KDD) there are some additional problems with handling ecological data, namely heterogeneity and uncertainty of these data. A fuzzy set approach can be used to handle these problems at some stages of the knowledge discovery process. Ecological data can be defined as fuzzy sets without sharp boundaries, which reflect better the continuous nature of ecological parameters. The paper focuses on one of the important methods of data reduction, namely clustering, and on the data transformation and construction of a combining operator. Two support systems developed at the University of Kiel and their applications are presented, namely the Fuzzy Clustering System ECOFUCS and the Fuzzy Evaluation and Kriging System FUZZEKS.*

Keywords— ecological data, fuzzy clustering, fuzzy data transformation, uncertainty of ecological data.

1 Some properties of ecological data

Dealing with a very large data basis (long time series, spatial data with high resolution, etc.) is a typical problem in data mining and knowledge discovery, but there are some additional problems with handling the environmental data. The first one is the heterogeneity of ecological data sources (e.g. sources of quantitative and qualitative information). The next problem in ecological data mining is the large inherent uncertainty of these data that results not only from the presence of random variables but also from the difficult comparability of these data, approximate estimations, and imprecision and the subjectivity of the information obtained from an expert [12, 13]. This paper deals with processing both "subjective" data derived from experts and "objective" measurement data. Statistical or stochastic aspects of the uncertainty problem are not taken into account.

Some requirements for methods of searching for ecological knowledge arise from the properties of ecological data mentioned above. Special methods of data analysis should be used to handle the uncertainty and heterogeneity of these data. The fuzzy approach, as a possible way to handle uncertainty, is particularly useful for processing imprecise or uncertain data. Ecological data can be defined as fuzzy sets or fuzzy clusters without sharp boundaries. That reflects better the continuous nature of environmental parameters.

Knowledge discovery is a very complex process, which includes data cleaning, data integration, data reduction, data transformation, data mining, pattern evaluation and knowledge presentation [9]. Collaboration with an expert in the data domain can be very useful at some stages of this process. The fuzzy approach enables us to integrate the expert knowledge in the knowledge discovery process. The paper focuses on the applications of fuzzy sets and fuzzy operators at some stages of this process, namely data reduction (section 2) and constructing a combining attribute (section 3).

2 Data reduction: a fuzzy clustering approach

Clustering belongs to the most popular methods of numerosity reduction of data sets, i.e. "replacing" the data set by smaller representations such as clusters in order to reduce the size of the data set. The clustering methods are based on the principle of maximizing the intraclass similarity of objects and minimizing the interclass similarity of these objects, i.e. objects within a cluster have a high similarity, but are very dissimilar to objects in other clusters. Conventional clustering methods based on Boolean logic ignore the continuous nature of ecological parameters and the uncertainty of data. That can result in misclassification.

Fuzzy clustering methods provide additional information, namely the distribution of the membership values which can be interpreted as a similarity measure. The common fuzzy clustering methods, like the fuzzy c-means method, work only with crisp data, that means they provide the fuzzy partition only for crisp data (e.g. exact measurement data). In ecology we have often to deal with data with a semblance of accuracy. In such cases it may only be possible to obtain estimates of data scatter which can be treated in the context of fuzzy sets and used for defining fuzzy data in the form of fuzzy vectors in a high dimension, the so-called conical fuzzy vectors [5]. They are defined by the apex and the so-called panderance matrix which describes the accuracy of the data. This matrix contains spreads of data in each dimension on its diagonal. Yang [15] defined the distance between two conical fuzzy vectors, $\tilde{A}$ and $\tilde{B}$ in (1), as follows:

$$d^2(\tilde{A},\tilde{B}) = \left\|\bar{a}-\bar{b}\right\|^2 + tr\left((A-B)^T(A-B)\right) \qquad (1)$$

where:

A and B are the so-called panderance matrices of $\tilde{A}$ and $\tilde{B}$,
$\left\|\bar{a}-\bar{b}\right\|$ is the distance norm (metric) between the apexes $\bar{a}$ and $\bar{b}$, and the trace
$tr\left((A-B)^T(A-B)\right)$ is the diagonal sum of $\left((A-B)^T(A-B)\right)$.

Yang proved that $d^2(\tilde{A},\tilde{B})$ defined above is a complete metric, which is an assumption for the convergence of the

ISBN: 978-989-95079-6-8

fuzzy c-means clustering procedure by Bezdek [3]. That means, we can define the well known objective function of the fuzzy c-means procedure for conical fuzzy vectors by

$$F_{(c)} = \sum_{i=1}^{n} \sum_{j=1}^{c} \left(\mu_{ij}\right)^m d_c^2\left(\tilde{A}_i, \tilde{B}_j\right) \qquad (2)$$

where:

$\tilde{A}_i$ is the *i*th object and $\tilde{B}_i$ is the *j*th cluster, both defined as conical fuzzy vectors,

n is the number of objects, and

c is the number of clusters.

The clustering algorithm for conical vectors proposed by Yang has been extended for the diagonal norm using the so-called z-transformation of the Euclidian norm and implemented for the Fuzzy Clustering System EcoFucs v.5.1 developed at the University of Kiel [8]. The diagonal norm is a highly recommendable distance measure in the case of heterogeneous ecological data with different domain scales. In such cases we can transform data in a uniform manner before we start the fuzzy c-means procedure for conical vectors:

$$z - trans\left(\tilde{A}_i\right) = \frac{\tilde{A}_i - \tilde{V}}{\tilde{S}} \qquad (3)$$

where:

$\tilde{V} = \frac{1}{n}\sum_{j=1}^{n} \tilde{A}_j$ is the mean vector of all fuzzy conical vectors of the input data set, and

$\tilde{S}$ is the vector of spreads from the panderance matrix.

To obtain back the coordinates of the cluster centers in the real scale we have to apply the inverse transformation of the results of the fuzzy c-means procedure. EcoFucs works also with crisp data and offers four different distance norms as a measure of similarity between an object and a respective cluster (the Euclidean-, the Diagonal-, the Mahalonobis- and L1-norms) and a set of methods for calculating the start partition (WARD, conventional c-means, maximum-distance-algorithm, sharp or fuzzy random partitions). The choice of the distance norm depends on the data set. The partition efficiency indicators (entropy, partition coefficient, payoff and non-fuzziness index) available in EcoFucs can be very helpful in searching for the optimal partition and finding the objects which can serve as the representatives of each cluster (see the example below) [10].

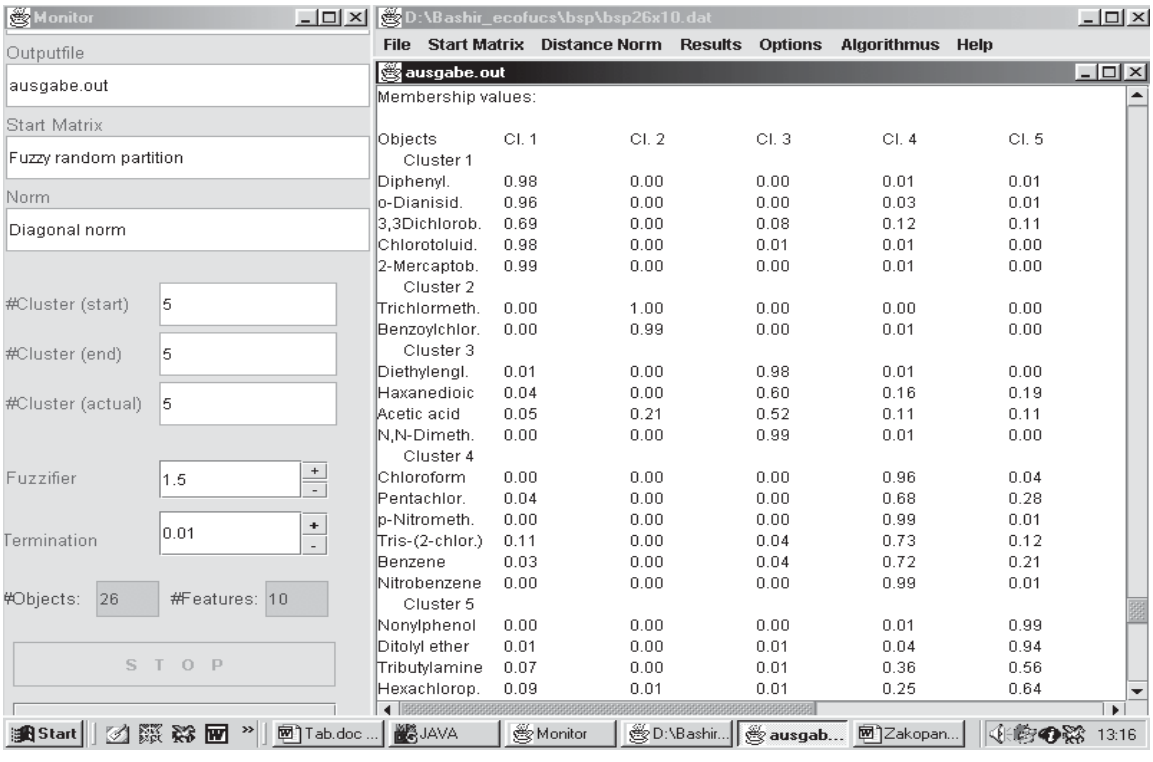


Figure 1: The main window of EcoFucs [8] with a small part of the results of the clustering of ecotoxicological data into 5 clusters.

The fuzzy clustering of chemicals according to their ecotoxicological properties [7] can be mentioned here as an example of data reduction. Both the large number of existing chemicals and the costs of ecotoxicological testing procedures make it necessary to select representative chemicals which faithfully reflect the relevant properties of possibly a major group of compounds. So the main tasks of this application were to find distinguishable clusters with characteristic properties and to find chemicals representative for each cluster. These representatives can be used for further data mining and other steps of the knowledge discovery process instead of all chemicals which belong to the cluster. Compared to conventional clustering methods a fuzzy clustering technique is more appropriate to handle the uncertainty of ecotoxicological data. The degree of uncertainty of these data is very high and can arise, for

ISBN: 978-989-95079-6-8

example, from a mixture of quantitative and qualitative data or from the difficult comparability of these data because of different measurement or test conditions (e.g. test results for different animals). The distribution of the membership values provides information from which the degree of similarity between the properties of a particular chemical and the properties characterising particular clusters can be deduced. This is particularly important since there are many chemicals with more or less overlapping properties. That would not be recognised by conventional clustering methods.

The analysis of the membership values helps us to find the representatives of each cluster. Their membership values are close (or equal) to 1. Figure 1 presents a small example of the results of the clustering of chemicals into 5 clusters for 10 ecotoxicological features (toxicity indicators and the potentials for biodegradability and hydrolysis). Trichlormethylbenzene and 2-Mercaptobenzothiazole, for example, can be taken as the representatives of clusters 2 and 1, respectively. We can also see some chemicals (e.g. Tributylamine and Hexachloropentadiene in cluster 5) with membership values strongly divided between different clusters. These values can be interpreted as the degree of similarity to the respective clusters.

3 Data transformation and the construction of a combining operator

One of the important stages of the knowledge discovery process is the data transformation needed to combine the data. In the case of heterogeneous data we have to normalise these data in a uniform manner before we construct the combining operator [4]. In order to do this we can scale data so that they fall within the same range. We can use the membership functions of fuzzy sets to transform the data into the interval 0.0 to 1.0. The definition of the membership functions of these fuzzy sets should express the evaluation criterion formulated by an expert.

We can consider an analysis of hydrogeological spatial data as an example. Figure 2 presents the membership function of the feature “large water table depth” used for the data normalization in the analysis of the suitability of a specified land unit as a waste disposal site [2, 14]. The shape of the defined membership function corresponds to the evaluation criterion. The values of the water table depth lower than 2 m are not suitable (the membership values equal 0); the values bigger than 5 m are very suitable for a waste disposal site (the membership values equal 1).

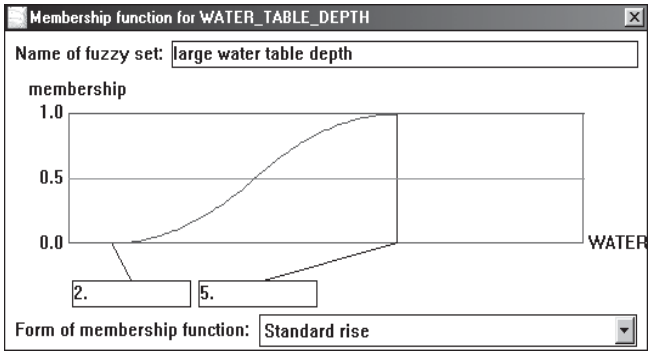


Figure 2: The membership function of the fuzzy set “large water table depth” used for the data transformation (the membership function window of FUZZEKS [2]).

Four land characteristics, namely water table depth, hydrologic conductivity, clay content and Cl concentration, are taken into account in this example. The values of these four features were transformed into a common scale 0.0 to 1.0 using suitable membership functions, like “large water table depth” or “high Cl concentration”. Now, we can combine these data by constructing a new combining attribute (Fig. 3), namely the joint degree of land suitability for a specified utilization (in this case, as a waste disposal site). Different logical and arithmetical operators can be used for the construction of such a combining attribute. Arithmetical operators (like the sum operator) can be weighted and that makes them particularly useful to express the degree of importance of a particular parameter. The distribution of weights is subjectively determined by a domain expert. In our example we combined the four land characteristics by means of a weighted sum operator and the “and”- operator using the Fuzzy Evaluation and Kriging System FUZZEKS developed at the University of Kiel (Fig. 3) [2].

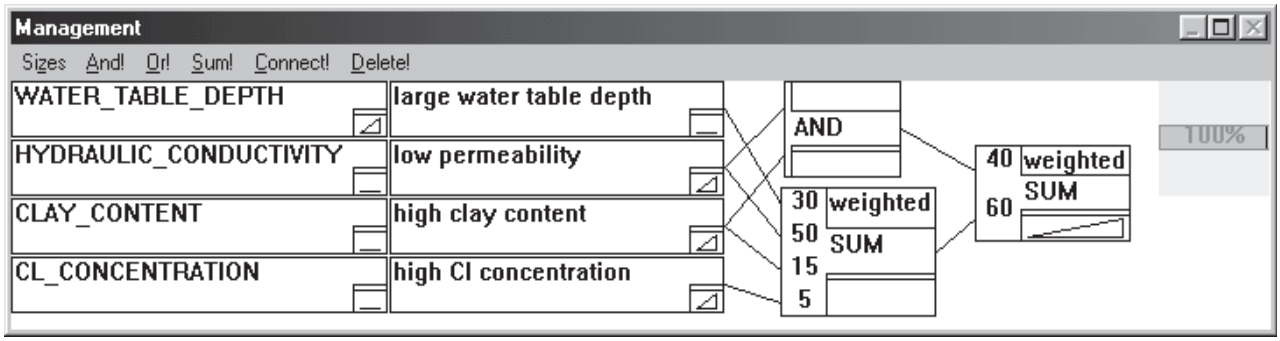


Figure 3: Constructing a combining operator (the management window of FUZZEKS [2]).

The calculated values of the joint attribute can be presented in the form of isolines using the fuzzy interpolation procedure, the so-called fuzzy kriging (Fig. 4). Fuzzy kriging is an extension of the conventional kriging procedure [1,6]. The application of the conventional methods of spatial interpolation is often restricted owing to insufficient amounts

ISBN: 978-989-95079-6-8

of data [11]. If the gathering of new data is too expensive or impossible, we can consider the use of additional imprecise data subjectively estimated by an expert. The fuzzy kriging procedure utilizes exact (crisp) measurement data as well as imprecise estimates obtained from an expert. These imprecise data can be defined as fuzzy numbers and taken as additional input data for the kriging procedure implemented in FUZZEKS. To simplify the preparation of the input data file a special ASCII-file format was implemented, combining both exact (crisp) and fuzzy data (fuzzy numbers) in one unified form [2]. Fuzzeks supports a user in the preparation of the so-called experimental variogram and in the interactive fitting of the crisp theoretical variogram, which is a basis for the interpolation procedure, to the fuzzy experimental variogram (see the small window left in Fig. 4).

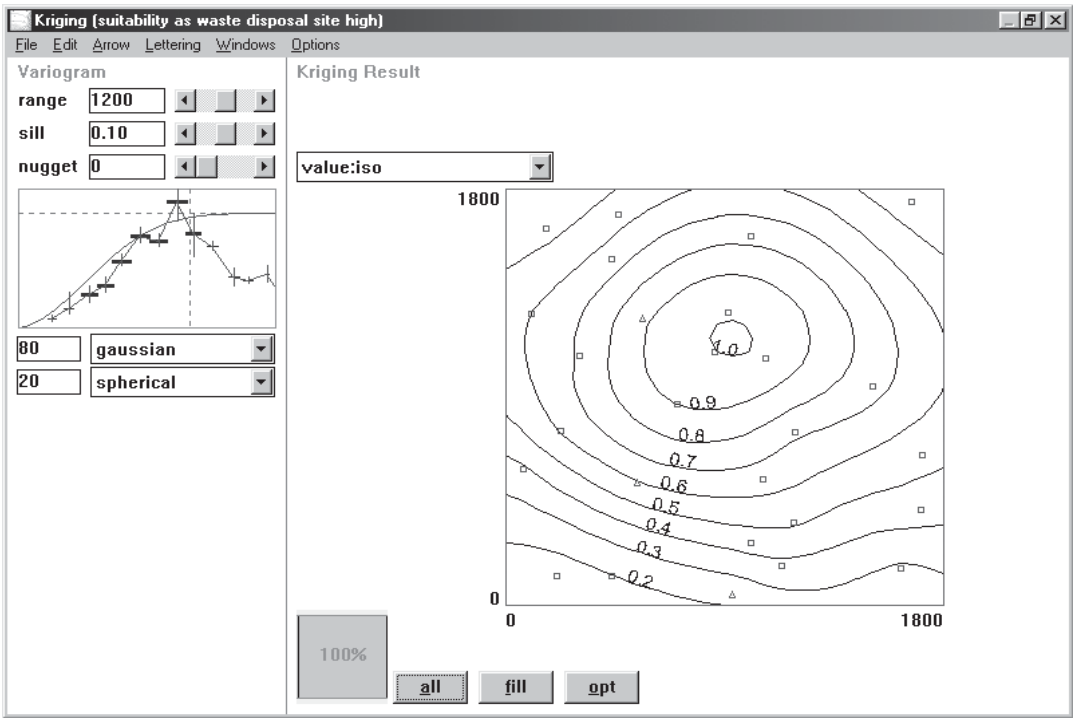


Figure 4: The presentation of the interpolated values of the joint attribute in the form of isolines (on the right in the main window of FUZZEKS,); interactive fitting of a theoretical variogram to an experimental variogram (small window left).

The logical structure of this fuzzy kriging procedure is shown in Fig. 5. The zigzag lines mark the stages with fuzzy data input in the form of fuzzy numbers. At two stages fuzziness is introduced into the calculation. First, fuzziness in the input values causes fuzziness in the experimental variogram. An expert takes the experimental variogram and its fuzziness into account when fitting the crisp theoretical variogram. Second, the fuzzy input values are used at the final step of kriging. Therefore, the kriging results are also fuzzy.

The main fuzzy kriging estimation is a linear combination of the input values and can be calculated using the extension principle and the α -cut-representation of fuzzy sets:

$$Z^*(x)_\alpha = \sum_{i=1}^{n} \delta_i(x) Z(x_i)_\alpha \qquad (4)$$

where:

$Z^*(x)_\alpha$ is is the α -cut of the interpolated value $Z^*(x)$ at the position *x*,

$Z(x_i)_\alpha$ are the α -cuts of the input values $Z(x_i)$, and

$\delta_i(x)$ are the crisp minimizing parameters.

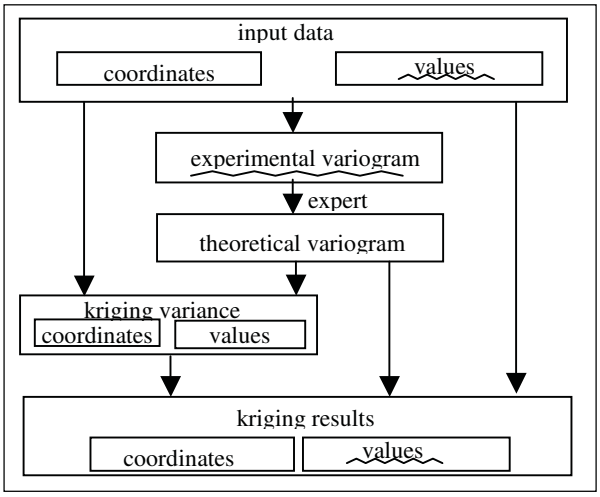


Fig. 5. Logical structure of fuzzy kriging (zigzag lines indicate the fuzziness of data).

ISBN: 978-989-95079-6-8

The kriging estimation (4) formulated by α-cut-representation of fuzzy sets has been used for the implementation of the fuzzy kriging procedure for the Fuzzy Kriging and Evaluation System FUZZEKS.

4 Final remarks

The fuzzy approach can support users in discovering interesting knowledge in uncertain ecological data at different stages of this process. The searching for representatives of groups of ecological objects by means of clustering methods can be very useful not only in numerosity reduction but also in dimensionality reduction of the data set. Fuzzy clusters of objects without sharp boundaries reflect better the continuous nature of ecological features. That enables a better interpretation of the data structure.

Heterogeneous and imprecise ecological data and vague expert knowledge can be integrated more effectively using a fuzzy approach. The data transformation by means of fuzzy sets can be used for this integration. The membership functions of fuzzy sets can describe the evaluation criteria (e.g. "high clay content") defined by an expert. The construction of a combining attribute in the next step reduces the dimensionality and the size of the data set.

And finally, the intention of the author was to draw the reader's attention to a large field for applications, namely ecology. The paper illustrates it briefly using some short application examples. The development of the easy to use tools (like EcoFucs and FUZZEKS developed at the University of Kiel) can be very helpful for the promotion of fuzzy methods in ecological applications.

References

[1] A. Bárdossy, I. Bogardi and W. E. Kelly, Geostatistics utilizing imprecise (fuzzy) information. *Fuzzy Sets and Systems*, 31/3: 311-327, 1989.

[2] F. Bartels, *Ein Fuzzy-Auswertungs- und Krigingsystem für raumbezogene Daten*. Diplomarbeit, Inst. für Informatik und Praktische Mathematik, Universität Kiel, 1997.

[3] J. C. Bezdek, A convergence theorem for the fuzzy c-means clustering algorithms. *IEEE Trans. PAMI*, PAMI-2(1): 1-8, 1980.

[4] Burrough P.A., Macmillan R.A. and Van Deursen, Fuzzy classification methods for determination land suitability from soil profile observations and topography. *Journal of Soil Science,* 43: 193-210, 1992.

[5] Celmiņš, A., Least Squares model fitting to fuzzy vector data. *Fuzzy Sets an Systems*, 22: 245-269, 1987.

[6] Diamond, P., Fuzzy kriging. *Fuzzy Sets and Systems*, 33/3: 315-332, 1989.

[7] Friderichs M., Fränzle O. and Salski A., Fuzzy clustering of existing chemicals according to their ecotoxicological properties. *Ecological Modelling*, 85/1: 27-40, 1996.

[8] Gömann, M., *Erweiterung des Fuzzy-Clustering-Systems EcoFucs*, Studienarbeit, Institut für Informatik, Christian-Albrechts-Universität zu Kiel, 2002.

[9] Han J. and Kamber M., *Data Mining. Concepts and Techniques*. Morgan Kaufman Publishers, San Francisco, 2001.

[10] Krenawi M., *Entwicklung eines plattformunabhängigen Systems zur Fuzzy-Clusteranalyse*. Diplomarbeit, Inst. für Informatik und Praktische Mathematik, Universität Kiel, 2001.

[11] Piotrowski, J.A., Bartels, F., Salski, A. and Schmidt, G., Geostatistical regionalization of glacial aquitard thicknessin northwestern Germany, based on fuzzy kriging., *Mathematical Geology*, 28/4: 437-452, 1996.

[12] Salski, A., Ecological modeling and data analysis. In: H.-J. Zimmermann: *Practical application of fuzzy technologies. The Handbooks of Fuzzy Sets*, Kluwer, , pp. 247-265, 1999.

[13] Salski, A., Ecological applications of fuzzy logic. In: F. Recknagel (ed): *Ecological Informatics,* Springer, pp. 3-14, 2002.

[14] Salski, A., Bartels, F., A fuzzy approach to land evaluation. *IASME Transactions,* 5/2: 774-780, 2005.

[15] Yang, M.-S. and Liu, H-H, Fuzzy clustering procedures for conical fuzzy vector data, *Fuzzy Sets and Systems*, 106: 189-200, 1999.

ISBN: 978-989-95079-6-8

Numerical Solution of Interval Differential Equations with generalized Hukuhara differentiability

Barnabas Bede, Luciano Stefanini
Department of Mathematics, University of Texas-Pan American, USA. email: bedeb@utpa.edu.
Department of Economics and Quantitative Methods, University of Urbino "Carlo Bo", Italy. email: lucste@uniurb.it

Abstract— *In the present paper we propose novel numerical methods for interval differential equations with generalized Hukuhara differentiability. The algorithms proposed here are based on characterization theorems of interval differential equations with ODEs. Usng the characterizations we can translate an interval problem into two ODE systems. We observe that an interval problem may have several solutions under the interpretation considered. Several examples are presented and discussed.*

Keywords— Generalized Hukuhara Differentiability, Interval Differential Equations, Interval Analysis.

1 Introduction

Interval Analysis was introduced as an attempt to handle interval (nonstatistical) uncertainty that appears in our mathematical or computer models of some deterministic real-world phenomena. The first monograph dealing with interval analysis is the book of R. Moore, [10]. Since then, Reliable Computing, Validated Numerics and Interval problems with Differential Equations are discussed in several monographs and research papers ([11], [2]).

Another major approach to a set of similar problems is that of differential inclusions and multivalued analysis ([1], [5]). This approach is also able to deal with discontinuous dynamical systems which do not fully fit into the Interval Analysis topic.

Hukuhara derivative of a set-valued mapping was first introduced by Hukuhara and it has been studied in several works. The paper of Hukuhara was the starting point for the topic of Set Differential Equations and later also for Fuzzy Differential Equations [8], [12], [6]. Recently, several works as e.g.[9], have brought back into the attention of the nonlinear analysis community, the topics of set differential equations and the Hukuhara derivative.

Recently a generalization of the Hukuhara difference for compact convex sets (gH-difference) was presented in [13] and the strongly and weakly generalized (Hukuhara) differentiability concepts have been proposed in [3]. Then, in [14] a combination (gH-derivative) was proposed and it was shown to be a useful concept in the interval setting. Our approach to interval differential equations is different from the approaches based on differential inclusions or interval analysis, however we can see that there is some connection among these approaches and the presented one, some connections with differential inclusions we discuss here.

After recalling some results of [14], we propose a numerical solution method for IDEs. Also, a numerical method for differential inclusions is discussed and several examples are proposed.

2 Generalized Hukuhara difference

Consider the space $\mathbb{I}$ of real intervals. Given two elements $A, B \in \mathbb{I}$ and $k \in \mathbb{R}$, the usual interval arithmetic operations, i.e. Minkowski addition and scalar multiplication, are defined by $A+B = \{a+b | a \in A, b \in B\}$ and $kA = \{ka | a \in A\}$. It is well known that addition is associative and commutative and with neutral element $\{0\}$. If $k = -1$, scalar multiplication gives the opposite $-A = (-1)A = \{-a | a \in A\}$ but, in general, $A + (-A) \neq \{0\}$, i.e. the opposite of A is not the inverse of A in Minkowski addition (unless $A = \{a\}$ is a singleton). A first implication of this fact is that, in general, additive simplification is not valid, i.e. $(A + C = B + C) \nRightarrow A = B$ or $(A + B) - B \neq A$ (the Minkowski difference is $A - B = A + (-1)B$).

To partially overcome this situation, the Hukuhara H-difference has been introduced as a set C for which $A \ominus B = C \iff A = B + C$ and an important property of $\ominus$ is that $A \ominus A = \{0\}\, \forall A \in \mathcal{K}_C^n$ and $(A+B) \ominus B = A, \forall A, B \in \mathcal{K}_C^n$.

The H-difference is unique, but it does not always exist (a necessary condition for $A \ominus B$ to exist is that A contains a translate $\{c\} + B$ of B).

A generalization of the Hukuhara difference proposed in [13] aims to overcome this situation.

Definition 1 *([13]) The generalized Hukuhara difference of two sets $A, B \in \mathbb{I}$ (gH-difference for short) is defined as follows*

$$A \ominus_g B = C \iff \begin{cases} (a) & A = B + C \\ \textit{or} \quad (b) & B = A + (-1)C \end{cases} \tag{1}$$

The following property ensures the existence of gH-differences.

Proposition 1 *([13]) The gH-difference $C = A \ominus_g B$ of two intervals $A = [a^-, a^+]$ and $B = [b^-, b^+]$ always exists and*

$$[a^-, a^+] \ominus_g [b^-, b^+] = [c^-, c^+] \tag{2}$$

with

$$c^- = \min\{a^- - b^-, a^+ - b^+\}$$
$$c^+ = \max\{a^- - b^-, a^+ - b^+\}.$$

Conditions (a) and (b) in (1) are satisfied simultaneously if and only if the two intervals have the same length and $c^- = c^+$.

3 Differentiation of interval valued functions

There exist several alternative definitions for the derivative of an interval-valued function. The first concept is a partic-

ISBN: 978-989-95079-6-8

ularization of the fuzzy concepts presented in [3] to the interval case, based on the generalized fuzzy differentiability concept, more general than one-sided derivatives, using the usual Hukuhara difference $\ominus$.

Definition 2 *([3]) Let $f :]a,b[\to \mathbb{I}$ and $x_0 \in]a,b[$. We say that f is strongly generalized (Hukuhara) differentiable at x_0, if there exists an element $f'(x_0) \in \mathbb{I}$, such that, for all $h > 0$ sufficiently small,*

(i) $\exists f(x_0+h) \ominus f(x_0)$, $f(x_0) \ominus f(x_0-h)$ and

$$\lim_{h \searrow 0} \frac{f(x_0+h) \ominus f(x_0)}{h} = \lim_{h \searrow 0} \frac{f(x_0) \ominus f(x_0-h)}{h} = f'(x_0),$$

or (ii) $\exists f(x_0) \ominus f(x_0+h)$, $f(x_0-h) \ominus f(x_0)$ and

$$\lim_{h \searrow 0} \frac{f(x_0) \ominus f(x_0+h)}{(-h)} = \lim_{h \searrow 0} \frac{f(x_0-h) \ominus f(x_0)}{(-h)} = f'(x_0),$$

or (iii) $\exists f(x_0+h) \ominus f(x_0)$, $f(x_0-h) \ominus f(x_0)$ and

$$\lim_{h \searrow 0} \frac{f(x_0+h) \ominus f(x_0)}{h} = \lim_{h \searrow 0} \frac{f(x_0-h) \ominus f(x_0)}{(-h)} = f'(x_0),$$

or (iv) $\exists f(x_0) \ominus f(x_0+h)$, $f(x_0) \ominus f(x_0-h)$ and

$$\lim_{h \searrow 0} \frac{f(x_0) \ominus f(x_0+h)}{(-h)} = \lim_{h \searrow 0} \frac{f(x_0) \ominus f(x_0-h)}{h} = f'(x_0).$$

Based on the gH-difference in [14] the following definition was proposed.

Definition 3 *Let $x_0 \in]a,b[$ and h be such that $x_0 + h \in]a,b[$, then the gH-derivative of a function $f :]a,b[\to \mathbb{I}$ at x_0 is defined as*

$$f'(x_0) = \lim_{h \to 0} \frac{1}{h}[f(x_0+h) \ominus_g f(x_0)]. \quad (3)$$

If $f'(x_0) \in \mathbb{I}$ satisfying (3) exists, we say that f is generalized Hukuhara differentiable (gH-differentiable for short) at x_0.

The next result is a characterization of the derivative concept proposed in (3) and it was proved in [14].

Theorem 2 *Let $f : [a,b] \longrightarrow \mathbb{I}$ be such that $f(x) = [f^-(x), f^+(x)]$. The function $f(x)$ is gH-differentiable if and only if $f^-(x)$ and $f^+(x)$ are differentiable real-valued functions and*

$$f'(x) = [\min\{(f^-)'(x), (f^+)'(x)\}, \quad (4)$$
$$\max\{(f^-)'(x), (f^+)'(x)\}]. \quad (5)$$

The following definition will be helpful in the discussion of the numerical methods for IDEs.

Definition 4 *Let $f : [a,b] \longrightarrow \mathbb{I}$ be gH-differentiable at $x_0 \in]a,b[$. We say that f is (i)-gH-differentiable at x_0 if*

$$\textit{(i.)} \qquad f'(x_0) = [(f^-)'(x_0), (f^+)'(x_0)] \quad (6)$$

and that f is (ii)-gH-differentiable at x_0 if

$$\textit{(ii.)} \qquad f'(x_0) = [(f^+)'(x_0), (f^-)'(x_0)]. \quad (7)$$

It is an interesting problem to see how the switch between the two cases (i.) and (ii.) in Definition 4 can occur.

Definition 5 *We say that a point $x_0 \in]a,b[$ is an l-critical point of f if it is a critical point for the length function $len(f(x)) = f^+(x) - f^-(x)$.*

Definition 6 *We say that a point $x_0 \in]a,b[$ is a switching point for the differentiability of f, if in any neighborhood V of x_0 there exist points $x_1 < x_0 < x_2$ such that*

type I) at x_1 (6) holds while (7) does not hold and at x_2 (7) holds and (6) does not hold, or

type II) at x_1 (7) holds while (6) does not hold and at x_2 (6) holds and (7) does not hold.

First, let us observe that any switching point is also an *l-critical* point. Indeed, if x_0 is a switching point then $[(f^-)'(x_0), (f^+)'(x_0)] = [(f^+)'(x_0), (f^-)'(x_0)]$ and so $(f^+)'(x_0) = (f^-)'(x_0)$ and $len(f(x_0))' = 0$. Clearly, not all *l-critical* points are also switching points.

The relation which exists between the two differentiability concepts was given in [14].

Theorem 3 *Let $f :]a,b[\to \mathbb{I}$ be a function $f(x) = [f^-(x), f^+(x)]$. The following affirmations are equivalent:*

(1) f is GH-differentiable

(2) f is gH-differentiable and the set of switching points is finite.

If the number of l-critical points is finite then the number of switching points is also finite. So, from now on we will use the GH-differentiability concept or equivalently the gH-differentiability with a finite number of switching points.

4 Interval differential equations

In this section we consider an interval valued differential equation

$$y' = f(x,y)\ , y(x_0) = y_0 \quad (8)$$
where
$$f : [a,b] \times \mathbb{I} \longrightarrow \mathbb{I}$$
with $f(x,y) = [f^-(x,y), f^+(x,y)]$ for $y \in \mathbb{I}$
$$y = [y^-, y^+],\ y_0 = [y_0^-, y_0^+].$$

We consider only GH-differentiable solutions, i.e. such that there exists $\delta > 0$ such that there are no switching points in $[x_0, x_0 + \delta]$.

An existence result for the solutions of an IDE was obtained in [14]. Let us consider $\overline{B}(y_0, q) \subset \mathbb{I}$ be a closed ball with center y_0 and radius q.

ISBN: 978-989-95079-6-8

Theorem 4 *Let $R_0 = [x_0, x_0 + p] \times \overline{B}(y_0, q)$, $y_0 \in \mathbb{I}$ nontrivial and $f : R_0 \to \mathbb{I}$ be continuous, nontrivial (i.e., y nontrivial interval gives $f(x, y)$ a nontrivial interval). If f satisfies the Lipschitz condition $D(f(x, y), f(x, z)) \leq L \cdot D(y, z)$, $\forall (x, y), (x, z) \in R_0$ then the interval problem*

$$\begin{cases} y' = f(x, y) \\ y(x_0) = y_0 \end{cases}$$

has two unique solutions $y^i, y^{ii} : [x_0, x_0 + r] \to \overline{B}(y_0, q)$ and the successive iterations in

$$y_0(x) = y_0$$

$$y_{n+1}(x) \ominus_g y_0 = \int_{x_0}^{x} f(t, y_n(t))dt,$$

or more precisely,

$$y^i_{n+1}(x) = y_0 + \int_{x_0}^{x} f(t, y^i_n(t))dt \qquad \textit{i.e.}$$

$$(a) \begin{cases} y^{i-}_{n+1}(x) = y_0^- + \int_{x_0}^{x} f^-(t, y^i_n(t))dt \\ y^{i+}_{n+1}(x) = y_0^+ + \int_{x_0}^{x} f^+(t, y^i_n(t))dt \end{cases}$$

and

$$y_0 = y^{ii}_{n+1}(x) - \int_{x_0}^{x} f(t, y^{ii}_n(t))dt \qquad \textit{i.e.}$$

$$(b) \begin{cases} y^{ii-}_{n+1}(x) = y_0^- + \int_{x_0}^{x} f^+(t, y^{ii}_n(t))dt \\ y^{ii+}_{n+1}(x) = y_0^+ + \int_{x_0}^{x} f^-(t, y^{ii}_n(t))dt \end{cases}$$

converge to these two solutions y^i and y^{ii} respectively.

Considering the interval problem

$$\begin{cases} y' = f(x, y) \\ y(x_0) = y_0 \end{cases} \tag{9}$$

we set

$$\begin{cases} \varphi^-(x, y^-, y^+) = f^-(x, y) \\ \varphi^+(x, y^-, y^+) = f^+(x, y) \end{cases} \tag{10}$$

with $\varphi^- \leq \varphi^+$ defined on a subset of $\mathbb{R}^3$.

Finally, we obtain two situations:

ODE(a): $y'^-(x) \leq y'^+(x)$; the differential equations are

$$\begin{cases} y'^-(x) = \varphi^-(x, y^-(x), y^+(x)) \\ y'^+(x) = \varphi^+(x, y^-(x), y^+(x)) \\ \text{s.t.} \\ y^-(x_0) = y_0^- \\ y^+(x_0) = y_0^+ \end{cases} \tag{11}$$

and if φ^- depends only on y^- and φ^+ depends only on y^+, the two equations are independent.

ODE(b): $y'^-(x) \geq y'^+(x)$; the differential equations are

$$\begin{cases} y'^-(x) = \varphi^+(x, y^-(x), y^+(x)) \\ y'^+(x) = \varphi^-(x, y^-(x), y^+(x)) \\ \text{s.t.} \\ y^-(x_0) = y_0^- \\ y^+(x_0) = y_0^+ \end{cases} \tag{12}$$

and if φ^+ depends only on y^- and φ^- depends only on y^+, the two equations are independent.

Now we present a characterization result obtained in [14]:

Theorem 5 *Let $R_0 = [x_0, x_0 + p] \times \overline{B}(y_0, q)$, $y_0 \in \mathbb{I}$ nontrivial and $f : R_0 \to \mathbb{I}$ be nontrivial and continuous. If f satisfies the Lipschitz condition $D(f(x, y), f(x, z)) \leq L \cdot D(y, z)$, $\forall (x, y), (x, z) \in R_0$ then the interval problem*

$$\begin{cases} y' = f(x, y) \\ y(x_0) = y_0 \end{cases} \tag{13}$$

is equivalent to the union of the ODEs (11) and (12) on some interval $[x_0, x_0 + q]$. Here, by equivalent, we understand that $y = [y^-, y^+] : [x_0, x_0 + q] \to \mathbb{I}$ is a solution of 13 if and only if $(y^-, y^+) : [x_0, x_0 + q] \to \mathbb{R}^2$ is a solution of one of the problems (11) or (12).

Remark 1 *We can write the IDE (13) in terms of the midpoint representations of y, y' and $f(x, y)$. Let $y = \langle \widehat{y}, \overline{y} \rangle$, $y_0 = \langle \widehat{y}_0, \overline{y}_0 \rangle$, $f(x, y) = \left\langle \widehat{f}(x, y), \overline{f}(x, y) \right\rangle$; then $y' = \langle \widehat{y}', |\overline{y}'| \rangle$ and (13) becomes*

$$\begin{cases} \widehat{y}' = \widehat{f}(x, y) \\ |\overline{y}'| = \overline{f}(x, y) \\ \widehat{y}(x_0) = \widehat{y}_0 \\ \overline{y}(x_0) = \overline{y}_0; \end{cases} \tag{14}$$

they represent a system of differential algebraic equations (DAEs) and can be investigated by the help of the corresponding theory.

All the preceding results hold for general interval-valued functions so possibly they depend on interval parameters, i.e. interval extensions of real-valued functions of the form $f(x, y, p_1, ..., p_n)$ with $p_i \in P_i \in \mathbb{I}$.

Let us analyze the case when $f(x, y)$ is the interval extension of a real valued function $f(x, p)$, i.e.

$$f(x, y) = [\inf_{p \in y} f(x, p), \sup_{p \in y} f(x, p)]. \tag{15}$$

Let us suppose further that f is monotonic with respect to p.

1) if $f(x, p)$ is increasing with respect to p,

$$case(a) : y'^-(x) \leq y'^+(x)$$
$$\begin{cases} y'^-(x) = f(x, y^-(x)) \\ y'^+(x) = f(x, y^+(x)), \\ \text{s.t. } y^-(x_0) = y_0^-, \\ y^+(x_0) = y_0^+ \end{cases} \tag{16}$$

$$case(b) : y'^-(x) \geq y'^+(x)$$
$$\begin{cases} y'^+(x) = f(x, y^-(x)) \\ y'^-(x) = f(x, y^+(x)), \\ \text{s.t. } y^-(x_0) = y_0^-, \\ y^+(x_0) = y_0^+ \end{cases} \tag{17}$$

2) if $f(x, p)$ is decreasing with respect to p

$$case(a) : y'^-(x) \leq y'^+(x)$$
$$\begin{cases} y'^-(x) = f(x, y^+(x)) \\ y'^+(x) = f(x, y^-(x)), \\ \text{s.t. } y^-(x_0) = y_0^-, \\ y^+(x_0) = y_0^+ \end{cases} \tag{18}$$

$$case(b) : y'^-(x) \geq y'^+(x)$$
$$\begin{cases} y'^+(x) = f(x, y^+(x)) \\ y'^-(x) = f(x, y^-(x)), \\ \text{s.t. } y^-(x_0) = y_0^-, \\ y^+(x_0) = y_0^+ \end{cases} \tag{19}$$

ISBN: 978-989-95079-6-8

Also, in this case, a very interesting result from [4] connects these cases to differential inclusions. In the followings we formulate a particularization of the result of Chalco-Cano and Roman-Flores. Namely we have:

Theorem 6 *[4] Let $f(x,p)$ be a monotonic function with respect to $p \in \mathbb{R}$ and let $y_0 \in \mathbb{I}$ nontrivial interval.*

a) If $f(x,y)$ is the interval extension of f and f is an increasing function, then the solution in case (a) of (16) and the attainable set of the differential inclusion $y' = f(x,y)$, $y(x_0) \in y_0$ coincide on some interval $[x_0, x_0+q]$.

b) If $f(x,y)$ is the interval extension of f and f is a decreasing function, then the solution in case (b) of (18) and the attainable set of the differential inclusion $y' = f(x,y)$, $y(x_0) \in y_0$ coincide on some interval $[x_0, x_0+q]$.

5 Solution methods for interval differential equations

From the results and discussion of previous sections, the interval differential equation concept presented in this paper does not coincide with the concept of a differential inclusion. It is a different, new approach to model interval uncertainty in dynamical systems. It is related (as it is seen from the above theorems) to differential inclusions but we do have in our case more than one solution. The existence of several solutions can be an advantage when we search for solutions with specific properties like e.g., periodic, almost periodic, asymptotically stable, etc. Also, it can be very useful when we have unknown correlations between the parameters. In those situations, the uncertainty about the correlation introduces supplementary uncertainty in our system, so the existence of several solutions appears to be natural in this case.

The above characterization theorems, together with the existence results, lead easily to a numerical algorithm to solve interval differential equations.

5.1 General description of the solution methods

First, let us remark that a switch between the cases (i) and (ii) of gH-differentiability is possible if and only if $y'(x)$ is a singleton as in fact possible switch-points are l-critical points, i.e., critical points of the length of $y(x)$. Let us remark that if at a point $y(x_0)$ is a singleton for some x_0, then this point is enforcing a switch to the case (i), because, according to the existence result in Theorem 4, the second solution does not exist in this case. All other l-critical points make possible a switch from (i) to (ii) case or vice versa, but they do not enforce a switch, so at each l-critical point x_0 with $y(x_0)$ nontrivial interval, two new local solutions arise. One is (i)-differentiable, the other is (ii)-differentiable.

These remarks, together with the characterization Theorem 5, lead to the following general approach to numerically solve interval differential equations. We solve basically ODEs (11) and (12) on subintervals of the time domain having the initial value y_0 updated at all the possible critical points. In this way on a bounded time interval we obtain a finite number of solutions. For the solution of (11) and (12) by the characterization theorem above, any efficient numerical method for ODEs can be used. This is an advantage of the method presented here, as we do not need to reinvent numerical methods for interval differential equations. Instead we can use the classical ones on the ODE translations of the interval differential equations.

This algorithm generates a tree structure for the solutions of the IDE. All the nodes of the tree will be critical points except the terminal nodes. All the branches represent local solutions of (i) or (ii) kind between two nodes. This is illustrated in Figure 1. Each node represents an l-critical point with a switching and each branch corresponds to one of the cases (i) or (ii) of differentiability.

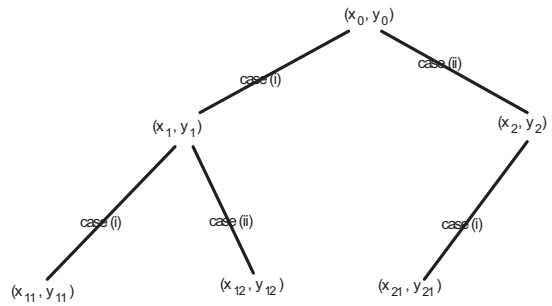


Figure 1. The structure of the tree of local solutions.

Our algorithm which generates all the solutions on a bounded interval is as follows.

Algorithm 1. (Find solutions of an interval differential equation - IDE)

Let us consider y_0 be any interval.

- Step 1. If y_0 is a singleton then we solve (11) and we obtain solution y_1.
- Step 2. Else we solve both (11) and (12) and we obtain solutions y_1, y_2.
- Step 3. Let $x_1 = \inf\{x > x_0 : y_1'(x) \in \mathbb{R}\}$ and $x_2 = \inf\{x > x_0 : y_2'(x) \in \mathbb{R}$ or $y_2(x) \in \mathbb{R}\}$ be the nearest l-critical point. (Let us remark here that since the algorithm applies to general interval-valued functions, it is possible that $y_2(x) \in \mathbb{R}$ however $y_2'(x) \notin \mathbb{R}$)
- Step 4. We insert the solutions which are found in a tree structure: y_1 in the left branch and y_2 in the right branch (the root will be simply (x_0, y_0)).
- Step 5. If $x_1 < X$ (a preset maximum value for x) then let $x_0 = x_1$ and $y_0 = y_1(x_1)$ and go to step 1.
- Step 6. If $x_2 < X$ then let $x_0 = x_2$ and $y_0 = y_2(x_2)$ and go to step 1.
- Step 7. Else Return.
- Step 8. Using a standard backtracking algorithm we generate all the solutions from the tree structure generated above.

The presented algorithm will generate a finite number of solutions on the interval $[x_0, X]$ provided that there are a finite number of critical points. Later we can extract those solutions which are closely reflecting the phenomenon which we have to model.

An approach to interval and fuzzy differential equations has been suggested in [7], based on differential inclusions.

ISBN: 978-989-95079-6-8

When $f(x,y)$ is the interval extension of a continuous function $f(x,p)$, $p \in \mathbb{R}$ and if we are interested in finding the *attainable set* for the differential inclusions $y' = f(x,y)$, $y(x_0) \in y_0$, $x \in [x_0, X]$, then we have a simpler algorithm based on the characterization Theorem 5 and the results shown in Theorem 6.

Let us consider y_0 be any interval. Let $f(x,y)$ be the interval extension of a real function $f(x,p)$. We use the same notation for both functions, and from the context we can identify them.

Algorithm 2. (Find the *attainable set* of a differential inclusion)

- Step 1. We find the points where the real function f changes its monotonicity w.r.t. p (if f is differentiable then we solve $\frac{\partial f}{\partial p} = 0$ and we find the critical points where monotonicity w.r.t. the second variable is changed). Let (x_1, y_1) be a critical point such that in $]x_0, x_1[$ there are no other critical points.
- Step 2. If f is increasing w.r.t. p on $[x_0, x_1] \times [y_0, y_1]$ we solve (11)
- Step 3. Else if f is decreasing w.r.t. p on $[x_0, x_1] \times [y_0, y_1]$ we solve (12)
- Step 4. If $x_1 < X$ then let $x_0 = x_1$ and $y_0 = y(x_1)$ and go to step 2.
- Else Stop.

This algorithm leads to the unique solution (*attainable set*) of the differential inclusion $y' = f(x,y)$, $y(0) \in y_0$.

It is easy to see that the proposed methods are very efficient from the numerical point of view, since for the local solutions we can use any standard algorithm.

5.2 Examples

The above algorithms were implemented in MATLAB. We have used MATLAB's standard ODE solver *ode45*, which is based on a Runge-Kutta (4,5) formula. Surely, any other solver could be used. Let us remark here that the critical points were in all the cases *a priori* determined. The critical points cannot be easily found on the run. The problem is that if we detect a critical point, due to the machine precision the algorithm finds further points which are close to the correct critical point as critical ones. Also, if we have a solution in the case (ii) with length decreasing asymptotically to zero, due to the machine precision the program detects them as critical points.

We will start with a simple example, which is easy to be solved analytically and we can compare the analytical solution to the numerical solution.

5.2.1 Example 1

Let us consider the interval differential equation

$$\begin{cases} y' = -y + [1,2]x \\ y(0) = [0,1] \end{cases} \quad x \in [0,4]. \tag{20}$$

We denote $y = [u, v]$, where u, v are real-valued functions. The systems (11) and (12) are respectively

$$\begin{cases} u' = -v + x \\ v' = -u + 2x \\ u(0) = 0 \\ v(0) = 1 \end{cases} \text{ and } \begin{cases} u' = -u + 2x \\ v' = -v + x \\ u(0) = 0 \\ v(0) = 1 \end{cases} \tag{21}$$

This equation (20) has exactly two solutions. One of them starts with the case (i) of differentiability

$$y_1(x) = [2x - e^x + 2e^{-x} - 1, x + e^x + 2e^{-x} - 2]$$

and there are no critical points in the trajectory.

If we start with case (ii) then we have a critical point of type II at $x = 1$ (i.e., $y(1)$ is a singleton). In this case we have to switch to case (i) of differentiability since the (ii)-differentiable solution does not exist. We obtain

$$u(x) = \begin{cases} 2x + 2e^{-x} - 2 & \text{if} \quad 0 \le x \le 1 \\ 2x - e^{x-1} + 2e^{-x} - 1 & \text{if} \quad 1 \le x \end{cases}$$

$$v(x) = \begin{cases} x + 2e^{-x} - 1 & \text{if} \quad 0 \le x \le 1 \\ x + e^{x-1} + 2e^{-x} - 2 & \text{if} \quad 1 \le x \end{cases}$$

The analytic solution and the numerical solutions obtained by the proposed algorithms are shown in Figs 2. and 3. respectively. We can see that the proposed method is very accurate. The error is controlled by the precision of *ode45* algorithm in MATLAB, and it is less than 10^{-6}.

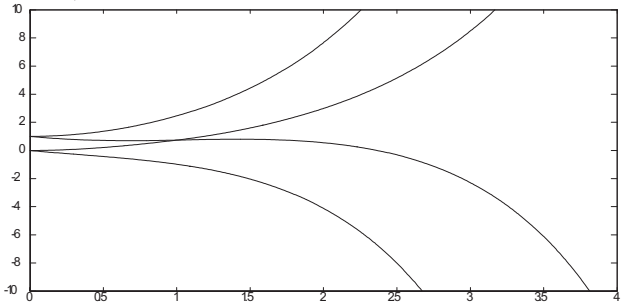


Figure 2. Analytic (exact) solutions of the IDE (20)

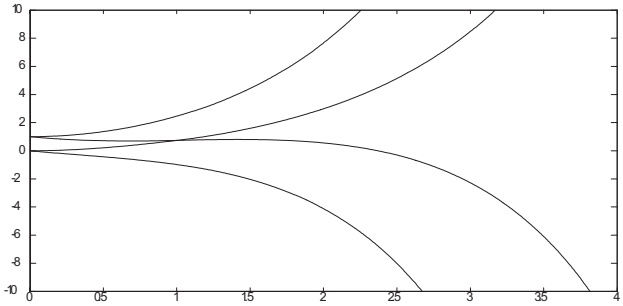


Figure 3. Numerical solutions of (20) by the proposed *Algorithm 1*.

5.2.2 Example 2

The structure of solution of an interval DE can be well illustrated on the problem

$$\begin{cases} y' = y \sin x + 2x, \\ y(0) = [1,3] \end{cases} \tag{22}$$

According to the characterization theorems it can be written as

$$\begin{cases} u'(x) = u(x)\sin(x) + 2x \\ v'(x) = v(x)\sin(x) + 2x \end{cases}$$

ISBN: 978-989-95079-6-8

or

$$\begin{cases} u'(x) = v(x)\sin(x) + 2x \\ v'(x) = u(x)\sin(x) + 2x \end{cases}.$$

The *l-critical points* are at $x = k\pi$, $k \in \mathbb{Z}$. If we solve the problem in the interval $[0, X]$ with $X < 2\pi$ we have only one *l-critical point* in the interval. Then we may have *four* solutions, as illustrated in Figure 4.

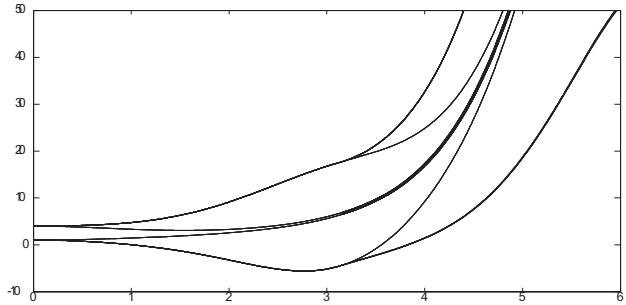


Figure 4. The four solutions of IDE (22) on the interval $[0, 6]$

If we consider time intervals containing more critical points, we may have more solutions. In Figure 5. we illustrate the solutions found by *Algorithm 1* to IDE (22) for $x \in [0, 11]$.

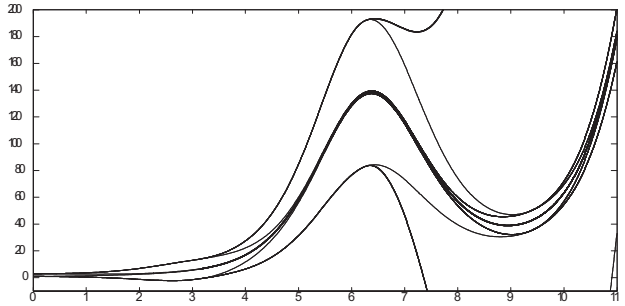


Figure 5. Existence of several solutions for IDE (22) on $[0, 11]$

5.2.3 Example 3

Regarding the second algorithm for solving differential inclusions when the function $f(x, y)$ is the interval extension of a crisp function we present the following problem

$$\begin{cases} y' = y\sin x + 2x, x \in [0, 6] \\ y(0) \in [1, 2]. \end{cases} \quad (23)$$

We observe that the function $y \sin x$ changes monotonicity at the points $k\pi$, $k \in \mathbb{Z}$. Also, we observe that in the interval $[0, \pi]$ it is increasing w.r.t. y. Using the *Algorithm 2* proposed for differential inclusions we obtain the solution presented in Figure 6.

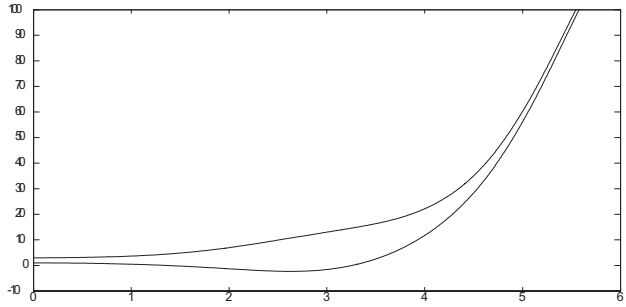


Figure 6. Solution of differential inclusion (23) using the proposed *Algorithm 2*

Remark: It is interesting to remark that the solution (attainable set) of the differential inclusion (23) is obtained by solving one of the ordinary differential equations (11) or (12) in steps 2. and 3. derived from the use of the generalized Hukuhara derivative.

References

[1] J.P. Aubin, A. Cellina, *Differential Inclusions*, Springer-Verlag, Berlin, Heidelberg, New York, Tokyo, 1984.

[2] R. Baker Kearfott, V. Kreinovich (Eds.) *Applications of interval computations*, Kluwer Academic Publishers, 1996.

[3] B. Bede, S.G. Gal, Generalizations of the differentiability of fuzzy-number-valued functions with applications to fuzzy differential equations, *Fuzzy Sets and Systems*, 151(2005), 581-599.

[4] Y. Chalco-Cano, H. Román-Flores, On new solutions of fuzzy differential equations, *Chaos, Solitons & Fractals*, 38(2008), 112-119.

[5] K. Deimling, *Multivalued Differential Equations*. W. De Gruyter, Berlin, 1992.

[6] P. Diamond, P. Kloeden, *Metric Spaces of Fuzzy Sets*, World Scientific, New Jersey, 1994.

[7] H. Hullermeier, An approach to modeling and simulation of uncertain dynamical systems, *Int. Journal of Uncertainty, Fuzziness and Knowledge-based Systems*, 5 (1997) 117-137.

[8] O. Kaleva, Fuzzy differential equations, *Fuzzy Sets and Systems*, 24 (1987) 301-317.

[9] V. Lakshmikantham, A.A. Tolstonogov, Existence and interrelation between set and fuzzy differential equations, *Nonlinear Analysis*, 55 (2003), no. 3, 255-268.

[10] R.E. Moore, *Interval Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1966.

[11] A. Neumaier, *Interval Methods for Systems of Equations*, Cambridge University Press (1990).

[12] M. Puri, D. Ralescu, Differentials of fuzzy functions, *J. Math. Anal. Appl.*, 91(1983), 552-558.

[13] L. Stefanini, A generalization of Hukuhara difference for interval and fuzzy arithmetic, in D. Dubois et Al. (Eds), *Soft Methods for Handling Variability and Imprecision*, Advances in Soft Computing (vol 48), Springer, 2008.

[14] L. Stefanini, B. Bede, Generalized Hukuhara Differentiability of Interval-valued Functions and Interval Differential Equations, *Nonlinear Analysis*, (2009) to appear. doi:10.1016/j.na.2008.12.005.

ISBN: 978-989-95079-6-8

Phi-Calculus fuzzy arithmetic in control: Application to model based control

Laurent Vermeiren[1,2,3], Thierry Marie Guerra[1,2,3], Hakim Lamara[1,2,3]

[1] Univ Lille Nord de France, F-59000 Lille, France
[2] UVHC, LAMIH, F-59313 Valenciennes, France
[3] CNRS, UMR 8530, F-59313 Valenciennes, France
Email: {laurent.vermeiren, guerra}@univ-valenciennes.fr

***Abstract*—*This paper aims at applying a fuzzy arithmetic of intervals calculus and fuzzy quantities to automatic control. This one called φ–calculus fuzzy arithmetic is more practical than the extension principle one and α-cut based methods. It comes from a different representation of fuzzy numbers. The present paper follows up work in introducing. The present paper is interested in its use for a fuzzy internal model control scheme based.*

Keywords— Fuzzy arithmetic, fuzzy model inversion, fuzzy number, weighted fuzzy fusion, model based control.

1 Introduction

The fuzzy set theory elaborated by L.A. Zadeh [1] has been shown to be used in the characterization of fuzziness and/or uncertainty using a coherent mathematical model. Various applications came at hand, in particular in the fuzzy control area [2, 3]. The uncertain calculus was undergoing quite a boom these last years. Several works exhibit interesting results, often referring to intervals theory [4, 5] or arithmetic based on fuzzy numbers such as Triangular or Trapezoidal Fuzzy Numbers [6, 7]. In the fuzzy arithmetic case, the different approaches are generally based either on Zadeh's extension principle [1, 6], either on fuzzy relation [8], or finally use the α-cuts [9]. However, no general approach allowing common arithmetic operations to be used on fuzzy numbers is available. This work proposes the use of a so-called φ–calculus fuzzy arithmetic based on a "different" modeling of fuzzy numbers [10, 11]. For this algebra, the modeling of fuzzy numbers is considered through the distribution function instead of the classical membership function. The first part presents some generalities on this φ–calculus arithmetic.

One of the main interests of this algebra is to provide some nice properties for what is called "exact calculus" with fuzzy numbers. These properties can be used in order to invert a fuzzy model. Therefore, the second part presents a possible application in control to fuzzy internal model control scheme.

2 Fuzzy arithmetic: φ-calculus

In the literature, many modeling approaches of imprecision, which is involved in many applications and domains, use fuzzy numbers and fuzzy arithmetic. In many works, the methods have led to the development of various membership functions for representing fuzzy numbers. Fuzzy numbers are often represented in applications by LR fuzzy sets and in particular, triangular and trapezoidal fuzzy sets. There exist two approaches for fuzzy arithmetic, on one hand the Zadeh's extension principle, on the other hand, the alpha-cuts and intervals arithmetic. Our works concern the first case, using a new modeling for fuzzy numbers [11], and it allows including a large part of the results existing in the domain of arithmetic [4, 6, 8, 12].

2.1 Fuzzy numbers modeling

Usually, a fuzzy number is modeled by its membership function $\mu_{\tilde{a}}$ not null on a bounded set $Supp(\tilde{a}) \subset \mathbb{R}$. For example, a triangular fuzzy number (TFN) $\tilde{a}$ can be represented by the shorthand symbol (b, m, c) with $\mu_{\tilde{a}}(x)=1$ for $x=m$ (mode) and the kernel defined by the interval $[m-b, m+c]$.

Herein, instead of the classical membership function, the modeling used for the representation of fuzzy numbers is based on the distribution function $\varphi_{\tilde{a}}$ defined by the following expression:

$$\varphi_{\tilde{a}}(x)=\frac{\int_{-\infty}^{x}\mu_a(t)\cdot dt}{\int_{-\infty}^{+\infty}\mu_a(t)\cdot dt} \qquad (1)$$

Convergence of $\int_{-\infty}^{+\infty}\mu_a(t)\cdot dt$ (finite cardinality of $\tilde{a}$) is assumed by considering only membership functions on a compact support $I \subset \mathbb{R}$. Thus, the major interest is that the distribution function for all fuzzy number $\tilde{a}$ is always an increasing monotone function, from $Supp(\tilde{a})$ to $[0,1]$. Thus, an inverse function $\varphi_{\tilde{a}}^{-1}$ from $[0,1]$ to $Supp(\tilde{a})=[\underline{a},\overline{a}]$ can always be defined and its definition is very important for the operations. The set of fuzzy numbers represented by a distribution function is noted Φ.

It should be noticed that the distribution function for a singleton is also a singleton and the one for an interval is a line between its bounds.

ISBN: 978-989-95079-6-8

2.2 Fuzzy realization and extension

Similarly to defuzzification and fuzzification concepts, it is necessary to define a relation between a crisp value and a fuzzy number. Therefore, an application from Φ to $\mathbb{R}$, which associates a crisp number to a fuzzy number, is called *fuzzy realization*. Conversely, an application from $\mathbb{R}$ to Φ is called *fuzzy extension*.

The choice of a realization depends obviously on the application. The most frequently encountered are:

- The median realisation noted $R_{med}(\tilde{a})$, it associates to a distribution function $\varphi_{\tilde{a}}$ the number a_0 such as $\varphi_{\tilde{a}}(a_0)=0.5$. Through this median realization, the following equivalence relation $\Re$ in Φ can be defined: the set of the functions $\varphi_{\tilde{a}}$ such as $\varphi_{\tilde{a}}(a_0)=0.5$ defines the class of equivalence of $a_0 \in \mathbb{R}$. Every element of this class will be called "fuzzy a_0", and denoted $\tilde{a}_0$.
- The modal realisation noted $R_{\mathrm{mod}}(\tilde{a})$ is defined as: $a_0 = R_{\mathrm{mod}}(\tilde{a}) \Leftrightarrow \varphi_{\tilde{a}}(a_0) = \underset{x\in\tilde{a}}{Max}\left[\mu_{\tilde{a}}(x)\right]=1$.
- The mean realization noted $R_{mean}(\tilde{a})$ is defined as: $a_0 = R_{mean}(\tilde{a}) = \int_0^1 \varphi_{\tilde{a}}^{-1}(y).dy$.

2.3 arithmetic operations

Consider two fuzzy numbers $\tilde{a}$ and $\tilde{b}$ using their distribution $\varphi_{\tilde{a}}$ and $\varphi_{\tilde{b}}$. For a fuzzy number $\tilde{a}$ whose support includes 0 let us also define its "negative" $\varphi_{\underline{\tilde{a}}}$ and "positive" $\varphi_{\overline{\tilde{a}}}$ parts as $\varphi_{\tilde{a}} = \varphi_{\underline{\tilde{a}}} + \varphi_{\overline{\tilde{a}}}$ with:

$$\varphi_{\underline{\tilde{a}}}(x)=\begin{cases}0 & \text{if } x<\underline{a}\\ \varphi_{\tilde{a}}(x) & \text{if } x\in[\underline{a},0]\\ 1 & \text{if } x>0\end{cases}, \qquad (2)$$

$$\varphi_{\overline{\tilde{a}}}(x)=\begin{cases}0 & \text{if } x<0\\ \varphi_{\tilde{a}}(x) & \text{if } x\in[0,\overline{a}]\\ 1 & \text{if } x>\overline{a}\end{cases} \qquad (3)$$

The classical arithmetic operations on fuzzy numbers (addition $\tilde{a}+\tilde{b}$, pseudo-opposite $-\tilde{a}$, subtraction $\tilde{a}-\tilde{b}$, multiplication $\tilde{a}\times\tilde{b}$, pseudo-inverse only for a non mixed-type fuzzy number $\tilde{b}^{-1}=1/\tilde{b}$ and division $\tilde{a}/\tilde{b}$) are defined hereinafter [9].

$$\begin{aligned}
&\varphi^{-1}_{\tilde{a}+\tilde{b}}(x)=\varphi^{-1}_{\tilde{a}}(x)+\varphi^{-1}_{\tilde{b}}(x)\\
&\varphi_{-\tilde{a}}(x)=1-\varphi_{\tilde{a}}(-x)\ \forall x\in\mathbb{R}\\
&\varphi^{-1}_{\tilde{a}-\tilde{b}}(x)=\varphi^{-1}_{\tilde{a}}(x)+\varphi^{-1}_{-\tilde{b}}(x)\\
&\varphi^{-1}_{\tilde{a}\times\tilde{b}}(x)=\varphi^{-1}_{\underline{\tilde{a}}\times\underline{\tilde{b}}}(x)+\varphi^{-1}_{\overline{\tilde{a}}\times\overline{\tilde{b}}}(x)\\
&\varphi_{\tilde{a}^{-1}}(x)=1-\varphi_{\tilde{a}}(1/x)\ \forall x\in\mathbb{R}^*\\
&\varphi^{-1}_{\tilde{a}/\tilde{b}}(x)=\varphi^{-1}_{\tilde{a}\times\tilde{b}^{-1}}(x)
\end{aligned} \qquad (4)$$

with $\varphi^{-1}_{\underline{\tilde{a}}\times\underline{\tilde{b}}}=\min\left[\varphi^{-1}_{\underline{\tilde{a}}}\times\varphi^{-1}_{\underline{\tilde{b}}},\varphi^{-1}_{\overline{\tilde{a}}}\times\varphi^{-1}_{\overline{\tilde{b}}}\right]$ and $\varphi^{-1}_{\overline{\tilde{a}}\times\overline{\tilde{b}}}=\max\left[\varphi^{-1}_{\underline{\tilde{a}}}\times\varphi^{-1}_{\underline{\tilde{b}}},\varphi^{-1}_{\overline{\tilde{a}}}\times\varphi^{-1}_{\overline{\tilde{b}}}\right]$.

Let us notice that these four classical operations are compatible with the Moore's interval operations.

2.4 Weighted fuzzy fusion operator

This part deals with a fusion operator introduced within the framework of this Φ-calculus algebra. This operator is called the weighted fuzzy fusion (WFF) [13] and is useful for the next sections.

Consider two fuzzy numbers $\tilde{a},\tilde{b}\in\Phi$ and their distribution $\varphi_{\tilde{a}},\varphi_{\tilde{b}}: x\in\mathbb{R}\rightarrow y\in[0,1]$. The weighted fuzzy fusion (WFF) of the two fuzzy numbers is defined as:

$$\varphi_{WFF}(x)=\frac{p_{\tilde{a}}\cdot\varphi_{\tilde{a}}(x)+p_{\tilde{b}}\cdot\varphi_{\tilde{b}}(x)}{p_{\tilde{a}}+p_{\tilde{b}}},\quad \forall x\in\mathbb{R} \qquad (5)$$

with $p_{\tilde{a}},p_{\tilde{b}}\in\mathbb{R}$ and $p_{\tilde{a}}+p_{\tilde{b}}=1$.

3 Internal model control

3.1 Principle of the structure

As with any open-loop control scheme, the internal model control (IMC) structure applies only on stable systems with a minimum phase behavior, according to Fig. 1.

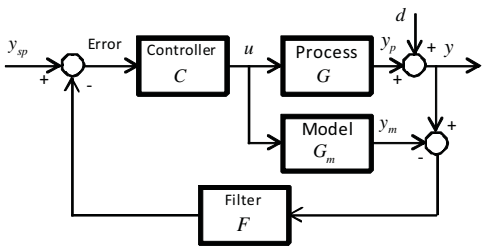


Figure 1: Internal model control scheme.

The filter F is introduced in order to filter out the measurement noise and to introduce some robustness in the loop. Its static gain is 1. Classically for linear models, it can be shown that the output equation y_k (Fig. 1) is given by:

$$y_k = F_{FB}(q^{-1})\cdot y_{sp_k} + S_{yd}(q^{-1})\cdot d_k \qquad (6)$$

with:

$$F_{FB}(q^{-1})=\frac{G(q^{-1})C(q^{-1})}{1+C(q^{-1})F(q^{-1})\left[G(q^{-1})-G_m(q^{-1})\right]}$$

$$S_{yd}(q^{-1})=\frac{1-G_m(q^{-1})C(q^{-1})F(q^{-1})}{1+C(q^{-1})F(q^{-1})\left[G(q^{-1})-G_m(q^{-1})\right]}$$

and the control law equation is given by:

ISBN: 978-989-95079-6-8

$$u_k = \frac{C\left(q^{-1}\right)}{1-C\left(q^{-1}\right)F\left(q^{-1}\right)G_m\left(q^{-1}\right)}\left(y_{c_k} - F\left(q^{-1}\right)y_k\right) \quad (7)$$

The minimum purpose of the internal model structure is to guarantee closed-loop stability and disturbance rejection, i.e. $F_{FB}(1)=1$ and $S_{yd}(1)=0$.

Recall also that the internal model control performs a kind of model simplification. Therefore its extension to fuzzy internal control needs the definition of an inverse model. The next part treats this issue.

3.2 Model inversion

Let us consider a stable time-invariant controllable SISO linear process model whose mathematical description is given in the form of recurrent equation:

$$y_k = -\sum_{i=1}^{N}\left(a_i \cdot y_{k-i} + b_i \cdot u_{k-i}\right) \quad (8)$$

where y_k and u_k correspond to the process output and input at sample k, $a_i, b_i, i \in \{1,\ldots,N\}$, denote the model parameters.

Suppose now that these parameters are fuzzy numbers, $\tilde{a}_i, \tilde{b}_i, i \in \{1,\ldots,N\}$ leading to:

$$y_k = -\sum_{i=1}^{N}\left(\tilde{a}_i \cdot y_{k-i} + \tilde{b}_i \cdot u_{k-i}\right) \quad (9)$$

For model inversion, the following question arises [14]:

knowing the description of uncertainties for the model, is it possible to synthesize a controller, based on the inverse model, able to maintain the model output within a tolerance envelope around the exact trajectory y_{sp} ?

Which means:

$$\text{Model output} \in \left[y_{sp} - \Delta,\ y_{sp} + \Delta\right] \quad (10)$$

where Δ is the accepted tolerance around the nominal trajectory.

To answer to this question, this work will use an approach developed in [14]. Notice that the "fuzzy" part of the uncertainties is not really taken into account, i.e. this problem could be solved using interval techniques, for example [15, 16, 17]. Nevertheless, the arithmetic provided therein gives an "easy" way to solve non exact calculus by means of probability distribution.

To compute the control law u_k at sample k, the terms defined at previous samples $\{y_{k-1}, i=1,\ldots,N\}$ and $\{u_{k-1}, i=1,\ldots,N\}$ are known. As $b_1 \neq 0$, equation (6) can be rewritten as:

$$u_k = \frac{1}{b_1}\left(y_{k+1} + \sum_{i=1}^{N}\left(a_i \cdot y_{k-i+1}\right) - \sum_{i=2}^{N}\left(b_i \cdot u_{k-i+1}\right)\right) \quad (11)$$

Or, considering the unknown term at sample k, y_{k+1}, expression (9) can be rewritten as follow:

$$u_k = \frac{1}{b_1}\left(y_{k+1} + \psi\left(z(k)\right)\right) \quad (12)$$

with

$$\psi\left(z(k)\right) = \sum_{i=1}^{N} a_i \cdot y_{k-i+1} - \sum_{i=2}^{N} b_i \cdot u_{k-i+1} \text{ and}$$

$$z(k) = \left[y_k \ \ \ldots \ \ y_{k-N+1} \ \ u_{k-1} \ \ \ldots \ \ u_{k-N+1}\right]^T \in \mathbb{R}^{2N-1}.$$

For a linear model without time delay, if y_{k+1} is replaced with the desired reference trajectory the ideal result is: $y_{k+1} = y_{sp_k}$. Therefore, with a perfect compensation, the model output follows the desired trajectory with a pure time delay corresponding to one sample.

Going back to the question (cf. (10)), we want to guarantee that $y_{k+1} \in \left[\underline{y}_{k+1}, \overline{y}_{k+1}\right]$. With equations (11) and (12), for a fuzzy model (9), we have straightforwardly:

$$\tilde{u}_k = \frac{1}{\tilde{b}_1}\left(\tilde{y}_{k+1} + \tilde{\psi}\left(\tilde{z}(k)\right)\right) \quad (13)$$

$$\tilde{\psi}\left(\tilde{z}(k)\right) = \sum_{i=1}^{N} \tilde{a}_i \cdot \tilde{y}_{k-i+1} - \sum_{i=2}^{N} \tilde{b}_i \cdot \tilde{u}_{k-i+1} \text{ and}$$

$$\tilde{z}(k) = \left[\tilde{y}_k \ \ \ldots \ \ \tilde{y}_{k-N+1} \ \ \tilde{u}_{k-1} \ \ \ldots \ \ \tilde{u}_{k-N+1}\right]^T \in \mathbb{R}^{2N-1}.$$

Equation (13) gives a model inversion description using fuzzy numbers. The term $\tilde{y}_{k+1}$ is replaced with a desired trajectory defined by a fuzzy number $\tilde{y}_{sp_k}$ to obtain a causal control law. For example, if $\tilde{y}_{sp_k}$ is a triangular fuzzy number (TFN) written as:

$$\tilde{y}_{sp_k} = triple\left(y_{sp} - \Delta,\ y_{sp},\ y_{sp} + \Delta\right) \quad (14)$$

the control law is:

$$\tilde{u}_k = \frac{1}{\tilde{b}_1}\left(\tilde{y}_{sp_k} + \tilde{\psi}\left(\tilde{z}(k)\right)\right) \quad (15)$$

$$\tilde{\psi}\left(\tilde{z}(k)\right) = \sum_{i=1}^{N} \tilde{a}_i \cdot \tilde{y}_{sp_k-i+1} - \sum_{i=2}^{N} \tilde{b}_i \cdot \tilde{u}_{k-i+1} \text{ and}$$

$$\tilde{z}(k) = \left[\tilde{y}_{sp_k} \ \ \ldots \ \ \tilde{y}_{sp_k-N+1} \ \ \tilde{u}_{k-1} \ \ \ldots \ \ \tilde{u}_{k-N+1}\right]^T \in \mathbb{R}^{2N-1}.$$

Of course, the vector $\tilde{z}(k)$ depends on the control part meaning that all $\tilde{y}_i$ are replaced with desired set point $\tilde{y}_{sp_i}$ which corresponds to the schematic diagram of Fig. 2.

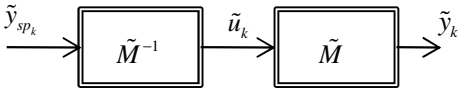


Figure 2: Fuzzy model inversion.

Naturally, perfect cancellation, i.e. $\tilde{y}_{k+1} = \tilde{y}_{sp_k}$ is not possible. Moreover using arithmetic of fuzzy numbers, or intervals or φ–calculus will introduce over-estimations. To illustrate this important point, an example issued from [14] is discussed hereinafter.

Let us consider the second-order:

$$\tilde{y}_k = -\tilde{a}_1 \cdot \tilde{y}_{k-1} - \tilde{a}_2 \cdot \tilde{y}_{k-2} + \tilde{b}_1 \cdot \tilde{u}_{k-1} + \tilde{b}_2 \cdot \tilde{u}_{k-2} \quad (16)$$

where $\tilde{a}_1$, $\tilde{a}_2$, $\tilde{b}_1$ and $\tilde{b}_2$ are fuzzy numbers defined as:

ISBN: 978-989-95079-6-8

$$\begin{aligned}\tilde{a}_1 &= triple(-0.6, -0.55, -0.5)\\ \tilde{a}_2 &= triple(0.05, 0.1, 0.15)\\ \tilde{b}_1 &= triple(0.6, 0.625, 0.65)\\ \tilde{b}_2 &= triple(0.15, 0.2, 0.25)\end{aligned} \quad (17)$$

The desired trajectory is:

$$y_{sp_k} = 0.5 \cdot \left(\sin\left(\frac{2\pi k}{50}\right) + \sin\left(\frac{2\pi k}{75}\right)\right) \quad (18)$$

and $\tilde{y}_{sp_k}$ is chosen as

$$\tilde{y}_{sp_k} = triple\left(y_{sp_k} - 1,\ y_{sp_k},\ y_{sp_k} + 1\right) \quad (19)$$

Using (15), the evolutions of the set point $\tilde{y}_{sp_k}$ envelopes (minimum $\underline{y}_{sp_k}$ and maximum $\overline{y}_{sp_k}$ of $Supp\left(\tilde{y}_{sp_k}\right)$) and the output model $\tilde{y}_k$ envelopes are illustrated in Fig. 3. Obviously, like interval calculus or α-cut fuzzy arithmetic, it clearly appears that the φ–calculus arithmetic generates overestimated results.

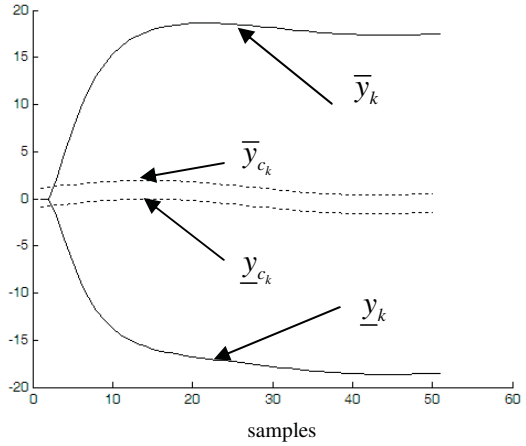


Figure 3: Set point $\tilde{y}_{sp_k}$ and output model $\tilde{y}_k$ envelopes.

We claim that this overestimation is not justified in such cases. Therefore, we need to introduce some extra constraints in order to take into account some knowledge and to reduce the pessimism of the results.

3.3 "Exact" inverse computation

Equation (15) corresponds to resolution of a fuzzy affine equation:

$$\tilde{b} \times \tilde{x} + \tilde{a} = \tilde{y} \quad (20)$$

The idea developed in [15] is to exactly solve this equation according to the two following steps.

Step1: Solve (20) with respect to $\tilde{d} = \tilde{b} \times \tilde{x}$.

The objective is to compute an "acceptable" solution $\tilde{d}$ of the equation $\tilde{d} + \tilde{a} = \tilde{y}$; i.e. the solution being exact when possible and approximated otherwise.

Notice that the exact solution can always be computed according to the arithmetic used:

$$\varphi_{\tilde{d}}^{-1}(u) = \varphi_{\tilde{y}}^{-1}(u) - \varphi_{\tilde{a}}^{-1}(u) \quad (21)$$

The problem arises when the result $\varphi_{\tilde{d}}^{-1}(u)$ is not an increasing function, resulting to $\tilde{d} \notin \Phi$. Nevertheless we can provide an approximate solution to the problem $\tilde{d}_{app} \in \Phi$. The fact of proposing an alternative to exact calculus is the key idea. Contrary to what generally authors accept, i.e. no solution for interval arithmetic or α-cut fuzzy arithmetic cases [14], we provide an approximate solution based on the initial $\varphi_{\tilde{d}}^{-1}(u)$.

For that purpose, we propose to use the weighted fuzzy fusion (WFF) operator on the set of pairs $\left(\varphi_{\tilde{d}}^{-1}(u_i), u_i\right)$, $i \in \{1, \ldots, N\}$. It results in a re-ordering of the pairs that constructs an increasing function $\tilde{d}_{app}$, thus $\tilde{d}_{app} \in \Phi$ is a feasible solution to the problem.

Step2: Finding a solution to equation $\tilde{d} = \tilde{b} \times \tilde{x}$ with $0 \notin Supp\left(\tilde{b}\right)$. If $\tilde{b} > 0$ and $\tilde{d} > 0$, the exact calculus uses a point by point division:

$$\varphi_{\tilde{x}}^{-1}(u) = \varphi_{\tilde{d}}^{-1}(u) \big/ \varphi_{\tilde{b}}^{-1}(u) \quad (22)$$

In the general case, we use a "positive-negative" decomposition (see section 2.3) for $\tilde{d}$ and $\tilde{x}$. We define:

$$\begin{cases}\varphi_{\tilde{d}}^{-1}(u) = \varphi_{\tilde{d}_-}^{-1}(u) + \varphi_{\tilde{d}_+}^{-1}(u)\\ \varphi_{\tilde{x}}^{-1}(u) = \varphi_{\tilde{x}_-}^{-1}(u) + \varphi_{\tilde{x}_+}^{-1}(u)\end{cases} \quad (23)$$

with $\varphi_{\tilde{d}_-}^{-1}(u)$, $\varphi_{\tilde{d}_+}^{-1}(u)$, $\varphi_{\tilde{x}_-}^{-1}(u)$ and $\varphi_{\tilde{x}_+}^{-1}(u)$ defined equations (3) and (4). At last, the exact solution is given by:

$$\begin{cases} if\ \tilde{b} > 0, \begin{cases}\tilde{x}_+ = \tilde{d}_+ \big/ \tilde{b},\ \tilde{x}_- = -\left(\left(-\tilde{d}_-\right) \big/ \tilde{b}\right),\\ \tilde{x} = \tilde{x}_+ + \tilde{x}_-\end{cases}\\ if\ \tilde{b} < 0, \begin{cases}\tilde{x}_+ = \tilde{d}_+ \big/ \left(-\tilde{b}\right),\ \tilde{x}_- = -\left(\left(-\tilde{d}_-\right) \big/ \left(-\tilde{b}\right)\right),\\ \tilde{x} = -\left(\tilde{x}_+ + \tilde{x}_-\right)\end{cases}\end{cases} \quad (24)$$

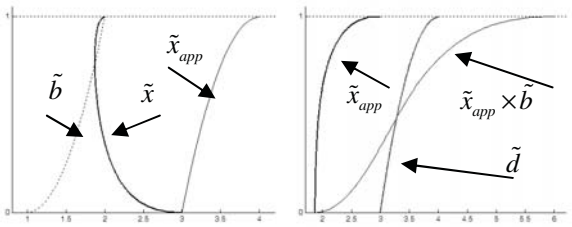
Once again the problem arises when the result $\varphi_{\tilde{x}}^{-1}(u)$ is not an increasing function, resulting to $\tilde{x} \notin \Phi$. Therefore approximate solution to the problem $\tilde{x}_{app} \in \Phi$ is generated in the same way using the WFF operator on the set of pairs $\left(\varphi_{\tilde{x}}^{-1}(u_i), u_i\right)$, $i \in \{1, \ldots, N\}$.

Consider an example to illustrate this second step with the fuzzy numbers $\tilde{b} = triple(1, 2, 2)$ and $\tilde{d} = triple(3, 3, 4)$. As shown Fig. 4(a), the exact solution $\tilde{x} \notin \Phi$, i.e. $\varphi_{\tilde{x}}^{-1}$ is not a monotone increasing function. Using the weighted fuzzy fusion (WFF) operator, an approximate solution $\tilde{x}_{app}$ is generated by re-ordering $\varphi_{\tilde{x}}^{-1}$. Fig. 4(b) shows the comparison between $\tilde{d}$ and $\tilde{b} \times \tilde{x}_{app}$. In a sense, the difference between these two fuzzy numbers exhibits the pessimism induced by the method when no exact solution in Φ is available. This pessimism can be shown comparing the supports of the numbers:

$$Supp\left(\tilde{d}\right) = [3, 4] \subset Supp\left(\tilde{b} \times \tilde{x}_{app}\right) = [1.87, 6].$$

ISBN: 978-989-95079-6-8

The next example demonstrates very clearly the interest of the complete algorithm used to find solutions to the problem $\tilde{b}\times\tilde{x}+\tilde{a}=\tilde{y}$. Consider again the inverse model of section 3.2 with the fuzzy numbers (17) and the desired trajectory (18), (19). Fig. 5 shows the same trial as for Fig. 3. For this test, we obtain the result considering, i.e. $\tilde{y}_{k+1}=\tilde{y}_{sp_k}$. This result indicates that an "exact solution" has been found at each time k.



a) « Exact » result b) φ–calculus result

Figure 4: Results with the numbers $\tilde{b}$ and $\tilde{d}$.

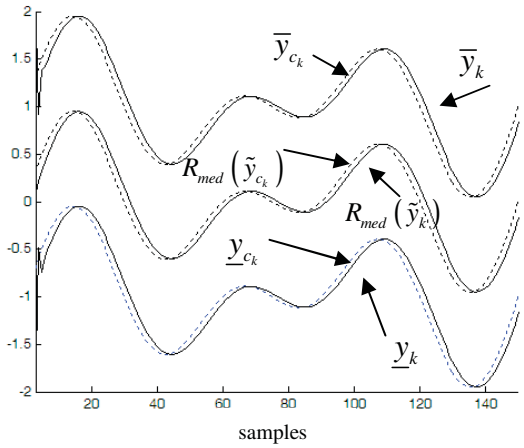


Figure 5: Set point $\tilde{y}_{sp_k}$ and output model $\tilde{y}_k$ envelopes.

In order to show the interest of the method, we will change the constraints level. To do so, the values of model uncertainties (17) are increased in order to deal with cases where an exact inversion is not always possible. New bounds for the fuzzy numbers (17) are defined. They correspond to the greatest possible values with pole-zero assignment located inside the unit circle – for evident stability purpose. Moreover, the fuzzy numbers are not more symmetric and correspond to:

$$\begin{aligned}\tilde{a}_1 &= triple(-0.825,-0.4,0.3)\\ \tilde{a}_2 &= triple(-0.175,0.25,0.95)\\ \tilde{b}_1 &= triple(0.415,0.625,0.835)\\ \tilde{b}_2 &= triple(-0.01,0.2,0.41)\end{aligned} \qquad (25)$$

Fig. 6 repeats the same trial with the desired trajectory (18), (19). The figure exhibits that an exact simplification, i.e. $\tilde{y}_{k+1}\neq\tilde{y}_{sp_k}$ is not always possible. Contrary to have no solution in such cases, the proposed approximate solutions seem perfectly adapted.

With the help of this inverse model control strategy based on constraints, justified in the control framework, it is now possible to consider a fuzzy internal model control.

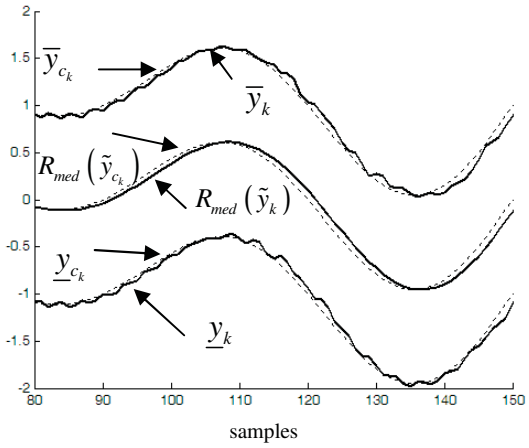


Figure 6: Set point $\tilde{y}_{sp_k}$ and output model $\tilde{y}_k$ envelopes.

3.4 Fuzzy internal model control

From Fig. 6 showing IMC scheme and different equations (6) and (7), it is possible to define several solutions.

First solution: Let $G_m(q^{-1})=q^{-r}G_m^+(q^{-1})$, $G_m^+(q^{-1})$ corresponds to function with an inverse (exactly proper and stable). We chose $C(q^{-1})=\dfrac{1}{G_m^+(q^{-1})}$ and propose the equivalent model given Fig. 7.

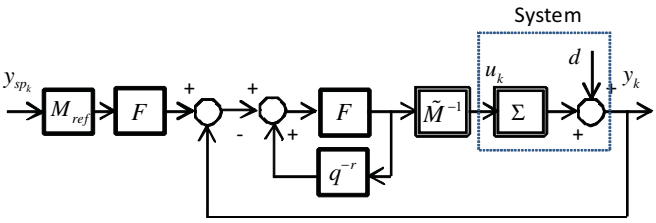


Figure 7: IMC1, equivalent scheme.

Indeed, equation (7) can be rewritten as follow:

$$u_k=\frac{1}{G_m^+(q^{-1})\left(1-q^{-r}\cdot F(q^{-1})\right)}\left(y_{c_k}-F(q^{-1})\,y_k\right) \qquad (26)$$

By analogy, we have the fuzzy inverse model $\left[\tilde{G}_m^+(q^{-1})\right]^{-1}$ multiplied by linear transfer function $\dfrac{F(q^{-1})}{1-q^{-r}\cdot F(q^{-1})}$.

For every case, it is possible to add a model reference control M_{ref} in order to attenuate the control signal.

ISBN: 978-989-95079-6-8

3.5 Example

Consider a nominal second-order transfer function:

$$\frac{y\left(q^{-1}\right)}{u\left(q^{-1}\right)}=\frac{B\left(q^{-1}\right)}{A\left(q^{-1}\right)}=\frac{0.0233\cdot q^{-1}+0.0197\cdot q^{-2}}{1-1.5729\cdot q^{-1}+0.6037\cdot q^{-2}} \quad (27)$$

and the following definition for the fuzzy numbers:

$$\begin{aligned}\tilde{a}_1 &= triple\left(-1.574,-1.573,-1.497\right)\\ \tilde{a}_2 &= triple\left(0.602,0.603,0.652\right)\\ \tilde{b}_1 &= triple\left(0.013,0.023,0.033\right)\\ \tilde{b}_2 &= triple\left(0.01,0.02,0.03\right)\end{aligned} \quad (28)$$

The control structure used is the IMC1 Fig. 7 with a reference model defined as:

$$\frac{B_m\left(q^{-1}\right)}{A_m\left(q^{-1}\right)}=\frac{0{,}1784\cdot q^{-1}+0{,}1071\cdot q^{-2}}{1-0{,}9315\cdot q^{-1}+0{,}2169\cdot q^{-2}} \quad (29)$$

Fig. 8 shows the results for different set points. The first set-point leads to results around the nominal model (27) (represented by the value 2). Secondly, two successive variations are made (set-point value of 3 and set-point value of 4), lastly a return to nominal position.

During both successive changes the model is changed in order to reach the bounds on $\tilde{a}_1$ and $\tilde{b}_2$. Two disturbances have been also added at samples 900 and 2300. The responses show a good robustness according to the parametric variations of the model.

This work presents a first possible track to use fuzzy arithmetic for control purpose. Next steps would be to prove stability and evaluate the robustness of these approaches. We can think to several possibilities including Lyapunov approach that are insightfully used for Takagi-Sugeno models [18] or Kharitonov polynomials [19].

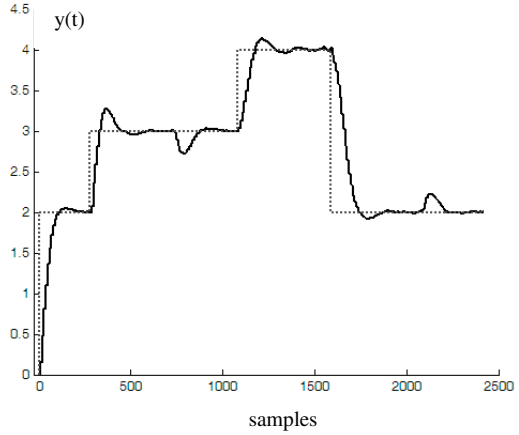


Figure 8: IMC1 results.

4 Conclusions

This paper attempts to show possible applications of the φ–calculus arithmetic for automatic control. As a first step, a new characterization of fuzzy numbers by the distribution function instead of the classical membership function has been presented. For this algebra, it is possible to use a set of arithmetic operators (addition, opposite and subtraction, multiplication, inverse and quotient) compatible with the classical algebra by using the median realization, or the Moore's calculus by using the fuzzy number support. A procedure allowing an "exact" calculus of an inverse model was proposed on the basis of φ–calculus. This approach was successfully implemented on internal model control structures.

Thus, the application of φ–calculus arithmetic to the stability of system seems to open some new prospects for other basic concepts (observation, stabilization…).

Acknowledgment

This work was supported in part by International Campus on Safety and Intermodality in Transportation, the European Community (through the FEDER European Funds for Regional Development), the Délégation Régionale à la Recherche et à la Technologie, the Ministère de l'Enseignement supérieur et de la Recherche, the région Nord Pas-de-Calais and the Centre National de la Recherche Scientifique (CNRS): the authors gratefully acknowledge the support of these institutions.

References

[1] L A. Zadeh, *Fuzzy sets*, Information & Control, 8:338-353, 1965.

[2] A. Sala, T. M. Guerra, R. Babuška. *Perspectives of fuzzy systems and control*, Fuzzy Sets and Systems, 156:432-444, 2005.

[3] L. Jaulin, M. Kieffer, O. Didrit, E. Walter. *Applied Interval Analysis*, Springer Edition, 2001.

[4] R.E. Moore. *Interval analysis*, Prentice-Hall-Englewood Cliffs, New Jersey, 1966.

[5] U.W. Kulish, W.L. Miranker. *Computer arithmetic in theory and practice*. Academic Press, 1981.

[6] D. Dubois, H. Fargier, J. Fortin. *A generalized vertex method for computing with fuzzy intervals*, Proc. of the International Conference on Fuzzy Systems (IEEE, ed.), 541-546, 2004.

[7] D. Dubois & H. Prade. *Fuzzy real algebra: some results*. Fuzzy Sets and Systems, 2:327-383, 1979.

[8] E. Sanchez. Solutions of fuzzy equations with extended operators. Fuzzy Sets and Systems, 12:237-248, 1984.

[9] Z. Zhao, R. Govind. Solutions of algebric equations involving generalised fuzzy numbers. Information sciences, 56:199-243, 1991.

[10] D. Roger, J.-M. Lecomte. *φ-calcul : une arithmétique pratique pour les nombres flous*. LFA'98, Rennes, 1998.

[11] H. Lamara, L. Vermeiren & D. Roger. « φ-calculus*: A new fuzzy arihtmetic* », 11th IEEE ETFA'06, Prague, Czech Republic, 2006.

[12] A. Kaufmann, M.M. Gupta. Introduction to fuzzy arithmetic theory and application. Van Nostrand Reinhold, New York, 1991.

[13] H. Lamara, L. Vermeiren, D. Roger. *Aggregation and fuzzification by weighted fuzzy fusion operator*. IEEE-FUZZ'07, London, United Kingdom, 2007.

[14] R. Boukezzoula, S. Galichet, L. Foulloy. *Nonlinear Internal Model Control: Application of inverse Model Based Fuzzy Control*, IEEE T. Fuzzy Systems, 11(6):814-829, 2003.

[15] R. Boukezzoula, L. Foulloy, S. Galichet. *Inverse Controller Design for Fuzzy Interval Systems*, IEEE T. Fuzzy Systems, 14(1):111-124, 2006.

[16] E. Gardeñes, M.A. Sainz, L. Jorba, R. Calm, R. Estela, H. Mielgo, A.A. Trepat. *Modal intervals*, Reliable Computing, 7:77-111, 2001, Kluwer Academic Publishers.

[17] J.E.F. Diaz. *Improvements in the Ray Tracing of Implicit Surfaces based on Interval Arithmetic*. Doctoral Thesis, Girona, Spain, 2008.

[18] T.M. Guerra, L. Vermeiren. LMI-based relaxed non quadratic stabilization conditions for non-linear systems in the Takagi-Sugeno's form, Automatica, 40(5) :823-829.

[19] C.W. Tao and J.S. Taur. Robust Fuzzy Control for a plant with fuzzy linear model, IEEE T. Fuzzy Systems, 13(1):20-41, 2005.

ISBN: 978-989-95079-6-8

Optimising the Fuzzy Granulation of Attribute Domains

Marcos E. Cintra[1] Heloisa A. Camargo[2] Trevor Martin[3]
1.Mathematics and Computer Science Institute, São Paulo University(USP)
P.O. Box 668, 13561-970, São Carlos-SP, Brazil
2.Computer Science Department, Federal University of São Carlos(UFSCar)
P.O. Box 676, 13565-905 São Carlos-SP
3.Mathematics Engeneering Department, University of Bristol
Queen's Building, University Walk, Bristol - UK - BS8 1TR
Email: cintra@icmc.usp.br, heloisa@dc.ufscar.br, Trevor.Martin@bristol.ac.uk

Abstract *— The definitions of the number of fuzzy sets and their proper distribution on their domains are fundamental issues for fuzzy systems since these basic parameters deeply affect the quality of the systems results, both in terms of performance rates and interpretability. Several methods have been proposed in the literature to define these parameters, although it is common to find works in which the number of fuzzy sets is defined empirically, distributing them equally in the domains. This paper presents a fast and easy method to estimate the number of fuzzy sets for each attribute and compares three methods for the distribution of fuzzy sets. Two of them are non-supervised methods, using same width and same frequency, the third one is an adaptation of the 1-R supervised method to discretize attributes. Experiments with 10 datasets for classification problems, 10-fold cross validation, using the Wang & Mendel method, and the classic and general fuzzy reasoning methods are presented and discussed.*

Keywords— Fuzzy systems, granulation of attributes, Wang & Mendel method, fuzzy data base generation.

1 Introduction

Fuzzy set theory and fuzzy logic [1] proposed by professor Loft A. Zadeh are the base for the fuzzy systems. Fuzzy set theory is used to represent and process information. The main characteristic of fuzzy sets, contrasting with crisp sets, is the progressive transition from one set to another. This natural characteristic of the fuzzy sets provides automatic mechanisms to deal with imprecision and uncertainty, which are inherent to real world knowledge. Moreover, the fuzzy logic theory prevents the creation of unnatural frontiers in the partitioning of attributes domains [2].

A fundamental issue, due to its direct impact on fuzzy systems, is the definition of the fuzzy sets that model the linguistic variables of a given domain, both in terms of shape (triangular, trapezoidal, S-function, etc) and partitioning of the attributes domains (number of fuzzy sets and their distribution). Unfortunately, there are no general rules or guidelines for these tasks that suit every domain [3]. In fact, many methods may have to be tested in order to find the appropriate definitions for a given application.

It is also common to find studies in which the definition of the membership functions is done empirically. In fact, the Gaussian, trapezoidal and triangular shapes are the most used in the literature, probably because they produce comparable results to other shapes and are easily interpretable [4]. Regarding the definition of the number of fuzzy sets and their distribution on the partitions, most of the papers in the literature use from 2 to 10 fuzzy sets for each attribute, equally distributed in their domains, *i.e.*, all fuzzy sets have same width.

Another aspect to be considered when defining fuzzy sets is personal interpretations, which make this task non trivial. For instance, let us consider a linguistic variable *temperature*; the interpretation of the linguist term *low temperature* is closely related to the region where one lives (consider a continental country such as Brazil, for example). Thus, personal interpretations may generate strong variations for the same concept: since each person may have different and personal interpretations on the meaning of a concept, it is natural that different membership functions may be created to define the same concept. This flexibility on the subjective selection of membership functions and, as previously stated, in the distribution of the fuzzy sets in their partitions shows the robustness of the fuzzy logic, which is closely related to the inherent characteristics of fuzzy sets [5].

Although this flexibility exists, the task of defining membership functions is largely ignored. The choice of the method, in fact, depends on the particular application and domain [6]. While the shape of the fuzzy sets may not present expressive differences in the results of fuzzy systems, the number of fuzzy sets and their distribution are relevant parameters that affect the system in terms of performance and in terms of interpretability. Thus, this paper presents a fast and easy method to estimate the number of fuzzy sets, in order to use this estimated number with more costly approaches, such as genetic algorithms. This paper also presents experiments carried out using three distinct methods for the definition of the distribution of fuzzy sets, in an attempt to provide further insight on the task, focusing specifically on classification problems. One of these methods uses same width distribution for the fuzzy sets, the second uses same frequency distribution, and the third takes the classes into consideration in the process.

This paper is organized as follows: Section 2 reviews methods for the automatic definition of fuzzy membership functions; Section 3 describes the three methods used in the experiments for the definition of fuzzy sets. Section 4 describes the heuristic method for the definition of the number of fuzzy sets proposed here. Section 5 presents the experiments and results, followed by the conclusions in Section 6.

ISBN: 978-989-95079-6-8

2 Methods for the Automatic Definition of Fuzzy Sets

There is a wide variety of methods applied for the definition of fuzzy sets, ranging from heuristic methods, genetic algorithms, artificial neural networks, clustering algorithms, the use of indexes, to the adaptation of classic machine learning methods, such as the K-NN algorithm. Next, we present some of these methods. More details on these and on other related methods can be found in [7] and on the provided references.

Several methods explore the use of special parameters for the distribution of fuzzy sets. In [2], for example, the authors use the concept of 2-tuple to define linguistic terms. The 2-tuple strategy is used to define both, the center of a fuzzy set (support), and a value for a symbolic translation, which enables the fuzzy sets to be moved sideways, always keeping the same base width. This symbolic translation of a linguistic term is a number within the interval $[-0.5, 0.5)$. A genetic algorithm is then used to define the best position for the fuzzy sets. The experiments used a fixed number of triangular shaped sets.

In [6] and [8] the authors use the Fuzzy C-Means algorithm for the definition of fuzzy membership functions. In the first method [6], a modified version of the Xie-Beni index [9] is used to define the number of clusters and, therefore, the number of fuzzy sets, for the Fuzzy C-Means algorithm. Once the number of fuzzy sets is defined, the centers of the sets are initialized based on two methods, the equalized universe method, which divides the partition equally by the number of sets, and the subtractive clustering method, which classifies the points present in the training data by their number of neighboring points (thus, a data point will have a high potential value if it has many neighboring points). After the initialization of the centers, the Fuzzy C-Means algorithm is employed to search for the best positions for centers. A hybrid learning algorithm for refining the system parameters based on the ANFIS method [10] is then presented. The second method [8] proposes the use of a Fuzzy C-Means variant for the generation of fuzzy term sets with ½ overlap. This method uses the mean squared error criterion to determine the number of fuzzy sets and the optimal shape of the membership functions associated with each term.

Genetic algorithms are used to tune the parameters of fuzzy sets in [11]. The genetic algorithm is used to find membership functions suitable for mining problems and then uses the final best set of membership functions to mine fuzzy association rules. The fitness of each chromosome is evaluated by the number of large 1-itemsets generated and by the suitability of the membership functions.

The use of artificial neural networks has also been explored for the definition of fuzzy membership functions. The authors in [12] and [13] present an overview of several neuro-fuzzy methods for the definition of fuzzy sets. These methods, although some present good results, have the problem of tunning parameters for the neural networks, and, in some cases, might need further tests.

There are also works focusing on the use of special indexes to generate membership functions. For instance, the concept of entropy is used in [14] and in [15] to determine membership functions. In [16] the authors propose a method for the definition of fuzzy sets based on two indexes, the fuzzy entropy and the fuzziness index, adopting the S shaped fuzzy set. These papers focus on image recognition applications. The Kappa measure was also used. In [17] the authors use the Kappa measure to calculate the *fuzzy Kappa* index in order to define a suitable segmentation of partitions.

Several other approaches can be found in the literature, such as heuristic methods [18], histograms [3], and even a fuzzy version of the classic machine learning K- nearest neighbor algorithm [19].

Although such a variety of methods exist, it is important to emphasize the fact that many studies define the number of fuzzy sets empirically, and employ the equalized universe method [6] to define the distribution of the fuzzy sets, which simply distributes the sets evenly in the partitions. Regarding the number of fuzzy sets, most researchers define this variable empirically using a range usually varying from 2 to 10 fuzzy sets per attribute. The main reasons why researchers decide to define such important variables without the aid of any formal method may include:

- The complexity of the methods available, which may require more time and effort to be implemented than the actual application in focus;
- The flexibility of the fuzzy logic for the definition of fuzzy variables and fuzzy terms, which allows the users to define their own fuzzy data base, since it will be the base for the generation of the fuzzy rules, which in turn will be adjusted to provide a suitable performance;
- The lack of consensus and/or guidelines on which of the available methods is the best for a given application;

The next section presents the methods used in the experiments conducted for this paper.

3 Evaluated Methods for Fuzzy Sets Distribution

The three methods tested and compared in this paper are the equalized universe method, same frequency method, and an adaptation of the 1-R method for attribute discretization. These methods are described next.

3.1 Equalized Universe Method

This method uses the same width for each fuzzy set [6]. Figure 1 presents an attribute described by five triangular fuzzy sets using the equalized universe method. Notice that, usually, the fuzzy sets at the extremes have half the width of the other fuzzy sets in the middle of the domain.

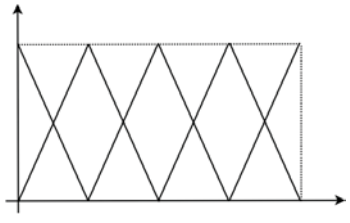


Figure 1: Attribute defined by 5 fuzzy sets using the equalized universe method.

This method is the most widely found in the literature. Some of the reasons for that may include the fact that it is

ISBN: 978-989-95079-6-8

a simple and easy method to use, and that it generates isosceles triangles, with ½ overlap, so no area of the domain will have a membership degree inferior to 0.5 [20]. Besides, they are easy to interpret.

3.2 Same Frequency Method

This method divides the domains of the attributes according to the frequency distribution of the examples. For our implementation, percentiles [21] were used to find the correct distribution for the sets. This way, if an attribute is described by 5 triangular fuzzy sets, the maximum membership degree for the first fuzzy set will be on the 16.7th percentile, on the 33.3th percentile for the second fuzzy set, on the 50.0th percentile for the third one, on the 66.7th percentile for the fourth one, and on the 83.3th percentile for the last fuzzy set. The triangles will not have the same base width, on the contrary, depending on the distribution of the examples, they are likely to be different. Notice that in the case of many examples with the same value, two or more fuzzy sets might be defined with exactly the same middle point. In order to avoid this problem, if the same example is chosen by two or more different percentile values, the neighboring examples of the left and right are chosen to replace the original value.

Figure 2 shows an attribute represented by 5 fuzzy sets distributed using the same frequency method, having 20 examples in the interval $[0, 10]$ (0, 0, 0, 1, 1, 2, 2, 2, 3, 5, 5, 6, 7, 7, 7, 9, 9, 9, 9, 9).

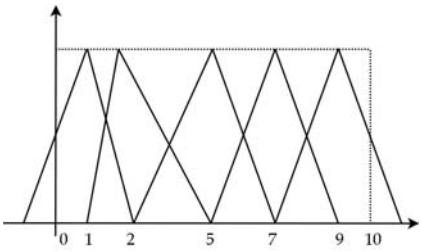


Figure 2: Attribute defined by 5 fuzzy sets using the same frequency method.

Notice that the fuzzy sets of the laterals are defined beyond the actual domain values. This strategy was used to avoid leaving any spaces of the domain with maximum membership degrees close to zero. The space between the actual beginning and end of the domain and the vertices of the triangles is calculated as the same space between the middle vertex of the triangle and the adjacent extremity of the domain.

3.3 1-R Supervised Method

The 1-R method is used for discretization of attributes. It basically divides the domain of an attribute in n intervals, in relation to the frequency of the examples, calculates the optimum class for each interval, and tries to find the best split point using the class. The original algorithm can be found in [22]. Algorithm 1 presents the adapted version of the 1-R algorithm used in the experiments.

Input: n = number of examples;
m = number of attributes;
i = number of fuzzy sets describing the attribute;
for a = *1* **to** m **do**
  Sort the examples by attribute a;
  Divide the examples in i intervals;
  for b = *1* **to** $i - 1$ **do**
    Calculate the optimal(majority) class for the interval;
    for c = *1* **to** *interval* $i - 2$ **do**
      for d = *index of first example of interval* $c + 1$ **to** *index of last example of interval* $c + 1$ **do**
        if *first example of interval* $c + 1$ *belongs to optimal class of interval* c **then** ;
        Adjust intervals so that the first example of interval $c + 1$ is transfered to interval c;
        else d = index of last example of interval $c + 1$;
        if d = *index of last example of interval* $c + 1$ **then** Split the examples of interval c between intervals c and $c + 1$;
      end
    end
  end
  For each interval, create a fuzzy set with core in the middle of the interval and laterals in the cores of the neighboring intervals;
end

Algorithm 1: 1-R adapted algoritghm.

The same strategy used for the same frequency method, to avoid leaving any spaces of the domain with maximum membership degrees close to zero, was used similarly for the 1-R adapted method.

Figure 3 shows an attribute represented by 5 fuzzy sets distributed by the same frequency method, having the following 20 examples (attribute value, class) in the interval $[0, 10]$: 0, 3; 0, 3; 0, 1; 1, 3; 1, 3; 2, 3; 2, 2; 2, 2; 3, 2; 5, 2; 5, 1; 6, 1; 7, 1; 7, 3; 7, 3; 9, 2; 9, 3; 9, 1; 9, 2; 9, 3.

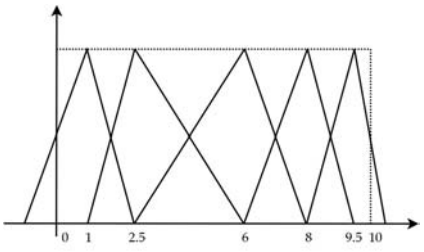


Figure 3: Attribute defined by 5 fuzzy sets using the 1-R adapted method.

4 Heuristic Method to Define the Number of Fuzzy Sets

The approaches used by the recently proposed methods to generate fuzzy rules usually demand a great computational effort, such as the genetic algorithms, that can take from minutes to days to converge to good solutions. Since the definition of the number of fuzzy sets is an open area of research, it is common to see researchers using a few different numbers of fuzzy sets for each attribute in order to validate their methods, usually ranging from 2 to 10, thus, repeating the complete round of experiments several times, according to the range of number of fuzzy sets used. If it were possible to establish an approximate optimal number of fuzzy sets, the total number of experiments would be greatly reduced. Thus, we propose a heuris-

ISBN: 978-989-95079-6-8

tic strategy to define this number using the Wang & Mendel method [23], which is able to generate fuzzy rule bases from data examples.

The Wang & Mendel method has been widely used due to its low complexity ($O(m \times n)$, m attributes, n examples) and the fact that it produces relatively small rule bases with good classification rates and no conflicting or redundant rules. However, nowadays it is possible to generate more precise fuzzy rule bases with even lower number of rules using other approaches, such as the genetic fuzzy systems or neural networks. The heuristic of the proposed method is based on the assumption that the Wang & Mendel method can be used as an indicator of the quality of the results that can be achieved by other methods, when using the same data base, *i.e.*, the same number, distribution, and shape of fuzzy sets.

Experiments shown in Section 5 were carried out using from 2 to 10 triangular shaped fuzzy sets defining each attribute of the domains. The idea is to run the experiments with the Wang & Mendel method before running the experiments with more time consuming approaches, and select the number of fuzzy sets that produced the best accuracy rates, or, in case of ties, the smaller number, due to interpretability reasons. Notice that a tolerance interval may be defined to allow the user to select the smaller number of fuzzy sets when a small difference between the accuracy rates is not significant, but the difference in the number of fuzzy sets is.

5 Experiments and Results

The experiments were conducted using 10-fold cross validation on 10 datasets. Table 1 shows a summary of the characteristics of the datasets used in the experiments, presenting the number of examples in each dataset (# Examples), number of attributes (# Attribs), number of classes (# Classes), and the majority class error (MCE).

Table 1: Characteristics of the datasets.

Dataset	# Examples	# Attribs	# Classes	MCE
Breast	684	10	3	34.99
Bupa	345	4	2	42.00
Credit	653	15	2	45.33
Diabetes	769	8	2	34.98
Glass	220	9	7	65.45
Heart	270	13	2	44.40
Iris	150	4	3	66.60
Segment	210	19	7	85.72
Vehicle	846	18	4	74.20
Wine	178	13	3	59.70

Two fuzzy reasoning methods, frequently employed in the fuzzy classification domain, were used, the classic and the general fuzzy reasoning methods. The classic fuzzy reasoning method uses the class of the rule with highest compatibility with an input pattern to classify it. The general fuzzy reasoning method, on the other hand, calculates the sum of compatibility degrees for each class and uses the class with highest sum to classify an input pattern. The accuracy was measured in terms of correct classification rates.

Table 2 presents the results for the experiments that were carried out using the classic fuzzy reasoning method. **Sets** describes the number of fuzzy sets used for each attribute in the experiments; **EU** represents the equalized universe method, which uses the same width for each fuzzy set, **F** represents the method that uses same frequency, and **1-R** represents the 1-R adapted method, which takes the classes into consideration when distributing the fuzzy sets.

Table 2: Classification rates - classic fuzzy reasoning method.

Sets	2			3			4		
Meth.	EU	F	1-R	EU	F	1-R	EU	F	1-R
Bre.	0.96	0.53	0.64	0.96	0.59	0.89	0.95	0.71	0.84
Bupa	0.49	0.46	0.43	0.52	0.51	0.55	0.55	0.53	0.57
Credit	0.54	0.58	0.60	0.54	0.58	0.73	0.53	0.55	0.63
Dia.	0.99	1.00	0.98	0.99	0.99	1.00	1.00	0.98	0.97
Glass	0.41	0.58	0.48	0.49	0.55	0.67	0.55	0.60	0.65
Heart	0.98	0.98	0.93	1.00	1.00	0.97	1.00	0.99	0.98
Iris	0.67	0.43	0.37	0.65	0.91	0.65	0.92	0.93	0.91
Seg.	0.47	0.14	0.26	0.74	0.15	0.54	0.71	0.15	0.50
Vehi.	0.99	0.95	1.00	1.00	0.98	1.00	1.00	0.97	1.00
Wine	0.94	0.99	0.99	0.99	0.98	0.97	1.00	1.00	0.98
Sets	**5**			**6**			**7**		
Meth.	EU	F	1-R	EU	F	1-R	EU	F	1-R
Bre.	0.92	0.69	0.76	0.89	0.66	0.71	0.91	0.65	0.70
Bupa	0.56	0.52	0.54	0.55	0.48	0.52	0.55	0.46	0.48
Credit	0.53	0.55	0.58	0.53	0.55	0.56	0.54	0.55	0.56
Dia.	1.00	0.94	0.98	1.00	1.00	0.97	0.99	1.00	0.97
Glass	0.61	0.56	0.61	0.61	0.53	0.57	0.63	0.52	0.55
Heart	1.00	0.98	1.00	1.00	0.97	1.00	0.93	0.95	0.99
Iris	0.91	0.90	0.91	0.87	0.83	0.87	0.93	0.80	0.87
Seg.	0.67	0.15	0.39	0.66	0.14	0.30	0.62	0.15	0.23
Vehi.	0.98	0.95	1.00	0.98	1.00	0.98	0.97	1.00	0.99
Wine	1.00	0.99	0.99	0.99	0.98	0.99	0.99	0.93	1.00
Sets	**8**			**9**			**10**		
Meth.	EU	F	1-R	EU	F	1-R	EU	F	1-R
Bre.	0.90	0.63	0.69	0.74	0.63	0.69	0.80	0.61	0.66
Bupa	0.61	0.45	0.42	0.58	0.43	0.42	0.55	0.43	0.45
Credit	0.54	0.55	0.55	0.60	0.55	0.55	0.57	0.55	0.55
Dia.	0.97	0.99	0.95	0.93	0.93	0.89	0.88	0.91	0.89
Glass	0.58	0.51	0.54	0.60	0.50	0.51	0.61	0.50	0.50
Heart	0.92	0.98	0.99	0.91	0.89	0.98	0.91	0.87	0.97
Iris	0.89	0.71	0.76	0.89	0.67	0.72	0.89	0.59	0.68
Seg.	0.54	0.14	0.21	0.53	0.15	0.18	0.45	0.14	0.16
Vehi.	0.95	0.95	0.99	0.94	0.97	0.97	0.93	0.95	0.98
Wine	0.93	0.88	0.95	0.81	0.85	0.93	0.73	0.81	0.91

Table 3 presents the results for the experiments that were carried out using the general fuzzy reasoning method. **Sets** describes the number of fuzzy sets used for each attribute; **EU** represents the equalized universe method, **F** represents the method that uses same frequency, and **1-R** represents the 1-R adapted method.

Table 4 shows the best estimated number of fuzzy sets, according to the proposed method, for the classic and general fuzzy reasoning methods, using the equalized universe method (**EU**), frequency method (**F**) and 1-R adapted method (**1-R**). The suffixes **C** and **G** were used to describe the classic and general fuzzy reasoning methods, respectively. A tolerance interval of 0.02 was used. Notice that the best classification rates used to define these values are light-gray shaded in Tables 2 and 3. The proposed method was able to estimate a small number of fuzzy sets for most of the datasets.

Figures 4 and 5 present the classification rates (vertical axis) for the classic and general fuzzy reasoning methods, respectively, for each of the 10 datasets, using from 2 to 10 fuzzy sets (horizontal axis) to define each of the attributes.

To test whether there is a significant difference among the methods, the Friedman test [24] was used with the null-hypothesis that the performance of the three methods, assessed in terms of the error rate, are comparable. As the null-hypothesis was rejected with a 95% confidence level, the Bonferroni-Dunn post-hoc test (to detect whether the differences among the methods are significant) was used [24]. Results showed that the equalized universe method is signifi-

ISBN: 978-989-95079-6-8

Table 3: Classification rates - general fuzzy reasoning method.

Sets	2			3			4		
Meth.	EU	F	1-R	EU	F	1-R	EU	F	1-R
Bre.	0.93	0.51	0.83	0.96	0.55	0.89	0.95	0.69	0.82
Bupa	0.57	0.55	0.44	0.53	0.59	0.57	0.55	0.61	0.57
Credit	0.49	0.48	0.53	0.49	0.48	0.65	0.48	0.47	0.51
Dia.	0.99	0.15	0.99	0.99	0.32	0.95	0.97	0.24	0.81
Glass	0.44	0.33	0.51	0.49	0.36	0.61	0.57	0.37	0.59
Heart	0.09	0.19	0.13	0.09	0.15	0.34	0.08	0.08	0.11
Iris	0.69	0.50	0.38	0.70	0.93	0.63	0.97	0.92	0.93
Seg.	0.41	0.14	0.20	0.70	0.15	0.56	0.71	0.16	0.51
Vehi.	0.98	0.44	1.00	0.98	0.44	0.95	0.97	0.28	0.65
Wine	0.93	0.52	0.99	0.89	0.22	0.81	0.73	0.05	0.36
Sets	**5**			**6**			**7**		
Meth.	EU	F	1-R	EU	F	1-R	EU	F	1-R
Bre.	0.92	0.67	0.73	0.89	0.66	0.71	0.91	0.65	0.69
Bupa	0.59	0.63	0.61	0.57	0.58	0.62	0.57	0.57	0.63
Credit	0.48	0.46	0.47	0.48	0.46	0.46	0.49	0.45	0.46
Dia.	0.93	0.11	0.51	0.88	0.06	0.25	0.80	0.02	0.11
Glass	0.54	0.27	0.50	0.56	0.25	0.39	0.55	0.17	0.29
Heart	0.06	0.02	0.04	0.05	0.01	0.04	0.04	0.01	0.02
Iris	0.94	0.90	0.91	0.89	0.80	0.84	0.94	0.77	0.83
Seg.	0.70	0.16	0.42	0.70	0.16	0.32	0.63	0.15	0.24
Vehi.	0.93	0.11	0.30	0.79	0.05	0.17	0.69	0.02	0.06
Wine	0.39	0.03	0.12	0.18	0.00	0.02	0.07	0.02	0.00
Sets	**8**			**9**			**10**		
Meth.	EU	F	1-R	EU	F	1-R	EU	F	1-R
Bre.	0.90	0.63	0.69	0.74	0.63	0.69	0.80	0.61	0.66
Bupa	0.60	0.60	0.61	0.58	0.59	0.61	0.61	0.60	0.59
Credit	0.48	0.45	0.45	0.57	0.45	0.45	0.50	0.45	0.45
Dia.	0.69	0.01	0.06	0.58	0.00	0.02	0.47	0.01	0.01
Glass	0.51	0.15	0.21	0.53	0.10	0.14	0.52	0.07	0.11
Heart	0.03	0.00	0.01	0.03	0.01	0.00	0.02	0.00	0.00
Iris	0.91	0.73	0.77	0.94	0.69	0.73	0.91	0.57	0.61
Seg.	0.60	0.14	0.21	0.59	0.15	0.18	0.51	0.14	0.17
Vehi.	0.41	0.00	0.02	0.31	0.01	0.01	0.17	0.01	0.01
Wine	0.02	0.01	0.01	0.00	0.01	0.01	0.00	0.00	0.01

Table 4: Best estimated number of fuzzy sets for each dataset.

Method	EU - C	EU - G	F - C	F - G	1-R - C	1-R - G
Breast	2	3	4	4	3	3
Bupa	8	5	3	4	3	5
Credit	9	9	2	2	3	3
Diabetes	2	2	2	3	2	2
Glass	5	4	2	3	3	3
Heart	2	2	2	2	4	3
Iris	4	4	3	3	4	4
Seg.	3	3	2	2	3	3
Vehicle	2	2	3	2	2	2
Wine	3	2	2	2	2	2

cantly better than the 1-R method and the frequency method, with a 95% confidence level. The results of the statistical tests also showed that the 1-R method performed significantly better than the frequency method with a 95% confidence level.

Table 5 shows the average ranks for each of the methods tested for the classic and general fuzzy reasoning methods.

Table 5: Average ranks.

Methods	EU	Frequency	1-R
Classic	1.694	2.411	1.894
General	1.350	2.711	1.939

One of the reasons for the good performance of the equalized universe method may be the fact that it generates fuzzy sets with the same distance amongst them, promoting the creation of rules that will cover equal areas of the attributes partitions, rather than promoting the creation of fuzzy sets in areas with a concentration of examples but with less differences regarding the classes, which could have a deeper effect on the classification rates. The 1-R method, since it uses the classes to adjust the fuzzy sets defining each attribute, should perform

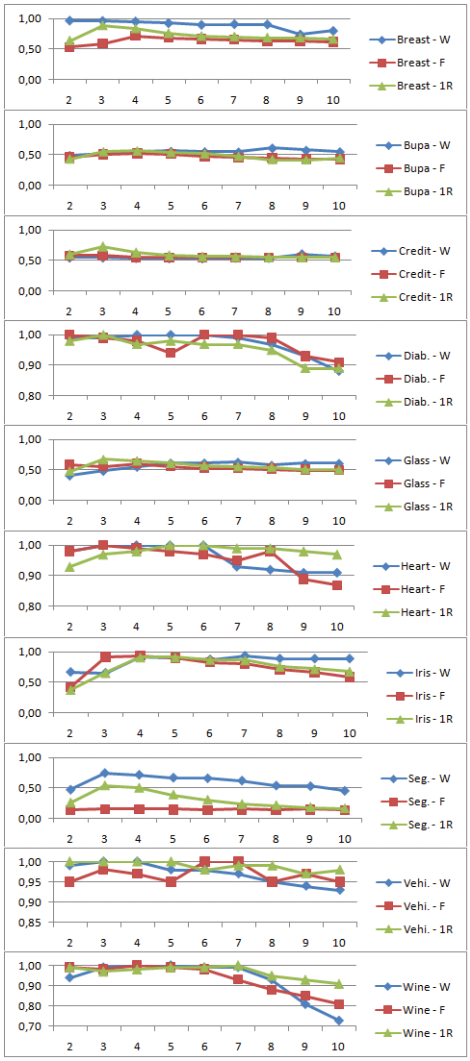


Figure 4: Results for the classic fuzzy reasoning method.

better than the frequency method.

6 Conclusions

In this paper we have tested and compared three different methods for the definition of fuzzy sets, the equalized universe method, the frequency method, and an adaptation of the 1-R supervised method. We have also proposed a heuristic method to estimate the number of fuzzy sets for each attribute using the Wang & Mendel method, to be used by other more costly approaches, able of finding smaller sets of rules with better accuracy. The experiments were carried out with 10 datasets for classification problems using 10-fold cross validation and the classic and general fuzzy reasoning methods. Each attribute of each dataset was defined using from 2 to 10 fuzzy triangular shaped sets. Statistical methods were carried out in order to find significant differences among the methods. The results show that the equalized universe method performed significantly better than the frequency and 1-R methods. The results also showed that the 1-R method performed better than the frequency method. Regarding the estimation of the number of fuzzy sets for the attributes, the proposed method provides support for the researcher to select small intervals of numbers of fuzzy sets that would be worthwhile using in more costly

ISBN: 978-989-95079-6-8

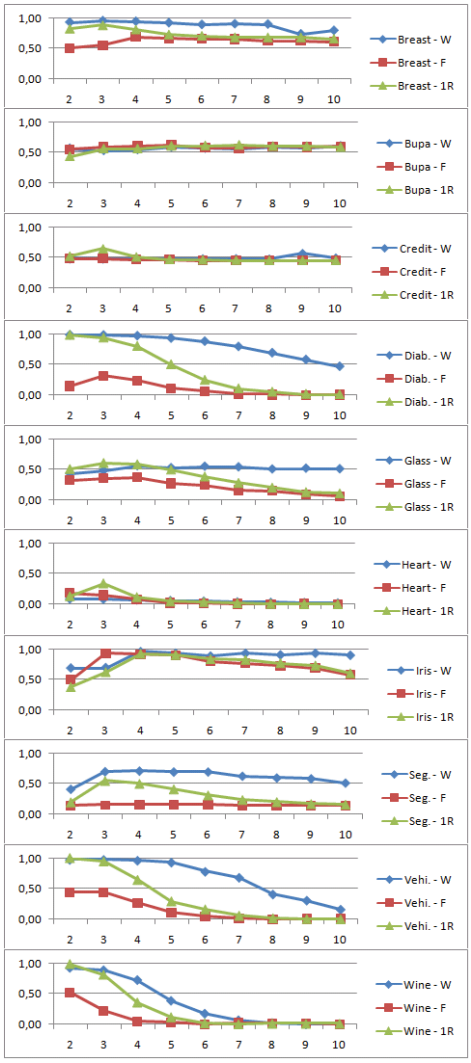


Figure 5: Results for the general fuzzy reasoning method.

approaches, based on the performance of the generated rule bases.

Acknowledgments

This research was supported by CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico, and by the São Paulo State Research Foundation - FAPESP (Proc. No. 2007/05390-0).

References

[1] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[2] Rafael Alcalá, Jesús Alcalá-Fdez, M.J. Gacto, and Francisco Herrera. Genetic learning of membership functions for mining fuzzy association rules. *IEEE International Conference on Fuzzy Systems*, 1:1538–1543, 2007.

[3] Swarup Medasani, Jaeseok Kim, and Raghu Krishnapuram. An overview of membership function generation techniques for pattern recognition. *International Journal of Approxiamate Reasoning*, 19:391–417, 1998.

[4] M. Kaya and R. Alhajj. A clustering algorithm with genetically optimized membership functions for fuzzy association rules mining. *12th IEEE International Conference on Fuzzy Systems*, 2:881–886, 2003.

[5] Ling Zhang and Bo Zhang. The structure analysis of fuzzy sets. *Int. Journal of Approx. Reasoning*, 40:92–108, 2005.

[6] M. S. Chen and S. W. Wang. Fuzzy clustering analysis for optimizing membership functions. *Fuzzy Sets and Systems*, 103:239–254, 1999.

[7] M. E. Cintra, H. A. Camargo, and M. C. Monard. A study on techniques for the automatic generation of membership functions for pattern recognition. *C3N - Congresso da Academia Trinacional de Ciências*, 1:1–10, 2008.

[8] T. Warren Liao, Aivars K. Celmins, and Robert J. Hammell. A fuzzy c-means variant for the generation of fuzzy term sets. *Fuzzy Sets and Systems*, 135:241–257, 2003.

[9] Xuanli Lisa Xie and Gerardo Beni. A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8):841–847, 1991.

[10] J.S. Jang. Adaptive-network-based fuzzy inference systems. *IEEE Trans. on SMC*, 23:665–685, 1993.

[11] Tzung-Pei Hong, Chun-Hao Chen, Yu-Lung Wu, and Yeong-Chyi Lee. A GA-based fuzzy mining approach to achieve a trade-off between number of rules and suitability of membership functions. *Soft Computing*, 10(11):1091–1101, 2006.

[12] Ajith Abraham. Neuro fuzzy systems: State-of-the-art modeling techniques. *Lec. Notes in CS*, 2084:269–276, 2001.

[13] Detlef Nauck, Frank Klawonn, and Rudolf Kruse. *Foundations of Neuro-Fuzzy Systems*. John Wiley & Sons, Inc., 1997.

[14] H. D. Cheng and Jim-Rong Chen. Automatically determine the membership function based on the maximum entropy principle. *Information Sciences*, 96(3-4):163–182, 1997.

[15] Hisao Ishibuchi, Takashi Yamamoto, and Tomoharu Nakashima. Fuzzy data mining: Effect of fuzzy discretization. In *ICDM '01: Proc. of the 2001 IEEE Int. Conf. on Data Mining*, pages 241–248, USA, 2001. IEEE Computer Society.

[16] Grzegorz Nieradka and Bohdan Butkiewicz. A method for automatic membership function estimation based on fuzzy measures. *International Fuzzy Systems Association World Congress - IFSA 2007*, 4529:451–460, 2007.

[17] Weibei Dou, Yuan Ren, Qian Wu, Su Ruan, Yanping Chen, Daniel Bloyet, and Jean-Marc Constans. Fuzzy kappa for the agreement measure of fuzzy classifications. *Neurocomputing*, 70(4-6):726–734, 2007.

[18] Hisao Ishibuchi, Ken Nozaki, and Hideo Tanaka. Efficient fuzzy partition of pattern space for classification problems. *Fuzzy Sets and Systems*, 59(3):295–304, 1993.

[19] J.M. Keller, M.R. Gray, and J.A. Given Jr. A fuzzy k-nearest neighbor algorithm. *IEEE Transactions on Systems, Man, and Cybernetics*, 15(4):580–585, 1985.

[20] Enrique H. Ruspini. A new approach to clustering. *Information and Control*, 15(1):22–32, 1969.

[21] Anthony J. Jakeman Jun and Michael McAleer Bai. *Effects of Misspecification in Estimating the Percentiles of Some Two- & Three-Parameter Distributions*. ANU, Australia, 1989.

[22] Robert C. Holte. Very simple classification rules perform well on most commonly used datasets. *Machine Learning*, 11(1):63–90, 1993.

[23] L. Wang. The WM method completed: a flexible fuzzy system approach to data mining. *IEEE International Conference on Fuzzy Systems*, 11:768–782, 2003.

[24] Janez Demšar. Statistical comparison of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7(1):1–30, 2006.

ISBN: 978-989-95079-6-8

Discriminatory Components for Pattern Classification

Nick J. Pizzi[1,2,3] Witold Pedrycz[4]

1 Institute for Bidiagnostics, National Research Council
Winnipeg, Canada
2 Department of Computer Science, University of Manitoba
Winnipeg, Canada
3 Department of Applied Computer Science, University of Winnipeg
Winnipeg, Canada
4 Electrical and Computer Engineering, University of Alberta
Edmonton, Canada
Email: pizzi@cs.umanitoba.ca, pedrycz@ece.ualberta.ca

Abstract—*The pattern recognition literature is replete with the use of principal component analysis in the interpretation and analysis of data. However, in the specific case of classification, especially of biomedical patterns, this pre-processing method, which transforms possibly correlated features into a new set of uncorrelated variables, must be used with caution since a principal component, which may account for significant variance in the data, is not necessarily discriminatory. To compensate for this deficiency, we present a novel classification method using an adaptive network of fuzzy logic connectives to select the most discriminatory principal components. We empirically demonstrate the effectiveness of this method using a benchmark combination of a conventional classifier and principal component analysis.*

Keywords— fuzzy logic network; principal component analysis; biomedical data; pattern classification.

1 Introduction

Today's biomedical instrumentation provides the acquisition of complex data rich in information content; however, its analysis and interpretation is often difficult and requires the latest pattern analysis methodologies [1,2,3]. This is particularly true for the domain of biomedical pattern classification, the prediction by a classifier of the class (for example, normal or abnormal) to which a pattern (for instance, an infrared spectrum of a biofluid) belongs. This prediction is validated against a gold standard, an external reference test such as a pathologist's expert assessment of the same biofluid. Although many classifier methods exist [4,5,6], the most successful approaches combine them with well designed pre-processing techniques, which simplify, in some sense, the feature space prior to presentation to a classifier, and a sound validation protocol to ensure realistic and clinically useful results.

A standard approach to biomedical pattern classification is to use simple linear classifiers coupled with pre-processing transformations that create new features (coordinates, parameters) ordered by the cumulative variance of the original features. A common pairing is linear discriminant analysis coupled with principal component analysis. The ordering allows for the reduction of the feature space by using only the first few components that account for the bulk of the data variance. While this combination is often successful when used in the classification of patterns, its success lies more in the feature reduction aspect rather than the exploitation of the data variance. This is because the feature variance is not concomitant with the discriminatory power of individual features.

We present a new classification method that uses an adaptive network of fuzzy logic connectives that operate on the entire set of principal components. This method exploits the discriminatory power of any principal component regardless of the amount of variance for which it accounts. We empirically demonstrate that this method produces superior classification performance compared to a benchmark using the conventional approach of linear discriminant analysis operating on the first few principal components of the original feature space. Section 2 presents a general discussion on classification including classifier validation, principal component analysis, linear discriminant analysis, and a recent fuzzy adaptive logic network on which the current method is based. Details of our novel approach are presented in Section 3. The synthetic and biomedical datasets, experiment design, and results are discussed in Section 4 followed by some concluding remarks and areas of future investigations.

2 Pattern classification

2.1 Classifier Validation

We begin by defining formal pattern classification notation: N is the number of patterns (samples, vectors, individuals, or cases); n is the number of features (dimensions, attributes, coordinates, or measurements); c is the number of classes (groups); and $\boldsymbol{X}=\{(\boldsymbol{x}_k,\omega_k),\ k=1,2,\ldots,N\}$ is a set of N labeled patterns where $\boldsymbol{x}_k\in\Re^n$ and $\omega_k\in\Omega=\{1,2,\ldots,c\}$. A classifier is a system that determines a mapping $f:\boldsymbol{X}\rightarrow\Omega$. If a classifier predicts that the class label for pattern $\boldsymbol{x}_i$ is ω_p then: a correct classification occurs if $\omega_p=\omega_i$; otherwise it is considered a misclassification.

Many investigations involving data classification are biased as they use the entire dataset to determine the mapping. This approach leads to unrealistic classification results that do not take into account the possibility of overfitting, wherein the

ISBN: 978-989-95079-6-8

mapping becomes a simple table lookup, between the patterns and class labels (that is, it possesses no generalized predictive power for new patterns). To compensate for this bias, it is essential to perform some type of validation. For instance, patterns in $\boldsymbol{X}$ may be randomly allocated to a design subset, $\boldsymbol{X}^D$ containing N^D patterns, or a validation subset, $\boldsymbol{X}^V$ containing N^V patterns ($N^D+N^V=N$). Now, a mapping is determined using only design patterns, $f':\boldsymbol{X}^D\rightarrow\Omega$, but the classification performance is measured using f' with the validation patterns.

Performance of a classification system is measured using the $c\times c$ confusion matrix, $\boldsymbol{C}$, of the desired class labels (as determined by an external reference test or "gold standard") versus the predicted class labels (as generated by the classifier). If the class prediction for $\boldsymbol{x}_i$ is ω_p then element $[\omega_p,\omega_i]$ of $\boldsymbol{C}$ is incremented by one. The conventional performance measure is the ratio of correctly classified patterns to the total number of patterns, $P_O=(\sum_i n_{ii})/N^V$ (i=1, 2,…, c), where n_{ij} is the number of class i validation patterns that are predicted to belong to class j.

A significant problem with P_O occurs when there is a large disparity between class sizes. For instance, if a validation set has many more patterns in one class than another, the classification accuracy relating to the former will outweigh any effects from the latter. For instance, a high value for P_O may occur due to few misclassifications with the larger class even though many misclassifications occur with the smaller class. This is especially problematic for biomedical patterns where it is relatively easy to acquire patterns from "normal" samples (tissues, biofluids, and so on) but clinically difficult to obtain "abnormal" samples. As a result, it is often better to forgo the use of P_O for these types of datasets and use the class-wise average accuracy, P_A

$$P_A = c^{-1}\sum_i^N \frac{n_{ii}}{\sum_j^N n_{ij}} (i,j=1,\ldots,c) \qquad (1)$$

2.2 Principal Component Analysis

The motivation behind principal component analysis, first described by Pearson [7] (with a practical computing method described by Hotelling [8]), is to find a set of directions (coordinates) that explain as much of the variability of the original data as possible. In other words, the principal components are a new set of orthogonal linear coordinates such that the variances of the original features with respect to these derived coordinates are in decreasing orders of magnitude [9]. As a result, each principal component is uncorrelated with all other principal components (in a normal distribution, they are statistically independent). Moreover, it can be shown [10] that no other set of m coordinates can account for more of the variability in the original data than the first m principal components.

The first principal component, $\boldsymbol{Y}_1$, of the original features x_1, x_2, …, x_n, is the linear combination

$$\boldsymbol{Y}_1 = \sum_i^n a_{1i}x_i, \sum_i^n a_{1i}^2 = 1 \qquad (2)$$

The constraint on the coefficients is necessary; otherwise the variance of $\boldsymbol{Y}_1$ can be increased simply by increasing the value of any coefficient. The second principal component, $\boldsymbol{Y}_2$, would be computed in a similar fashion to (1). Fig. 1 is a plot of some bivariate data and their two principal components. It is clear from this figure that an additional constraint, orthogonality to the first principal component, is required to compute the second principal component, otherwise it would simply be driven to the first principal component. Orthogonality is ensured by restricting the variables of the second principal component to those that are uncorrelated with the first principal component. As a result of this orthogonality constraint, if there are n features then there can be up to n principal components [11]. In fact, if the original features are completely uncorrelated, then all n principal components must be used to take into account the variance in the original features. In this case, principal component analysis serves little purpose, with respect to pattern classification, since the motivation behind the technique is to reduce the dimensionality of the original input (feature) space. However, in "real-world" high-dimensional biomedical patterns, the converse is usually true; features are highly correlated and hence only $1\leq m<<n$ principal components are required to account for all (or nearly all) of the variation.

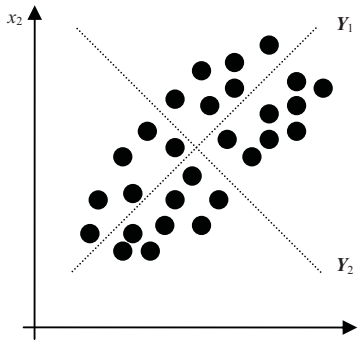


Figure 1: Principal components, $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$, for some bivariate data.

Determining the principal components is a straightforward process involving the computation of the eigensystem of the original data's covariance matrix, $\boldsymbol{V}$, whose element v_{lm} is the sample covariance between features x_l and x_m

$$v_{lm} = \frac{1}{N-1}\sum\nolimits_N \left(x_{il}-\mu_l\right)\left(x_{im}-\mu_m\right) \qquad (3)$$

where μ_j is the mean for feature x_j (cf. [12] for a derivation of the proof). The variances of the principal components are the eigenvalues of $\boldsymbol{V}$, $\lambda_1\geq\lambda_2\geq\ldots\geq\lambda_n\geq 0$ (the covariance matrix is quadratic and hence admits no negative eigenvalues). The variance of a principal component, $\boldsymbol{Y}_i$, is λ_i and its constants a_{i1}, a_{i2}, …, a_{in} are the elements of the corresponding eigenvector.

A standard strategy employed in biomedical pattern classification is to take the first m principal components whose cumulative variance exceeds some pre-defined threshold. This reduction is often significant ($m<<n$): for instance in some high-dimensional infrared spectra more than 80% of the cumulative variance may be accounted for by only the first one or two principal components [13].

ISBN: 978-989-95079-6-8

While this pre-processing strategy may be effective, it must be noted that the components are ordered by maximal variance. Unfortunately, this does not necessarily translate into maximal discriminatory power [14]. For instance, if the method used to acquire values for a particular feature is extremely prone to measurement error, then this feature will have a high variance. Now, assuming this variance is greater than other features, the first principal component will be approximately equal to this suspect feature, and hence, this principal component will be useless in discriminating between classes. Conversely, a highly discriminatory feature may have an extremely small variance and hence will not contribute to the first few principal components. In summary, maximal discriminatory power is not equivalent to maximal variance.

2.3 Linear Discriminant Analysis

Linear discriminant analysis (LDA) [15] is a standard pattern classification strategy that determines linear boundaries between c classes while taking into account between-class and within-class variances. If the error distributions for the classes are the same (identical covariance matrices), LDA constructs the optimal linear boundary between the classes. In real-world situations, this optimality is seldom achieved since different classes typically give rise to different distributions.

LDA assigns a pattern, $\boldsymbol{x}$, to class i for which the probability distribution, $p_i(\boldsymbol{x})$, is greatest. That is, $\boldsymbol{x}$ is allocated to class i, if $q_i p_i(\boldsymbol{x}) \geq q_j p_j(\boldsymbol{x})$ $(\forall j=1,2,\ldots,c\ [j\neq i])$, where q_i is the class' prior (or proportional) probability. The discriminant function for class i is

$$D_i(\boldsymbol{x}) = \log q_i + \boldsymbol{\mu}_i^{\mathrm{T}} \boldsymbol{W}^{-1} \left(\boldsymbol{x} - \tfrac{1}{2} \boldsymbol{\mu}_i \right) \qquad (4)$$

where $\boldsymbol{\mu}_i$ is the mean for class i and $\boldsymbol{W}$ is the covariance matrix of the patterns in $\boldsymbol{X}$. The feature space hyperplane separating class i from j is defined by $F_{ij}(\boldsymbol{x})=D_i(\boldsymbol{x})-D_j(\boldsymbol{x})=0$. As a classification system, LDA is often used with the first m principal components of a dataset rather than the original features. This combination will serve as the classification benchmark against which our novel pattern classification method will be evaluated.

2.4 Fuzzy Adaptive Logic Network

Our novel approach is partly based upon a modification to the fuzzy adaptive logic network (cf. [16] for a thorough description). This network, which can be used for pattern classification, combines two different subsystems within a general architecture. A neurocomputing subsystem uses a set of perceptrons to construct class boundaries. Through a set of weights and respective inputs, a perceptron is defined as $P(\boldsymbol{x},\boldsymbol{w})=f(\sum_i w_i x_i + w_0)$ (f is a sigmoidal transfer function) describing an n-dimensional hyperplane. This geometric information is presented to the logic processing subsystem composed of a layer of fuzzy conjunctions ("and" elements) and disjunctions ("or" elements). The intent is to use these fuzzy logic connectives to combine the hyperplanes from the neurocomputing subsystem to form convex hull-like topologies. For instance, a convex region delineated by p perceptrons may be represented by the compound logic predicate, $Q = P_1(\boldsymbol{x},\boldsymbol{w}_1)$ *and* $P_2(\boldsymbol{x},\boldsymbol{w}_2)$ *and* … *and* $P_p(\boldsymbol{x},\boldsymbol{w}_p)$, which produces values close to one (in other words, approaching the crisp notion of *true*) when all contributing predicates are *true* (that is, the respective perceptrons produce high outputs).

To capture the geometric notion of disjoint regions one may take a union (in the fuzzy set theoretic sense) of the individual regions described by the Q's, $R = Q_1$ *or* Q_2 *or* … *or* Q_q. To implement these fuzzy predicates, one uses t-norms to model the *and* logic connectives and s-norms to model the *or* logic connectives. A t-norm, $\wedge$, is a function $[0,1]^2 \rightarrow [0,1]$ that is commutative, symmetric, monotonic, and satisfies the boundary conditions $x\mathrm{t}0=0$ and $x\mathrm{t}1=x$, while the boundary conditions for the s-norm, $\vee$, are $x\mathrm{s}0=x$ and $x\mathrm{s}1=1$. The fuzzy *or* and *and* connectives may now be defined as

$$\begin{aligned} OR(\boldsymbol{x};\boldsymbol{w}) &= \wedge_i \left(w_i \vee x_i \right) \\ AND(\boldsymbol{x};\boldsymbol{w}) &= \vee_i \left(w_i \wedge x_i \right) \end{aligned} \qquad (5)$$

where $\boldsymbol{x}$ are the inputs and $\boldsymbol{w}$ are the corresponding adjustable weights (connections) confined to the unit interval. In the case of $OR(\boldsymbol{x};\boldsymbol{w})$, the greater the weight value the more relevant the respective input (if all weights are 1, it becomes a standard *or* gate). In the case of $AND(\boldsymbol{x};\boldsymbol{w})$, the greater the weight value the less relevant the respective input (if all weights are 0, it becomes a standard *and* gate). If we restrict ourselves to differentiable t- and s-norms, a gradient descent strategy can be used to train a fuzzy adaptive logic network (cf. [16] for details).

3 Fuzzy logic classification network using principal components

Building upon the concepts described in Section 2, we now describe a novel pattern classification algorithm, COAP (Classification using a fuzzy Or/And network with Principal components), a component of which extends the fuzzy logic network architecture developed by the authors [17]. There are three major steps to the COAP algorithm: (i) apply principal component analysis on the original features to find all n principal components; (ii) Use a genetic algorithm to determine the optimal weights for the fuzzy logic network given the principal component values found in (i); (iii) use the patterns from the validation subset to assess the classification performance using the selected feature regions and principal component values. Fig. 2 illustrates the architecture of the COAP system.

Let us now look at each algorithmic step in more detail. After applying principal component analysis to the feature space, we replace pattern, $\boldsymbol{x}_i=[x_1,x_2,\ldots,x_n]$, with its respective principal component values, $\boldsymbol{y}_i=[\boldsymbol{Y}_1(\boldsymbol{x}_i),\boldsymbol{Y}_2(\boldsymbol{x}_i),\ldots,\boldsymbol{Y}_n(\boldsymbol{x}_i)]$. These principal component values for all design set patterns are subsequently presented to the fuzzy logic network component.

COAP's fuzzy logic network uses the product $(x_1 \times x_2)$ and probabilistic sum $(x_1+x_2-x_1 \times x_2)$ for the t- and s-norms, respectively, with p (user selected) *AND* connectives and c *OR* connectives. There are two issues with this network that

ISBN: 978-989-95079-6-8

do not exist with the fuzzy adaptive logic network described in Section 2.4. First, while output from a perceptron maps onto the unit interval (due to the sigmoidal nature of its transfer function), which is necessary for input into a fuzzy logic *AND* connective, principal component values map onto $\Re$. This can be dealt with by rescaling the principal component values prior to presentation to the fuzzy logic network ((*x*–*min*)÷(*max*–*min*), where *min* and *max* are the respective minimum and maximum for all principal component values).

The second, more serious, issue is that a gradient descent strategy cannot be used to minimize the fuzzy logic network error (optimize the weights) since the weight adjustments are now based on sets of principal components rather than differentiable perceptron output. We deal with this issue by using a straightforward implementation of a genetic algorithm [18,19,20] to perform the structural optimization of the network. While much slower than gradient descent, this solution still provides adequate computational performance. In this study, we implemented a conventional genetic algorithm as described in [21], but other more sophisticated genetic algorithm variants could certainly be explored.

As mentioned in Section 2.1, $\Omega=\{1,2,\ldots,c\}$; however, it is often beneficial [13] to use 1-of-*c* encoding for the class labels for iterative classifiers such as artificial neural networks or fuzzy logic networks, namely, $\Omega=\{\gamma_1,\gamma_2,\ldots,\gamma_c\}$ where, for $\boldsymbol{x}_i$, $\gamma_{\omega_i}=1$ and $\gamma_{\omega_j}=0$ $(\forall\omega_i\neq\omega_j)$. Instead of using one output element to represent all *c* classes, we use *c* output elements.

Finally, all performance results using P_A are based on the class predictions of COAP using the biomedical patterns from the validation subset. That is, the principal component values are obtained for the validation patterns and presented to the fuzzy logic network component. Subsequently, the predictions are compared against the desired class labels (if the output element with the maximum value corresponds to the desired label's non-zero component then we have a successful classification, otherwise it is a misclassification).

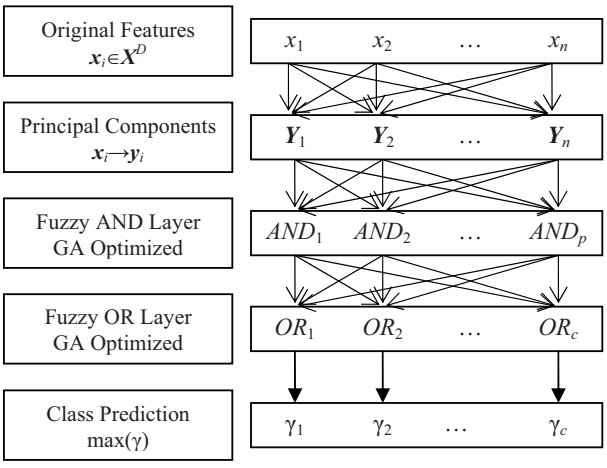


Figure 2: General architecture of COAP with the GA (genetic algorithm) optimized fuzzy logic network.

4 Experiments and discussion

4.1 Synthetic Datasets

We begin our experiments with the "exclusive or" dataset (*n*=2, *c*=2, *N*=4): patterns {{0,0},{1,1}} and {{0,1},{1,0}}; and respective principal component values of {{0,0},{1,–1}} and {{1,0},{0,–1}}. Intuitively, one expects that LDA would perform poorly in this case as no hyperplane can act as a class boundary to perfectly separate the patterns. Using LDA with the principal components, this is actually the case with P_A=0.5 (one misclassification for each class). [As this is a strictly pedagogical experiment, we skip validation.] Setting the initial genetic algorithm population to 50, the number of iterations to 5, and the number of *AND* connectives to 2, we now get perfect accuracy, P_A=1.0. The weights for the two *AND* connectives are {0.95,0.05} and {0.0,1.0}. The weights for the two *OR* connectives are {0.53,0.05} and {0.07,0.35}.

This next dataset is a variant of the exclusive or dataset described above (*n*=10, *c*=2, *N*=400, $\boldsymbol{x}\in[0,1]^n$). A pattern belongs to the first class, if all of its features are identical; otherwise it belongs to the second class. Fig. 3 is a plot of the first two features of this dataset. The initial genetic algorithm population is 100, the number of iterations is 5, and the number of *AND* connectives is 10. In this case, LDA with principal components also performed poorly, P_A=0.5, while COAP produced a higher classification accuracy, P_A=0.81.

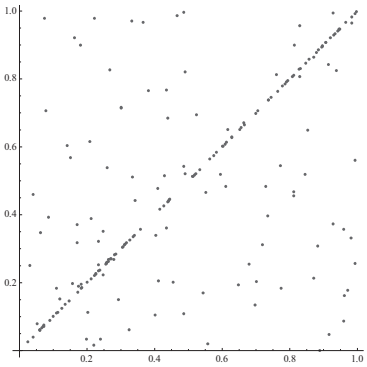


Figure 3: Plot of, x_1 and x_2, for the second synthetic dataset.

4.2 Biomedical Dataset

We used a biomedical dataset from the Machine Learning Repository at the University of California, Irvine [22]. This archive is used to evaluate machine learning algorithms. This dataset [23,24] comprises voice measurements acquired from two classes of subjects: those with Parkinson's disease (abnormal); and those without (normal). There are *N*=195 patterns from one of two classes: N_a=48 abnormal patterns; N_n=147 normal patterns. The *n*=22 features include: the maximum, minimum, and average fundamental vocal frequency (3 features); several measures of variation in fundamental frequency (5 features); several measures of variation in amplitude (6 features); measures of the ratio of noise to tonal components in the voice (2 features); nonlinear dynamical complexity measures (2 features); a

ISBN: 978-989-95079-6-8

signal fractal scaling exponent (1 feature); and nonlinear measures of fundamental frequency variation (3 features).

For this dataset, the following COAP parameters were used: p=7 (the number of *AND* connectives); crossover rate, 0.07; mutation rate, 0.005; size of genetic algorithm population, 200; and 20 iterations of the genetic algorithm. The patterns were randomly assigned to either a design subset (N^D=64 with 32 normal patterns and 32 abnormal patterns) or a validation subset (N^V=131 with 115 normal patterns and 16 abnormal patterns). Finally, given the class size disparity between the abnormal and normal patterns within the validation set, P_A was used to assess classification accuracy.

Table 1 lists the confusion matrices for COAP for the design patterns and validation patterns. For the design patterns, P_A=0.91, while P_A=0.85 for the validation patterns. While only 70% of the normal validation patterns were correctly classified, all abnormal validation patterns were correctly classified. Table 2 lists the confusion matrices for LDA using all n principal components. In the case of this benchmark, P_A=0.94 for the design patterns and P_A=0.70 for the validation patterns. The disparity between the design and validation results is a classic sign of potential overfitting. While 75% of the abnormal validation patterns were correctly classified, only 65% of the normal patterns were correctly classified.

Table 1: COAP confusion matrices for $\boldsymbol{X}^D$ and $\boldsymbol{X}^V$.

Desired vs	Design Set (P_A=0.91)		Validation Set (P_A=0.85)	
Predicted	Abnormal	Normal	Abnormal	Normal
Abnormal	29	3	16	0
Normal	3	29	35	80

Table 2: Benchmark confusion matrices for $\boldsymbol{X}^D$ and $\boldsymbol{X}^V$.

Desired vs	Design Set (P_A=0.94)		Validation Set (P_A=0.70)	
Predicted	Abnormal	Normal	Abnormal	Normal
Abnormal	31	1	12	4
Normal	3	29	40	75

Table 3 lists the confusion matrices (desired versus predicted class labels) using successive combinations of principal component in turn (ordered by variance) for both the design and validation sets. Note that: P_A is listed for the design set followed by the validation set; $\sum^{\alpha}\lambda_i$ refers to the cumulative variance of the first α principal components used to produce the corresponding results; and the last entry is the same as that described in Table 2. The best result, P_A=0.781, occurred when using the first 13 principal components. It is clear with this dataset that feature variance is not well correlated with discriminatory features (for example, the first three principal components, which account for 0.998 of the cumulative variance, produce a low accuracy score, P_A=0.60).

Fig. 3 summarizes these results by plotting the COAP validation result against the validation results using all successive combinations of principal components. It should also be noted that the last few entries in Table 3 clearly demonstrate a problem with overfitting with a disparity of approximately 0.24 between the design and validation results.

Table 3: Results using successive principal components.

	D vs P	$\boldsymbol{X}^D$		$\boldsymbol{X}^V$	
		A	N	A	N
$\sum^{1}\lambda_i$=0.729	A	24	8	8	8
P_A=0.77/0.63	N	7	25	27	88
$\sum^{2}\lambda_i$=0.947	A	22	10	7	9
P_A=0.69/0.61	N	10	22	25	90
$\sum^{3}\lambda_i$=0.998	A	23	9	7	9
P_A=0.70/0.60	N	10	22	27	88
$\sum^{4}\lambda_i$=1.000	A	22	10	9	7
P_A=0.73/0.62	N	7	25	36	79
$\sum^{5}\lambda_i$=1.000	A	28	4	10	6
P_A=0.83/0.69	N	7	25	29	86
$\sum^{6}\lambda_i$=1.000	A	28	4	10	6
P_A=0.83/0.69	N	7	25	29	86
$\sum^{7}\lambda_i$=1.000	A	28	4	10	6
P_A=0.84/0.69	N	7	25	29	86
$\sum^{8}\lambda_i$=1.000	A	28	4	11	5
P_A=0.83/0.72	N	7	25	32	83
$\sum^{9}\lambda_i$=1.000	A	27	5	11	5
P_A=0.81/0.70	N	7	25	32	83
$\sum^{10}\lambda_i$=1.000	A	27	5	12	4
P_A=0.81/0.74	N	7	25	32	83
$\sum^{11}\lambda_i$=1.000	A	29	3	13	3
P_A=0.92/0.74	N	2	30	38	77
$\sum^{12}\lambda_i$=1.000	A	29	3	13	3
P_A=0.92/0.75	N	2	30	37	78
$\sum^{13}\lambda$=1.000	A	30	2	14	2
P_A=0.92/0.78	N	3	29	36	79
$\sum^{14}\lambda$=1.000	A	30	2	14	2
P_A=0.92/0.78	N	3	29	37	78
$\sum^{15}\lambda_i$=1.000	A	30	2	14	2
P_A=0.92/0.78	N	3	29	37	78
$\sum^{16}\lambda_i$=1.000	A	32	0	12	4
P_A=0.94/0.71	N	4	28	39	76
$\sum^{17}\lambda_i$=1.000	A	32	0	14	2
P_A=0.95/0.76	N	3	29	41	74
$\sum^{18}\lambda_i$=1.000	A	31	1	12	4
P_A=0.94/0.70	N	3	29	40	75
$\sum^{19}\lambda_i$=1.000	A	31	1	12	4
P_A=0.94/0.70	N	3	29	40	75
$\sum^{20}\lambda_i$=1.000	A	31	1	12	4
P_A=0.94/0.70	N	3	29	40	75
$\sum^{21}\lambda_i$=1.000	A	31	1	12	4
P_A=0.94/0.70	N	3	29	40	75
$\sum^{22}\lambda_i$=1.000	A	31	1	12	4
P_A=0.94/0.70	N	3	29	40	75

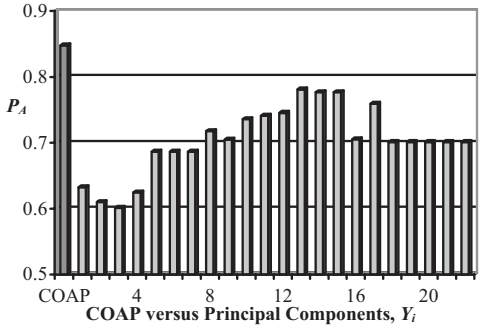


Figure 3: Summary of validation set accuracies.

ISBN: 978-989-95079-6-8

Finally, after examination of the weights of the fuzzy logic network, six principal components tended to significantly contribute to the COAP training, $\boldsymbol{Y}_1$, $\boldsymbol{Y}_4$, $\boldsymbol{Y}_5$, $\boldsymbol{Y}_{10}$, $\boldsymbol{Y}_{12}$, and $\boldsymbol{Y}_{16}$. This is further evidence that care must be taken when using principal component analysis as one may overlook discriminatory components simply because they account for little feature variance (for example, the variance for $\boldsymbol{Y}_{10}$, $\boldsymbol{Y}_{12}$, and $\boldsymbol{Y}_{16}$ is zero).

5 Conclusions

We have empirically demonstrated the effectiveness of a novel classification method that uses an adaptive network of fuzzy logic connectives to combine new features generated using principal component analysis. Using a "real world" biomedical dataset, COAP correctly classified significantly more patterns from a validation set compared to the benchmark.

While this novel classification method has demonstrated the utility of merging fuzzy logic connectives with multivariate statistical discrimination, the investigation has also led to the identification of future areas of research to potentially improve its overall effectiveness and computational performance. First, rather than setting the number of fuzzy *and* connectives by the user *a priori*, it would be worthwhile to investigate a cascade approach to determining an optimal number of *and* connections that would be completely data-driven. Second, alternative structural optimizations to the fuzzy logic network need to be examined beginning with more sophisticated evolutionary computational approaches or exploiting recent advances in stochastic optimization techniques. A final area of investigation is using more sophisticated component analysis methods for feature pre-processing (non-linear and local principal component methods, other kernel based approaches, and fuzzy set based feature selection/extraction approaches).

Acknowledgment

We are grateful to Max Little of the University of Oxford who, in collaboration with the National Centre for Voice and Speech, Denver, Colorado, recorded the speech signals for the contribution of the biomedical dataset within the publicly available machine learning repository. We thank Drs. Asuncion and Newman of the University of California, Irvine who maintain the highly useful and publicly available data repository.

Conrad Wiebe and Aleksander Demko are gratefully acknowledged for the implementation and testing of some of the software framework components used in the implementation of this pattern classification method and validation protocol.

This investigation was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1] D.L. Pavia, G.M. Lampman, G.S. Kriz and J.A. Vyvyan. *Introduction to Spectroscopy*. Fort Worth: Harcourt Brace College Publishers, 2008.

[2] H. Friebolin. *Basic One- and Two-Dimensional NMR Spectroscopy*. New York: Wiley and Sons, 1998.

[3] M. Akay. *Nonlinear Biomedical Signal Processing, Volume I: Fuzzy Logic, Neural Networks, and New Algorithms*. New York: IEEE Press, 2000.

[4] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. New York: Wiley Interscience, 2000.

[5] B. Bouchon-Meunier, G. Coletti, R.R. Yager. *Modern Information Processing: From Theory to Applications*. Amsterdam: Elsevier, 2006.

[6] I.H. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques (2nd edition)*. San Francisco: Morgan Kaufmann Publishers, 2005.

[7] K. Pearson. On lines and planes of closest fit to a system of points in space. *Philosophical Magazine*, 2:557–572, 1901.

[8] H. Hotelling. Analysis of a complex of statistical variables into principal components. *J Educational Psychology*, 24:417–441, 1933.

[9] I.H. Bernstein, C.P. Garbin, and G.K. Teng. *Applied Multivariate Analysis*. New York: Springer-Verlag, 1988.

[10] W. Krzanowski. *Principles of Multivariate Analysis*. New York: Oxford University Press, 1988.

[11] B.F.J. Manly. *Multivariate Statistical Methods: A Primer*. New York: Chapman & Hall, 1986.

[12] R. Gnanadesikan. *Methods for Statistical Data Analysis of Multivariate Observations*. New York: John Wiley, 1977.

[13] N.J. Pizzi, L.-P. Choo, J. Mansfield, M. Jackson, W.C. Halliday, H.H. Mantsch, R.L. Somorjai. Neural network classification of infrared spectra of control and Alzheimer's diseased tissue. *Artificial Intelligence in Medicine*, 7:67–79 (1995).

[14] J.D. Jobson. *Applied Multivariate Data Analysis: Categorical and Multivariate Methods*. New York: Springer–Verlag, 1992.

[15] G.A.F. Seber. *Multivariate Observations*. Hoboken: Wiley and Sons, 1984.

[16] W. Pedrycz, A. Breuer, and N.J. Pizzi. Fuzzy adaptive logic networks as hybrid models of quantitative software engineering. *Intelligent Automation and Soft Computing*, 12:189–209, 2005.

[17] N.J. Pizzi and W. Pedrycz. A fuzzy logic network for pattern classification. *Proceedings of the Annual Meeting of the North American Fuzzy Information Processing Society*, June 14–17, Cincinnati, USA, 2009 (submitted for review).

[18] K.A. De Jong. *Evolutionary Computation: A Unified Approach*. Cambridge: MIT Press, 2006.

[19] D.E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading: Addison-Wesley, 1989.

[20] R.L. Haupt and S.E. Haupt. *Practical Genetic Algorithms*. Hoboken: John Wiley & Sons, 2004.

[21] C. Jacob. *Illustrating Evolutionary Computation with Mathematica*. San Diego: Academic Press, 2001.

[22] A. Asuncion and D.J. Newman. *UCI Machine Learning Repository* [http://www.ics.uci.edu/~mlearn/MLRepository.html]. Irvine, CA: University of California, School of Information and Computer Science, 2007.

[23] M.A. Little, P.E. McSharry, E.J. Hunter, and L.O. Ramig. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Transactions on Biomedical Engineering*, 2008 (to appear).

[24] M.A. Little, P.E. McSharry, S.J. Roberts, D.A.E. Costello, and I.M. Moroz. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *Biomedical Engineering Online*, 6:23, 2007.

ISBN: 978-989-95079-6-8

Image reduction with interval-valued fuzzy sets and OWA operators

D. Paternain, H. Bustince, J. Sanz, M. Galar, C. Guerra
Departamento de Automática y Computación, Universidad Pública de Navarra
Pamplona, Spain
Email: {daniel.paternain,bustince,joseantonio.sanz,mikel.galar,carlos.guerra}@unavarra.es

Abstract— *In this paper we propose a generalization of Atanassov's operators and we prove that these generalized operators and OWA operators of dimension* 2 *provide the same numerical results. We apply Atanassov's operators to image compression and use different families of OWA operators in order to calculate the coefficient* α *of the Atanassov's operator.*

Keywords— Image compressing, OWA operators, Atanassov's operator, Interval-valued fuzzy set.

1 Introduction

Image compression consists of reducing the resolution of an image in order to display it, print it, or even speed up some other computations (see [10, 11, 18]). The target of image reduction algorithms is to reduce the size of the image with the smallest possible quality loss.

Atanassov's operator was given in 1983 ([1, 2]) allowing to associate a fuzzy set to each Atanassov's intuitionistic fuzzy set or interval-valued fuzzy set ([9, 12, 14]). Later, in 1988, Yager presented the Ordered Weighted Averaging (OWA) operators ([16])). Our paper is based on the idea that, under certain conditions, the numerical results of both operators are the same. For this reason, we define an operator generalizing Atanassov's operators, being an OWA operator of dimension 2([3, 4, 5]).

We apply this operator in image compression. In this way, we divide the image in blocks. We associate each block with an interval, getting an interval-valued fuzzy set associated with the image. We use the Atanassov's operator to obtain a single point from each interval. The result is a new (fuzzy) set representing the reduced image.

In order to calculate the α parameter of the Atanassov's operator, we focus on several families of OWA operators of dimension 2. We calculate the α coefficient according to the type of OWA operator, with the aim of getting different reductions from a single image.

This paper is organized in the following way. First, we start recalling some concepts related to interval-valued fuzzy sets and Atanassov's operator. In Section 3, we see the relation between OWA operators of dimension 2 and generalized Atannasov's operators. In Section 4 we present different methods to calculate α, based on different families of OWA operators. In Section 5 we propose an algorithm to compress images and in Section 6 we see different experimental results. We finish with some conclusions and references .

2 Preliminary definitions

We denote by L([0,1]) the set of all closed subintervals of the unit interval [0,1], that is,

$$L([0,1]) = \{\mathbf{x} = [\underline{x}, \overline{x}] | (\underline{x}, \overline{x}) \in [0,1]^2 \text{ and } \underline{x} \leq \overline{x}\}.$$

L([0,1]) is a partially ordered set with respect to the order relationship $\leq_L$ defined in the following way: given **x**,**y** $\in$ L([0,1])

$$\mathbf{x} \leq_L \mathbf{y} \text{ if and only if } \underline{x} \leq \underline{y} \text{ and } \overline{x} \leq \overline{y}.$$

With this order relationship, (L([0,1]),$\leq_L$) is a complete lattice ([7, 8, 3, 4]), where the smallest element is $0_L = [0,0]$ and the largest is $1_L = [1,1]$.

Definition 1. *An interval-valued fuzzy set (IVFS) A on the universe* $U \neq \emptyset$ *is a mapping* $A : U \to L([0,1])$.

M_A(u)=[$\underline{A}$(u),$\overline{A}$(u)] $\in$ L([0,1]) is the membership degree of u $\in$ U, with $\underline{A}$(u), $\overline{A}$(u) $\in$ [0,1] denoting the lower bound and the upper bound respectively of the membership associated to u.

In 1983, Atanassov proposed an operator to associate a fuzzy set with each interval-valued fuzzy set (see [1, 2]). This operator associates each interval with a point.

Definition 2. *Let* $\alpha \in [0,1]$. *The Atanassov's operator* K_α *is a mapping* $K_\alpha : L([0,1]) \to [0,1]$ *defined by*

1. $K_0(x) = \underline{x}$ *for all* $x \in L([0,1])$,
2. $K_1(x) = \overline{x}$ *for all* $x \in L([0,1])$,
3. $K_\alpha(x) = K_\alpha([K_0(x), K_1(x)]) = K_0(x) + \alpha(K_1(x) - K_0(x)) = \underline{x} + \alpha(\overline{x} - \underline{x})$ *for all* $x \in L([0,1])$.

To generalize this operator, the following definition was proposed in [4]:

Definition 3. *Let* $\alpha \in$ *[0,1]. An operator* D_α *is a mapping* $D_\alpha : L([0,1]) \to [0,1]$ *such that it satisfies the following conditions:*

1. *If* $\underline{x} = \overline{x}$, *then* $D_\alpha(\boldsymbol{x}) = \underline{x}$,
2. $D_0(\boldsymbol{x}) = \underline{x}$, $D_1(\boldsymbol{x}) = \overline{x}$ *for all* $\boldsymbol{x} \in L([0,1])$,
3. *If* $\boldsymbol{x} \leq_L \boldsymbol{y}$ *with* $\boldsymbol{x}, \boldsymbol{y} \in L([0,1])$, *then* $D_\alpha(\boldsymbol{x}) \leq D_\alpha(\boldsymbol{y})$,
4. $D_\alpha([0,1]) = \alpha$ *for any* $\alpha \in [0,1]$.

Example 1.

$$D_\alpha([\underline{x}, \overline{x}]) = \begin{cases} \overline{x} & \text{if } \overline{x} \leq \alpha \\ \underline{x} & \text{if } \underline{x} \geq \alpha \\ \alpha & otherwise \end{cases}$$

Proposition 1. *For any* $\alpha \in [0,1]$, K_α *is a* D_α *operator.*

Next, we propose a theorem to construct D_α operators using one-variable real functions.

ISBN: 978-989-95079-6-8

Theorem 1. *Let $\alpha \in [0,1]$ and let $f : [0,1] \to [0,1]$ be a continuous and strictly increasing function. Then the operator*

$$D_\alpha : L([0,1]) \to [0,1] \text{ given by}$$
$$D_\alpha(\boldsymbol{x}) = f^{-1}\Big(pf(\underline{x}) + (1-p)f(\overline{x})\Big)$$

with $p = \frac{f(1)-f(\alpha)}{f(1)-f(0)}$, is a D_α operator in the sense of Definition 3.

Proof:

1. If $\underline{x} = \overline{x}$, then $D_\alpha(\mathbf{x}) = f^{-1}(pf(\underline{x}) + (1-p)f(\underline{x})) = = f^{-1}(f(\underline{x})) = \underline{x}$.
2. If $\alpha = 0$, then $p = 1$. In these conditions $D_0([\underline{x}, \overline{x}]) = = f^{-1}(f(\underline{x})) = \underline{x}$.
   If $\alpha = 1$, then $p = 0$ and therefore $D_1([\underline{x}, \overline{x}]) = \overline{x}$.
3. If $[\underline{x}, \overline{x}] \leq_L [\underline{y}, \overline{y}]$, as f is continuous and strict, we have that $D_\alpha([\underline{x}, \overline{x}]) = f^{-1}(pf(\underline{x}) + (1-p)f(\overline{x})) \leq f^{-1}(pf(\underline{y}) + (1-p)f(\overline{y})) = D_\alpha([\underline{y}, \overline{y}])$.
4. If $\mathbf{x} = [0,1]$, then $D_\alpha([0,1]) = f^{-1}(p(f(0) + (1-p)f(1)) = f^{-1}(\frac{f(0)(f(1)-f(\alpha))-f(1)(f(1)-f(\alpha))+f(1)(f(1)-f(0))}{f(1)-f(0)}) = f^{-1}(\frac{f(\alpha)(f(1)-f(0))}{f(1)-f(0)}) = \alpha$.

Remark: In this paper, we take $f(x) = x$. Under this condition, by Theorem 1, we have that

$$D_\alpha([\underline{x}, \overline{x}]) = \underline{x} + \alpha(\overline{x} - \underline{x}) = K_\alpha([\underline{x}, \overline{x}])$$

3 K_α and OWA operators

In [16], Yager introduced the Ordered Weighted Aggregation Operator (OWA operator) in the following way:

Definition 4. *A mapping $F : [0,1]^n \to [0,1]$ is called an OWA operator of dimension n if there exists a weighting vector W,$W = (w_1, w_2, \ldots, w_n) \in [0,1]^n$ with $\sum_i w_i = 1$ and such that*

$$F(a_1, a_2, \ldots, a_n) = \sum_{j=1}^{n} w_j b_j$$

with b_j the j-th largest of the a_i.

In the original definition, Yager considered the OWA operators as a mapping from the whole euclidean space $\mathbb{R}^n$ to $\mathbb{R}$. However, for us it is better to reduce the domain only to $[0,1]^2$.

There exist three important special cases of OWA operators that coincide with well known aggregation functions, as we can see in [17].

1. F^*: The weighting vector, denoted as W^*, is defined as $w_1 = 1$ and $w_j = 0$ for all $j \neq 1$.
   $F^*(a_1, \ldots, a_n) = max\{a_1, \ldots, a_n\}$
2. F_*: The weighting vector, denoted as W_*, is defined as $w_n = 1$ and $w_j = 0$ for all $j \neq n$.
   $F_*(a_1, \ldots, a_n) = min\{a_1, \ldots, a_n\}$.
3. F_A: The weighting vector, denoted as W_A, is defined as $w_j = 1/n$ for all $j \in 1, \ldots, n$.
   $F_A(a_1, \ldots, a_n) = \frac{1}{n}\sum_{i=1}^{n} a_i$.

If we focus on two-dimensional OWA operators using as weighting vector $W = (\alpha, 1-\alpha)$, we can think of applying it to the bounds of an interval. In this case, the numerical result of applying the OWA operator on the bounds of an interval and the operator D_α acting on that interval is the same. While D_α operator acts over elements in L([0,1]), the domain of OWA operators is [0,1]x[0,1]. For this reason, OWA operators acting on the unit square require an ordering operation to ensure the elements to be the extremes of an interval defined in L([0,1]). That's the reason why we need the following theorems to study the relationship between these operators.

We define a new operator $\mathbb{D}_\alpha$ by composing the D_α operator with the map

$$\begin{array}{rcl} i : [0,1]^2 & \to & L([0,1]) \\ (x,y) & \to & [min(x,y), max(x,y)] \end{array}$$

Theorem 2. *1. Let $\alpha \in [0,1]$ and $\mathbb{D} = D_\alpha \circ i$ where D_α is the operator given in Definition 3. Then, if F is the OWA operator (of dimension 2) defined by the weighting vector $W = (\alpha, 1-\alpha)$, we have that*
$$\mathbb{D}_\alpha(x,y) = F(x,y) \text{ for all } x, y \in [0,1].$$

2. Let F be an OWA operator (of dimension 2) with weighting vector $W = (w_1, w_2)$. Then for any $(x,y) \in [0,1]^2$ we have that
$$F(x,y) = \mathbb{D}_\alpha(x,y), \text{ with } \alpha = w_1.$$

4 Calculation methods for α coefficient

In this section, we study how to calculate the α coefficient basing on different families of OWA operators. First, we study two measures defined in [17] and associated with any OWA operator: the orness and the dispersion.

Definition 5. *Let F be an OWA operator and W its weighting vector. The orness measure is defined as*

$$orness(F) = \frac{1}{(n-1)}\sum_{i=1}^{n}(n-i)w_i.$$

With this definition it is easy to see that $orness(F^*) = 1$, $orness(F_*) = 0$ and $orness(F_A) = 0.5$.

Proposition 2. $orness(\mathbb{D}_\alpha) = \alpha$

Proof:

$$orness(\mathbb{D}_\alpha(x)) = \frac{1}{2-1}\sum_{i=1}^{2}(2-i)w_i = w_1 = \alpha$$

Definition 6. *Let F be an OWA operator and W its weighting vector. The dispersion measure is defined as*

$$Disp(F) = -\sum_{i=1}^{n} w_i ln w_i.$$

Proposition 3. $Disp(\mathbb{D}_\alpha) = \alpha ln(\frac{1-\alpha}{\alpha}) + ln(1-\alpha)$

Proof:

$$Disp(\mathbb{D}_\alpha) = -\textstyle\sum_{i=1}^{2} w_i ln w_i =$$

$$= \alpha(ln(1-\alpha) - ln(\alpha)) - ln(1-\alpha) =$$

$$= \alpha ln(\tfrac{1-\alpha}{\alpha}) + ln(1-\alpha)$$

ISBN: 978-989-95079-6-8

Next, it is shown the relation between the coefficient α of the operator $\mathbb{D}_\alpha$ and some families of OWA operators.

Definition 7. *F is called a ME-OWA operator if F is an OWA operator such that given a desired value of orness β, it maximizes the dispersion (entropy). In particular we solve the following problem:*

$$\begin{array}{ll} \textit{Max} & -\sum_{i=1}^{n} w_i ln w_i \\ \textit{subject to} & \beta = 1/(n-1)\sum_{i=1}^{n}(n-i)w_i \\ \textit{where} & \sum_{i=1}^{n} w_i = 1,\ w_i \in [0,1]. \end{array}$$

Proposition 4. *The operator $\mathbb{D}_\alpha$ is a ME-OWA for all $\alpha \in [0,1]$, with $orness(\mathbb{D}_\alpha) = \alpha$.*

Definition 8. *F is called a generalized S-OWA operator if F is an OWA operator where*

$$\begin{array}{lcll} w_1 & = & a + \frac{1}{n}(1-(a+b)), & \\ w_i & = & \frac{1}{n}(1-(a+b)), & \textit{with } i = 2,\ldots,n-1, \\ w_n & = & b + \frac{1}{n}(1-(a+b)). & \end{array}$$

with $a, b \in [0,1]$ and $a+b \leq 1$.

Proposition 5. *Let $a, b \in [0,1]$ such that $a+b \leq 1$. If we take*

$$\alpha = \frac{1+a-b}{2}$$

then $\mathbb{D}_\alpha$ is a generalized S-OWA operator.

Definition 9. *F is called a BADD-OWA operator if F is an OWA operator where*

$$w_i = \frac{b_i^\beta}{\sum_{j=1}^{n} b_j^\beta}$$

with $\beta \geq 0$ and being b_i as in Definition 4.

Proposition 6. *Let $\beta \geq 0$. If we take*

$$\alpha = \frac{(\mathbb{D}_1(x,y))^\beta}{(\mathbb{D}_0(x,y))^\beta + (\mathbb{D}_1(x,y))^\beta}$$

with $x, y \in [0,1]$, then $\mathbb{D}_\alpha$ is a BADD-OWA operator.

Definition 10. *Let $\beta \geq 0$. F is called a modified BADD-OWA operator if F is an OWA operator where*

1. $w_i = \frac{(1/b_i)^\beta}{\sum_{j=1}^{n}(1/b_j)^\beta}$ *or*
2. $w_i = \frac{(1-b_i)^\beta}{\sum_{j=1}^{n}(1-b_j)^\beta}$ *or*
3. $w_i = \frac{1}{n-1}(1 - \frac{b_i^\beta}{\sum_{j=1}^{n} b_j^\beta})$ *or*
4. $w_i = \frac{(b_{n-i+1})^\beta}{\sum_{j=1}^{n} b_j^\beta}$

being b_i as in Definition 4.

Proposition 7. *Let $\beta \geq 0$. The following items are satisfied:*

1. *If we take*

$$\alpha = \frac{(1/\mathbb{D}_1(x,y))^\beta}{(1/\mathbb{D}_1(x,y))^\beta + (1/\mathbb{D}_0(x,y))^\beta} =$$

$$= 1 - \frac{(\mathbb{D}_1(x,y))^\beta}{(\mathbb{D}_1(x,y))^\beta + (\mathbb{D}_0(x,y))^\beta} = \frac{(\mathbb{D}_0(x,y))^\beta}{(\mathbb{D}_1(x,y))^\beta + (\mathbb{D}_0(x,y))^\beta}$$

then $\mathbb{D}_\alpha$ is a modified BADD-OWA operator in the sense of items 1, 3 y 4 of Definition 10.

2. *If we take*

$$\alpha = \frac{(1-\mathbb{D}_1(x,y))^\beta}{(1-\mathbb{D}_1(x,y))^\beta + (1-\mathbb{D}_0(x,y))^\beta}$$

then $\mathbb{D}_\alpha$ is a modified BADD-OWA operator in the sense of item 2 of Definition 10.

Proof:

1. Obviously, taking into account items 1,3 and 4 of Definition 10 and that $\mathbb{D}_\alpha$ is an OWA operator of dimension 2.
2. Direct $\square$.

5 Image reduction

We consider an image Q as a $N \times M$ matrix. Each coordinate of the pixels in the image Q is denoted as (i,j). The intensity or gray level of the pixel located in (i,j) is represented as q_{ij}, with $0 \leq q_{ij} \leq L-1$ for each $(i,j) \in Q$.

In our approach to image reduction we use interval-valued fuzzy sets and D_α operator. It has been shown (see [13, 6, 15]) that interval-valued fuzzy sets used in images allow the development of algorithms in several topics as edge detecion, contrast or thresholding with very good results. We propose the following algorithm:

1. Divide the image Q in blocks of size $n \times n$. If M or N are not multiple of n, we delete the minimum number of rows/columns in the boundary of the image until the new size of the image satisfies the property.
2. Associate each block with an interval in the following way: the lower bound of the interval is given by the minimum of the intensities in the block and the upper bound by the maximum.
3. Choose the α parameter in the operator D_α.
4. Associate each interval with the number obtained after applying the operator D_α.

Example: Let Q be a matrix of dimension 6×6 and let $n = 3$

$$\begin{pmatrix} q_{1,1} & q_{1,2} & q_{1,3} & q_{1,4} & q_{1,5} & q_{1,6} \\ q_{2,1} & q_{2,2} & q_{2,3} & q_{2,4} & q_{2,5} & q_{2,6} \\ q_{3,1} & q_{3,2} & q_{3,3} & q_{3,4} & q_{3,5} & q_{3,6} \\ q_{4,1} & q_{4,2} & q_{4,3} & q_{4,4} & q_{4,5} & q_{4,6} \\ q_{5,1} & q_{5,2} & q_{5,3} & q_{5,4} & q_{5,5} & q_{5,6} \\ q_{6,1} & q_{6,2} & q_{6,3} & q_{6,4} & q_{6,5} & q_{6,6} \end{pmatrix}$$

Then, the interval-valued fuzzy set associated with Q is formed by 4 elements:

$$\left(\begin{bmatrix} \bigwedge_{\substack{i=1,2,3 \\ j=1,2,3}} q_{i,j} & \bigvee_{\substack{i=1,2,3 \\ j=1,2,3}} q_{i,j} \\ \bigwedge_{\substack{i=4,5,6 \\ j=1,2,3}} q_{i,j} & \bigvee_{\substack{i=4,5,6 \\ j=1,2,3}} q_{i,j} \end{bmatrix} \begin{bmatrix} \bigwedge_{\substack{i=1,2,3 \\ j=4,5,6}} q_{i,j} & \bigvee_{\substack{i=1,2,3 \\ j=4,5,6}} q_{i,j} \\ \bigwedge_{\substack{i=4,5,6 \\ j=4,5,6}} q_{i,j} & \bigvee_{\substack{i=4,5,6 \\ j=4,5,6}} q_{i,j} \end{bmatrix} \right)$$

Remark: Notice that the symbols $\wedge$ and $\vee$ stand for minimum and maximum respectively.

Once the interval-valued fuzzy set associated with the image has been obtained, we build a fuzzy set. Applying the operator D_α to each interval we get the reduced image of size 2×2.

ISBN: 978-989-95079-6-8



Figure 1: Original Image Lena



Figure 2: Original image Cameraman

6 Experimental results

In this section we use the algorithm proposed in Section 5. Using the relation between the operator D_α and OWA operators of dimension 2, we propose several methods of calculating the α coefficient based on families of OWA operators.

We get two different situations according to the OWA operator used. In the first one, the value of α is constant for the whole image. This situation happens with ME-OWA operators and generalized S-OWA operators, due to their definition. In the second one, the value of α depends on the bounds of each interval, so its value varies for each block. This happens when we take BADD-OWA operators and modified BADD-OWA operators.

The tests of this section have been made with the image Lena (Figure 1) and Cameraman (Figure 2). The size of the block is $n = 3$ (submatrices 3×3). In this way, the size of the reduced image is 9 times smaller than the original.

6.1 Calculation with constant α

In this section we use ME-OWA operators and generalized S-OWA to calculate a constant value of α for the whole image. Once the value has been calculated, we apply the D_α operator to each interval in the image.

6.1.1 Reduction with ME-OWA operators

As we have seen in Section 4, the construction of ME-OWA operators is direct. For this reason, we analyze three specific cases: $\alpha = 0$, $\alpha = 0.5$ and $\alpha = 1$.

With $\alpha = 0$, we associate the lower bound of the interval to each block. With $\alpha = 0.5$, we take the mean point of the interval. Finally, with $\alpha = 1$, we associate the upper bound of the interval.

Obviously, the higher the value of α, the higher the membership degree and therefore the intensity of the image. The image is darker with $\alpha = 0$ than with $\alpha = 0.5$, which is also darker than with $\alpha = 1$.



$\alpha = 0$ $\alpha = 0.5$ $\alpha = 1$

Figure 3: Reduction using ME-OWA operators

6.1.2 Reduction with generalized S-OWA operator

To construct a generalized S-OWA operator, it is necessary, as we have seen in Definition 8, the election of two parameters $a, b \in [0, 1]$ such that $a + b \leq 1$. If we focus on the contribution of the parameters in the S-OWA operator, we can study several possibilities:

1. If $a = b$, then we have that $\alpha = 0.5$, getting the average of the bounds of the interval.
2. If $a = 1$ and $b = 0$, we get $\alpha = 1$.
3. If $a = 0$ and $b = 1$, we get $\alpha = 0$.
4. If $a + b < 1$ and $a \neq b$, when $a > b$, the upper bound of the interval is predominant over the lower bound, while when $a < b$, the importance is given to the lower.

Cases 1, 2 and 3 have been already analyzed in Section 6.1.1, so we focus in case 4. For this, we start taking $a = 0.5, b = 0.25$. As $a > b$, the importance is given to the upper bound of the interval, getting a value of $\alpha = 0.625$. If we increase the value of a, then α tends to 1. For example, if we take $a = 0.9, b = 0$, then $\alpha = 0.95$. With this value of α we get a reduced image very similar to the images studied in the last column of Figure 3. On the other side, if we take $a = 0.25, b = 0.5$, we get a value of $\alpha = 0.375$. If we increase the value of b, $\alpha \to 0$. We can see this if we take $a = 0, b = 0.9$ ($\alpha = 0.05$). Notice that as α increase, the image becomes lighter.

6.2 Calculation with variable α

In this section, the value of α is calculated by means of the bounds of the interval. This means that the value of α is variable. For this reason we use BADD-OWA and modified BADD-OWA operators.

ISBN: 978-989-95079-6-8



a=0,b=0.9 $\alpha = 0.05$ | a=0.25,b=0.5 $\alpha = 0.375$ | a=0.5,b=0.25 $\alpha = 0.625$ | a=0.9,b=0 $\alpha = 0.95$

Figure 4: Reduction using generalized S-OWA operators

6.2.1 Reduction with BADD-OWA operators

As we see in Definition 9, to construct a BADD-OWA operator it is necessary to select a value of $\beta \geq 0$. We know that if we take $\beta = 0$, then we get $\alpha = 0.5$, already studied in Section 6.1.1. In other case, if we take $\beta = 1$, the value of α is calculated as follows:

$$\alpha = \frac{\overline{x}}{\overline{x} + \underline{x}}$$

being $[\underline{x}, \overline{x}] \in L([0,1])$ the interval representing each block. We also know that $\alpha \geq 0.5$, and if we increase the value of β, $\alpha \to 1$. That is, the result tends to the upper bound of the interval. For this reason, with a high value of β, we get reduced images similar as the images analyzed in third column of Figure 3.



$\beta = 1$ | $\beta = 10$ | $\beta = 20$ | $\beta = 100$

Figure 5: Reduction using BADD-OWA operators

6.3 Reduction with modified BADD-OWA operators

If we use modified BADD-OWA operators, bearing in mind items 1 and 2 of Proposition 7, we can take two different expressions for α. As we have analyzed in BADD-OWA operators, the value of α depends on the interval and on the β parameter. As in Section 6.2.1, if $\beta = 0$, then $\alpha = 0.5$.

Under conditions of item 1 of Proposition 7 and taking $\beta = 1$, the value of α is as follows:

$$\alpha = \frac{\underline{x}}{\underline{x} + \overline{x}}$$

being $[\underline{x}, \overline{x}] \in L([0,1])$ the interval representing each block. In this case, $\alpha \leq 0.5$ and as β increases, $\alpha \to 0$ when $\beta \to \infty$ (notice that the image becomes darker as β increases and α decreases).



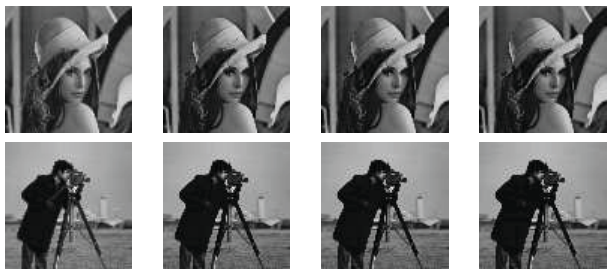
$\beta = 1$ | $\beta = 10$ | $\beta = 20$ | $\beta = 100$

Figure 6: Reduction using modified BADD-OWA operators (item 1)

If we base on item 2 of Proposition 7, for $\beta = 1$, α is as follows:

$$\alpha = \frac{1 - \overline{x}}{(1 - \overline{x}) + (1 - \underline{x})}$$

being $[\underline{x}, \overline{x}] \in L([0,1])$ the interval representing each block. In this case, the value of α also decreases when β increases.



$\beta = 1$ | $\beta = 10$ | $\beta = 20$ | $\beta = 100$

Figure 7: Reduction using modified BADD-OWA operators (item 2)

7 Conclusions and future research

In this paper, we have proved that we can obtain OWA operators of dimension 2 from Atanassov's operators. This fact has allowed us to use several families of OWA operators to get D_α operators for image reduction.

Some future lines of research can be:

1. Compare the different reductions obtained according to the value of α.
2. Compare our results with other reduction algorithms.
3. Consider different methods of reconstruction of the original image from the reduced one. Analyze which of the reduced images leads to the best reconstructed image.

Acknowledgment - This paper has been partially supported by the National Science Foundation of Spain, Reference TIN2007-65981.

ISBN: 978-989-95079-6-8

References

[1] K. Atanassov, Intuitionistic fuzzy sets, In: VIIth ITKR Session, Deposited in the Central Science and Technology Library of the Bulgarian Academy of Sciences, Sofia, Bulgaria, 1983, pp. 1684-1697.

[2] K. Atanassov, Intuitionistic fuzzy sets, Fuzzy Sets and Systems 20 (1986) 87-96.

[3] H. Bustince, E. Barrenechea, M. Pagola, Generation Of Interval-Valued Fuzzy And Atanassov's Intuitionistic Fuzzy Connectives From Fuzzy Connectives And From K(Alpha) Operators. Laws For Conjunctions And Disjunctions. Amplitude, International Journal of Intelligent Systems (2008) In Press

[4] H. Bustince, J. Montero, M. Pagola, E. Barrenechea, D. Gmez: A Survey On Interval-Valued Fuzzy Sets. In W. Pedrycz, A. Skowron, V. Kreinovichedrycz (Eds.): *Handbook of Granular Computing* John Wiley and sons, NewYork, (2008), Chapter 22.

[5] H. Bustince, J. Montero, E. Barrenechea, M. Pagola, Laws of Conjunctions and Disjunctions in Interval type 2 fuzzy sets. In: Proc. IEEE World Congress on Computational Intelligence, WCCI2008 , Hong Kong, 2008.

[6] T. Chaira, A.K. Ray, A new measure using intuitionistic fuzzy set theory and its application to edge detection, Applied Soft Computing 8(2) (2008) 919-927.

[7] C. Cornelis, G. Deschrijver, E.E. Kerre, Advances and challenges in interval-valued fuzzy logic, Fuzzy Sets and Systems 157 (2006) 622-627.

[8] G. Deschrijver,C. Cornelis, E.E. Kerre, On the representation of intuitionistic fuzzy T-norms and T-conorms, IEEE Transactions on Fuzzy Systems 12(1) (2004) 45-61.

[9] M.B. Gorzalczany, A method of inference in approximate reasoning based on interval-valued fuzzy sets, Fuzzy Sets and Systems 21 (1987) 1-17.

[10] E. Karabasssis, M. E. Spektsakis, An analysis of image interpolation, differentiation and reduction using local polynomial fits, Graphical models and image processing, 57 (3), (1995) 183-1995.

[11] H. Nobuhara, W. Pedrycz, S. Sessa, K. Hirota, A motion compression/reconstruction method based on max t-norm composite fuzzy relational equations, Information Sciences 176 (2006) 2526-2552.

[12] R. Sambuc, Function Φ-Flous, Application a l'aide au Diagnostic en Pathologie Thyroidienne, These de Doctorat en Medicine, University of Marseille (1975).

[13] H.R. Tizhoosh, Image thresholding using type-2 fuzzy sets, Pattern Recognition 38 (2005) 2363-2372.

[14] I. B. Turksen, Interval valued fuzzy sets based on normal forms, Fuzzy Sets and Systems 20(2) (1986) 191-210.

[15] I.K. Vlachos, G.D. Sergiadis, Intuitionistic fuzzy information - Applications to pattern recognition, Pattern Recognition Letters 28(2) (2007) 197-206.

[16] R.R. Yager, On ordered weighted averaging aggregation operators in multicriteria decisionmaking, IEEE Trans. Syst. Man Cybern, 18 (1988) 183-190.

[17] R.R. Yager, Families of OWA operators, Fuzzy Sets and Systems, 59 (1993) 125-148.

[18] G. Y. Yang, H.Z. Shu, C. Toumoulin, G.N. Han, L.M. Luo, Efficient Legendre moment computation for grey level images, Pattern Recognition, 39 (2006) 74-80.

ISBN: 978-989-95079-6-8

Fuzzy Transforms, Korovkin Theorems and the Durrmeyer Operator

Barnabas Bede, Imre J. Rudas
Department of Mathematics, University of Texas-Pan American,
Edinburg, Texas, USA. email: bedeb@utpa.edu.
Department of Intelligent Engineering Systems, John von Neumann Faculty of Informatics,
Budapest Tech, Budapest, Hungary. email: rudas@bmf.hu

***Abstract*— *In the present paper a Korovkin-type theorem is proposed for the approximation operators defined by the inverse F-transforms. These results allow us to choose between a variety of shapes to be used as atoms of the fuzzy partitions used within the F-transform's framework. In this way we can enlarge considerably the class of F-transforms proposed recently by I. Perfilieva. The new fuzzy partitions are shown to include, for example the Bernstein basis polynomials. The F-transform considered with the Bernstein basis polynomials as atoms of a fuzzy partition, is shown to be the classical Durrmeyer operator.*

***Keywords*—** F-transform, Korovkin Theorems, Bernstein Polynomial, Durrmeyer operator.

1 Introduction

The F-transform was proposed by I. Perfilieva in [4] and studied in several papers [5], [7], etc. As starting point for the F-transforms, one considers first a fuzzy partition [6]. In the present paper we enlarge the class of F-transforms by considering arbitrary shapes and not necessarily small support for the atoms of the fuzzy partition used in the F-transform as an approximation method. To study the possibility of such extension we exploit the classical Korovkin theorems, [1] which are particularized in the present paper for the F-transform. The Korovkin Theorems are very important results in classical Approximation Theory, stating that the convergence of a sequence of positive linear approximation operators on a finite set of test functions implies the convergence of the operator for any function. There are two versions of the Korovkin type results. A qualitative and a quantitative version. The quantitative versions provide error estimates, beyond proving the convergence results of qualitative type. We will adopt the quantitative approach. Further-on in the present paper we consider the F-transforms based on Bernstein basis polynomials. It turns out that the F-transform that uses Bernstein basis polynomials as atoms for the fuzzy partition under discussion, leads to the classical Durrmeyer operators [2]. In this way, the theory of fuzzy approximation and especially approximation by inverse F-transforms can be embedded and studied, as a refined part of the classical theory of approximation.

In the present paper a fuzzy partition is defined as a finite sequence of fuzzy sets $A_i : [a,b] \to [0,1]$, $i = 1, ..., k$, such that $\sum_{i=1}^{k} A_i(x) = 1$ for any $x \in [a,b]$. The following is a generalized version of the Fuzzy Transform (F-transform) proposed in [4], since the condition regarding the small support of the atoms of the fuzzy partitions occurring in the F-transform is released.

According to [4], the F-transform is given by

$$f_i = \frac{\int_a^b A_i(x) f(x) dx}{\int_a^b A_i(x) dx}, i = 1, ..., k \tag{1}$$

and the discrete F-transform is given by

$$f_i = \frac{\sum_{j=1}^{n} A_i(x_j) f(x_j)}{\sum_{j=1}^{n} A_i(x_j)}, i = 1, ..., k \tag{2}$$

where $x_j \in [a,b]$, $j = 1, ..., n$ are given data. The inverse discrete F-transform is

$$F(x) = \sum_{i=1}^{k} A_i(x) f_i. \tag{3}$$

We denote by $C(K)$ the space of continuous functions $f : K \to \mathbb{R}$ with K a compact metric space. Then the space $C(K)$ is a Banach space with the uniform norm

$$\|f\| = \sup_{x \in K} |f(x)|.$$

If $K = [a,b]$ then we use the notation

$$C[a,b] = \{f : [a,b] \to \mathbb{R}, f \text{ continuous}\}.$$

We will consider inverse F-transform as an approximation operator. Let $\mathcal{L}(C(K))$ denote the space of linear operators of the form $T : C(K) \to C(K), T(\alpha f + \beta g) = \alpha T(f) + \beta T(g)$, $\forall \alpha, \beta \in \mathbb{R}$ and $f, g \in C(K)$. An operator is said to be positive if $T(f(x)) \geq 0 \; \forall x \in K$ whenever $f(x) \geq 0, \forall x \in K$. In the present paper positive linear operators play a very important role.

The inverse F-transform will be regarded in the present paper as a sequence of positive linear operators. Indeed, let us consider $F_{n,k} : C[a,b] \to C[a,b]$, given by $F_{n,k}(f)(x) = F(x)$, where $F(x)$ is given by (3) and f_i given by (1) or (2). Then, it is easy to see that $F_{n,k}$ is a linear operator, i.e., $F_{n,k} \in \mathcal{L}(C[a,b])$. Moreover, if we assume $f(x) \geq 0$ $\forall x \in [a,b]$, we can see that both the continuous and discrete F-transforms $f_i, i = 1, ..., k$ satisfy $f_i \geq 0$, and as a consequence the inverse F-transform $F_{n,k}(f)(x) = F(x) \geq 0$ and as a conclusion $F_{n,k}$ is a positive linear operator.

In order to obtain an error estimate and the uniform convergency of the F-transforms that we intend to study, seen as approximation operators via a particularization of the Korovkin Theory related to positive linear operators. For this aim we need the modulus of continuity. Let $f : [a,b] \to \mathbb{R}$ be a continuous function. Then $\omega(f, \cdot) : [0, \infty) \to [0, \infty)$, defined by

$$\omega(f, \delta) = \bigvee \{|f(x) - f(y)|; x, y \in [a,b], \; |x - y| \leq \delta\}$$

ISBN: 978-989-95079-6-8

is called the first order modulus of smoothness (modulus of continuity) of f.

2 Korovkin type theorem for the F-transform

Let us recall Korovkin's Theorem in what follows. We can see that it is directly applicable for the F-transform, since the F-transform is a positive linear operator, however, we will deduce later particularizations of the Korovkin results for the F-transform to better understand the F-transform method and to be able to provide error estimates for them.

Theorem 1 *(Korovkin, [3]) Let $L_n \in \mathcal{L}(C[a,b])$, $n = 1,2,...$ be a sequence of positive linear operators. Then there exists a finite set of test functions S, such that $\lim_{n\to\infty} L_n(s) = s$ for any $s \in S$ implies $\lim_{n\to\infty} L_n(f) = f$, for any $f \in C[a,b]$. The finite set S is called a Korovkin set.*

If we consider a sequence of positive linear operators on $C[0,1]$, then a Korovkin set (it is not unique) consists of $e_0, e_1, e_2,$ where $e_i(x) = x^i$, $i = 0,1,2$. If we consider $\mathcal{L}(C[-\pi,\pi])$ a Korovkin set consists of e.g., $e_0(x) = 1$, $e_1(x) = \sin x$ and $e_2(x) = \cos x$.

The adaptation of Korovkin Theorem for the F-transform is as follows.

Theorem 2 *(Korovkin Theorem for the discrete inverse F-transform) Let $F_{n,k} \in \mathcal{L}(C[a,b])$, $n,k = 1,2,...$ be a sequence of inverse F-transforms. The following general error estimate holds true for any $\delta > 0$:*

$$\|F_{n,k}(f) - f\| \leq \left(1 + \frac{1}{\delta}\sqrt{e_2 - 2e_1F_{n,k}(e_1) + F_{n,k}(e_2)}\right)\omega(f,\delta),$$

where $e_1(x) = x, e_2(x) = x^2$.

Proof. The inverse discrete F-transform can be expressed as

$$F_{n,k}(f)(x) = \sum_{i=1}^{k} A_i(x)\frac{\sum_{j=1}^{n} A_i(x_j)f(x_j)}{\sum_{j=1}^{n} A_i(x_j)}.$$

Let us observe that for the inverse F-transform we have $F_{n,k}(e_0) = e_0$, where $e_0(x) = 1\ \forall x \in [a,b]$, so $f(x)$ can be written as

$$f(x) = \sum_{i=1}^{k} A_i(x)\frac{\sum_{j=1}^{n} A_i(x_j)f(x)}{\sum_{j=1}^{n} A_i(x_j)}.$$

Then we get

$$|F(x) - f(x)| \leq \sum_{i=1}^{k} A_i(x)\frac{\sum_{j=1}^{n} A_i(x_j)\,|f(x) - f(x_j)|}{\sum_{j=1}^{n} A_i(x_j)}.$$

Further, using the properties of the modulus of continuity, by standard reasoning we have

$$|F(x) - f(x)| \leq \left(1 + \frac{1}{\delta}\sum_{i=1}^{k} A_i(x)\frac{\sum_{j=1}^{n} A_i(x_j)|x - x_j|}{\sum_{j=1}^{n} A_i(x_j)}\right)\omega(f,\delta),$$

for any fixed $\delta > 0$. The error is controlled by the ratio

$$R_{n,k}(x) = \sum_{i=1}^{k} A_i(x)\frac{\sum_{j=1}^{n} A_i(x_j)|x - x_j|}{\sum_{j=1}^{n} A_i(x_j)}.$$

Using Cauchy-Schwarz inequality we have

$$\sum_{j=1}^{n} A_i(x_j)|x - x_i| \leq \left(\sum_{j=1}^{n} A_i(x_j)|x - x_j|^2\right)^{\frac{1}{2}} \cdot \left(\sum_{j=1}^{n} A_i(x_j)\right)^{\frac{1}{2}},$$

and we obtain

$$R_{n,k}(x) \leq \sum_{i=1}^{k} A_i(x)\frac{\left(\sum_{j=1}^{n} A_i(x_j)|x - x_j|^2\right)^{\frac{1}{2}}}{\left(\sum_{j=1}^{n} A_i(x_j)\right)^{\frac{1}{2}}} = \sum_{i=1}^{k} \sqrt{A_i(x)}\frac{\left(A_i(x)\sum_{j=1}^{n} A_i(x_j)|x - x_j|^2\right)^{\frac{1}{2}}}{\left(\sum_{j=1}^{n} A_i(x_j)\right)^{\frac{1}{2}}}.$$

Using Cauchy-Schwarz inequality again we obtain

$$\sum_{i=1}^{k} \sqrt{A_i(x)}\frac{\left(A_i(x)\sum_{j=1}^{n} A_i(x_j)|x - x_j|^2\right)^{\frac{1}{2}}}{\left(\sum_{j=1}^{n} A_i(x_j)\right)^{\frac{1}{2}}} \leq \left(\sum_{i=1}^{k} A_i(x)\right)^{\frac{1}{2}}\left(\sum_{i=1}^{k}\frac{A_i(x)\sum_{j=1}^{n} A_i(x_j)|x - x_j|^2}{\sum_{j=1}^{n} A_i(x_j)}\right)^{\frac{1}{2}}.$$

Taking into account that $A_i(x)\ is$ a fuzzy partition we have

$$R_{n,k}(x) \leq \left(\sum_{i=1}^{k} A_i(x)\frac{\sum_{j=1}^{n} A_i(x_j)(x^2 - 2xx_j + x_j^2)}{\sum_{j=1}^{n} A_i(x_j)}\right)^{\frac{1}{2}} = \left(x^2 - 2xF_{n,k}(e_1)(x) + F_{n,k}(e_2)(x)\right)^{\frac{1}{2}}$$

which completes the proof. ■

Corollary 3 *Let $F_{n,k} \in \mathcal{L}(C[a,b])$, $n,k = 1,2,...$ be a sequence of inverse F-transforms. If $\lim_{n,k\to\infty} F_{n,k}(e_i) = e_i$, $i = 1,2$, then $\lim_{n\to\infty} L_n(f) = f$, for any $f \in C[a,b]$.*

For the continuous case we have a similar result.

Theorem 4 *(Korovkin Theorem for the inverse F-transform) Let $F_k \in \mathcal{L}(C[a,b])$, $k = 1,2,...$ be a sequence of inverse F-transforms. The following general error estimate holds true for any $\delta > 0$:*

$$\|F_k(f) - f\| \leq \left(1 + \frac{1}{\delta}\sqrt{e_2 - 2e_1F_k(e_1) + F_k(e_2)}\right)\omega(f,\delta),$$

where $e_1(x) = x, e_2(x) = x^2$.

Corollary 5 *Let $F_k \in \mathcal{L}(C[a,b])$, $k = 1,2,...$ be a sequence of continuous inverse F-transforms. If $\lim_{k\to\infty} F_k(e_i) = e_i$, $i = 1,2$, then $\lim_{n\to\infty} L_n(f) = f$, for any $f \in C[a,b]$.*

ISBN: 978-989-95079-6-8

We observe that the inverse F-transform can be expressed as

$$F_k(f)(x) = \sum_{i=1}^{k} A_i(x) \frac{\int_a^b A_i(x) f(x) dx}{\int_a^b A_i(x) dx}$$

so, similar reasoning to the previous Theorem can be followed.

The Korovkin type Theorem above shows that if we are able to control $\sqrt{e_2 - 2e_1 F_{n,k}(e_1) + F_{n,k}(e_2)}$ for the discrete case, or $\sqrt{e_2 - 2e_1 F_k(e_1) + F_k(e_2)}$ for the continuous case, for a given sequence of fuzzy partitions A_i, $i = 1, ..., k$, then we are able to control the error for any continuous function.

3 Example

Using the Korovkin-type result shown in the previous section we consider the F-transforms, with Bernstein basis polynomials used as a fuzzy partition this time. The proposed construction leads to the well-known Durrmeyer operator. Let us recall that the Bernstein basis polynomials are

$$p_{k,i}(x) = \binom{k}{i} x^i (1-x)^{k-i},$$

$k = 1, 2, ...$ $i = 0, 1, ..., k$ $x \in [0, 1]$.

The inverse F-transform is given by

$$F_k(f)(x) = \sum_{i=0}^{k} p_{k,i}(x) \frac{\int_0^1 p_{k,i}(x) f(x) dx}{\int_0^1 p_{k,i}(x) dx},$$

that is the classical Durrmeyer operator [2]. Let us calculate $F_k(e_1)$ and $F_k(e_2)$ in this case. It is well known that $p'_{k,.i}(x) = k(p_{k-1,i-1}(x) - p_{k-1,i}(x))$. Integrating we obtain $\int_0^1 p_{k-1,i-1}(x)dx = \int_0^1 p_{k-1,i}(x)dx$, i.e., for fixed k integrals of all Bernstein basis polynomials are the same. Since $\sum_{i=0}^{k} p_{n,i}(x) = 1$ we get $\int_0^1 p_{k,i}(x)dx = \frac{1}{k+1}$. By direct calculation

$$\int_0^1 p_{k,i}(x) x dx = \int_0^1 \binom{k}{i} x^{i+1} (1-x)^{k-i} dx$$

$$= \frac{i+1}{k+1} \int_0^1 p_{k+1,i+1} dx = \frac{i+1}{(k+1)(k+2)}.$$

Also,

$$\int_0^1 p_{k,i}(x) x^2 dx = \int_0^1 \binom{k}{i} x^{i+2} (1-x)^{k-i} dx$$

$$= \frac{i+1}{k+1} \frac{i+2}{k+2} \int_0^1 p_{k+2,i+2} dx = \frac{(i+1)(i+2)}{(k+1)(k+2)(k+3)}.$$

Then we obtain

$$f_i(e_1) = \frac{\int_0^1 p_{k,i}(x) x dx}{\int_0^1 p_{k,i}(x) dx}$$

$$= \frac{i+1}{k+2} \sim \mathcal{O}\left(\frac{i}{k}\right)$$

and

$$f_i(e_2) = \frac{\int_0^1 p_{k,i}(x) x^2 dx}{\int_0^1 p_{k,i}(x) dx}$$

$$= \frac{(i+1)(i+2)}{(k+2)(k+3)} \sim \mathcal{O}\left(\frac{i^2}{k^2}\right).$$

$$F_k(e_1)(x) \sim \sum_{i=1}^{k} p_{k,i}(x) \frac{i}{k} = B_k(e_1)(x) = x,$$

where B_k denotes the usual Bernstein operator

$$B_k(f)(x) = \sum_{i=1}^{k} p_{k,i}(x) f\left(\frac{i}{k}\right)$$

and it is well known that $B_k(e_1)(x) = x$. We also have

$$F_k(e_2)(x) \sim \sum_{i=1}^{k} p_{k,i}(x) \left(\frac{i}{k}\right)^2$$

$$= B_k(e_2)(x) = x^2 + \frac{x(1-x)}{k},$$

and finally we obtain: $\sqrt{e_2 - 2e_1 F_k(e_1) + F_k(e_2)} = \sqrt{\frac{x(1-x)}{k}}$ and by the above Korovkin type results we have

$$\|F_k(f) - f\| \leq C\omega\left(f, \sqrt{\frac{x(1-x)}{k}}\right).$$

As a conclusion we observe that the F-transform is a generalization of the Durrmeyer operators. For the discrete F-transform with Bernstein basis polynomials a similar result can be proven.

4 Conclusion

Korovkin-type theorems for the discrete and continuous inverse F-transforms regarded as positive linear approximation operators were provided. As an application we have studied the F-transform with Bernstein basis polynomials as atoms of the underlying fuzzy partition. It turns out that the F-transform with Bernstein basis polynomials is the classical Durrmeyer operator. Surely one can imagine different fuzzy partitions as well, so the F-transforms can be seen as generalizations of the Durrmeyer operator.

References

[1] R.A. DeVore, G.G. Lorentz, Constructive Approximation, Polynomials and Splines Approximation, Springer-Verlag, Berlin, Heidelberg, 1993.

[2] J.L. Durrmeyer, Une formule d'inversion de la transformée de Laplace: Applications à la théorie des moments. Thèse de 3e cycle (1967) Paris.

[3] P.P. Korovkin, Linear operators and approximation theory, Hindustan Pub. Corp. 1960.

[4] I. Perfilieva, Fuzzy Transforms: theory and applications, Fuzzy Sets and Systems, 157(2006) 993-1023.

[5] I. Perfilieva, Fuzzy Transforms, in Transactions on Rough Sets II (J.F. Peters et al. Eds.) LNCS 3135 (2004), 63-81.

[6] H. Ruspini, E. (1969) A new approach to clustering. Inform. and Control, Vol. 15, 22–32.

[7] M. Stepnicka, O. Polakovi, A neural network approach to the fuzzy transform, Fuzzy Sets and Systems, to appear.

ISBN: 978-989-95079-6-8

Fuzzy Logic Applications in Wireless Communications

Maria Erman[1] Abbas Mohammed[2] Elisabeth Rakus-Andersson[3]

[1,2]Department of Signal Processing, Blekinge Institute of Technology
S-37225 Ronneby, Sweden
[3]Department of Mathematics and Science, Blekinge Institute of Technology
S-37179 Karlskrona, Sweden
Email: maria.erman@bth.se, abbas.mohammed@bth.se, elisabeth.andersson@bth.se

Abstract— *A survey of fuzzy logic applications and principles in wireless communications is presented, with the aim of highlighting successful usage of fuzzy logic techniques in applied telecommunications and signal processing. To the best of our knowledge, this is the first such study of its kind. This paper will focus firstly on discerning prevalent fuzzy logic or fuzzy-hybrid approaches in the areas of channel estimation, channel equalization and decoding, and secondly outlining what conditions and situations for which fuzzy logic techniques are most suited for these approaches. Furthermore, after insights gained from isolating fuzzy logic techniques applied to real problems, this paper proposes areas for further research targeted to practice-oriented researchers.*

Keywords— Channel Equalization, Channel Estimation, Decoding, Fuzzy-hybrid systems, Fuzzy Logic, Wireless Communications

1 Introduction

Fuzzy Logic has been successfully applied in various areas pertaining to wireless communication systems. As fuzzy logic is used to model systems and situations, taking into consideration uncertainty and ambiguity, it can be an efficient tool to be utilized in problems for which knowledge of all factors is insufficient or impossible to obtain. Methods furnished with fuzzy logic have been shown to be useful in difficult conditions with respect to non-linear and time-variant systems. Additionally, the often mentioned advantages of using fuzzy logic in practical applications is to reduce complexity as well as to add robustness to the system under study.

Fuzzy logic and, more specifically, fuzzy control traditionally incorporates human expert knowledge into a rule-based framework. It may, however, be further expanded with learning algorithms to derive the fuzzy control parameters from sample data. These parameters may be obtained by combining fuzzy logic with related soft computing disciplines such as, e.g., neural networks, evolutionary computation techniques etc. On the other hand, a method developed by Wang and Mendel [1] derives the fuzzy rule base by using a combination of human experience and numerical data.

Wireless communications is a rapidly evolving industry, constantly challenging researchers for new techniques in order to meet the demands of ever higher performance and efficiency. The most obvious products of wireless communications, e.g., the worldwide adoption of the mobile telephone, wireless local area networks etc., exert a strong influence on many people's lives today.

In a wireless communication system, the *channel* is the medium by which information-bearing signals are transferred from a transmitter to a receiver. The characteristics of the channel are generally unknown, and, barring any distortion imposed by the channel, i.e., in ideal conditions, the transmitted data will be received without any errors.

In practice, however, it is unavoidable for the channel to be affected by distortion, hence degrading the performance of the receiver, severely limiting the throughput of the system. The wireless channel poses tough challenges for achieving reliable and fast transfers. While interference typically is not a major concern in wired transmission, i.e., in predicting the behavior of the signal in the transmission channel, it poses a great challenge in wireless transmission.

When there is no line-of-sight between the transmitter and the receiver, distortions to the signal in the form of effects such as scattering and reflections, etc. will follow, all of them resulting in a phenomenon called multipath propagation. Due to multipath propagation, the receiver encounters many signal paths from the transmitter, where each of these paths is delayed by an arbitrary amount and attenuated by various factors. As a result there will be a superposition of the different copies of the signal being in different phases, hence causing an amplification or attenuation of the signal power – also referred to as fading. Multipath propagation will cause previously sent data bits to smear into current data bits, referred to as intersymbol interference. The aim of the receiver in the communication system is to overcome the disturbances of the channel, intersymbol interference and noise, and correctly decode the data having been transmitted.

2 Contemporary Uses of Fuzzy Logic in Wireless Communications

In this section we will focus on fuzzy logic applications in channel estimation, channel equalization and decoding. The purpose of channel estimation is to accurately describe the channel and track its variations, and with the aid of channel equalization and decoding recover the original transmitted data. In the case of time-varying channels, adaptive techniques have to be employed and it is in this area that fuzzy techniques and/or neural networks find their main uses.

2.1 Channel Estimation

In [2] channel estimation is performed by tracking the channel coefficients, applying a fuzzy tracking method in a multipath fading Code Division Multiple Access (CDMA) [3] channel. CDMA is a spread-spectrum technology that makes it possible for transmitters to share the same frequency range. The

ISBN: 978-989-95079-6-8

fuzzy tracking method is based on Kosko's fuzzy associative memory models [4]. The fuzzy associative memory models combines fuzzy logic and a single-layer feed-forward neural network that saves the fuzzy logic rule-base in matrix form. The tracking used in [2] is iterative with the estimated symbol being used in the prediction of the coefficient. There are two inputs to the fuzzy tracker: the difference between the measured and predicted coefficients, and the change of difference between current and previous differences between the measured and predicted coefficients. The output of the fuzzy tracker yields a correction term for the next coefficient.

The motivation in [2] for using fuzzy channel estimation is due to fuzzy estimation not needing exact process models. Comparisons are made between the fuzzy tracker and a non-fuzzy tracker, a.k.a. the alpha tracker, where it is shown that the fuzzy tracker performs better under noisier multipath conditions.

Channel estimation using a fuzzy approach has also been performed with a multi-carrier modulation technique called Orthogonal Frequency Division Multiplexing (OFDM) [5]. In OFDM, multiple orthogonal subcarriers are used for the same channel. The data stream is divided into lower bit-rate data streams, each modulating a separate subcarrier.

There are two different methods for estimating the channel parameters at each subcarrier: blind channel estimation techniques and pilot assisted channel estimation, with the pilot being a reference signal used by the transmitter and the receiver. The blind channel estimation techniques do not use pilot samples and are thus more spectrally efficient, but at the cost of higher computational complexity and slower convergence rate. Pilot assisted channel estimation has typically been based either on the Minimum Mean-Square-Error (MMSE), the LS (Least Square) or the LMS (Least Mean-Square) algorithms [6], with the MMSE algorithm being more robust and performing better in time-varying channels.

In [7] a Takagi-Sugeno-Kang (TSK) [8] model is used for the pilot assisted channel estimation in an OFDM system. The TSK method is similar to that of the traditional Mamdani method – the difference being that the output of the TSK model is a linear function of the input variables instead of fuzzy sets. Pilot symbols are used in order to train the TSK fuzzy model as well updating the fuzzy model in order to track the channel. A Gaussian membership function is chosen, as it has been shown that a fuzzy system with bell-shaped Gaussians can approximate any continuous functions on compact sets to any degree of accuracy [9]. The TSK learning algorithm consists of defining the center and width of the rules dependent on the number of rules chosen, after which an adjustable parameter of the TSK model is trained from the first snapshot of the pilot subcarriers. Finally the channel transfer function is estimated and the adjustable parameter updated to track the channel. Simulation results show that the proposed TSK channel estimation model performs closely to the ideal MMSE, but with lower computational complexity.

The TSK fuzzy modeling technique used in [7] has been updated for a multiple-input multiple-output (MIMO) OFDM system in [10] with two transmit and two receive antennas. MIMO [11] is a way to increase system throughput without the need for higher transmit power or bandwidth, and has hence become a highly popular research topic. MIMO achieves increased throughput by using multiple antennas at the transmitter and the receiver. It is shown in [10] with computer simulations that the Word Error Rate (WER) is close to the MMSE method with lower computational complexity.

In [12] an MMSE linear receiver was proposed in which a fuzzy inference system was inserted into the LMS algorithm. The motivation to use fuzzy logic was for convergence and stability reasons. The LMS algorithm was modified with a fuzzy logic controlled adaptive step size and partial update. This modified algorithm was then used in simulations of noise cancellation in a space-time joint direct-sequence (DS) CDMA system in a dynamic fading multipath channel. It was shown that the performance of the authors' modified LMS algorithm was superior to that of the LMS algorithm.

The proposed algorithm in [12] was further used in [13] for channel estimation and tracking in OFDM systems for a time-variant channel. Simulation results indicated that the modified LMS algorithm had lower steady-state Mean-Square Error (MSE) and faster convergence speed compared to the ordinary LMS algorithm.

In [14], an adaptive neuro-fuzzy inference system (ANFIS) was evaluated for channel estimation in OFDM systems. The ANFIS uses a hybrid learning algorithm based on the LS and the gradient descent methods in order to train the parameters of the membership functions of a TSK fuzzy inference system. Clustering is used in order to group data and from this generate the TSK fuzzy rule-base. From the results of the computer simulations performed in [14] it can be seen that the ANFIS performs very closely to that of the MMSE algorithm albeit with less computational complexity.

2.2 Channel Equalization

Channel equalization is the process of removing the degradation caused by the channel with the aim of reconstructing the transmitted data. The wireless channel is time-variant, and in this kind of channel, non-linear distortion is usually encountered. Due to linear equalization not performing very well in these channels, efficient equalization should be both adaptive and non-linear.

The first results that seem to appear in the literature with regards to fuzzy logic applications in channel equalization are [15, 16]. Here fuzzy adaptive filters based on both the LMS and the Recursive Least Squares (RLS) algorithms are constructed and applied to channel equalization. Initially the fuzzy sets are defined over the filter input space, after which linguistic information from human experts and numerical data are combined and incorporated into the filter. The algorithms are then used to update the free parameters. The objective of using fuzzy adaptive filters is to improve the adaptation speed of the algorithms with the extra help of linguistic inference rules. The results indicate that the bit error rates of the fuzzy equalizer is close to that of the optimal equalizer.

Whereas Wang and Mendel in [15, 16] assumed a fixed delay, in [17] the delay and the membership functions are derived from training data. Two membership functions and two rules are defined that correspond to the input of the channel, which is binary, i.e. it takes on two different values. Together they form an output grounded on the conclusion of a fuzzy rule, from which an independent decision results. Test symbols are transmitted and the correlation between the de-

ISBN: 978-989-95079-6-8

sired outputs and the received signals are determined and form weights assigned to fuzzy outputs. A weighted sum of all the rules is used in the fuzzy inference, after which the defuzzified value is fed through a thresholding device for the final decision of the equalizer. The results from the simulations performed in [17] show that the fuzzy logic equalizer outperforms the LMS in non-linear channels as well as a neural network equalizer using the backpropagation algorithm. The fuzzy logic equalizer also needs fewer training samples compared to the LMS in linear channels for the same error performance.

The Wang-Mendel RLS Fuzzy Adaptive Filter [16] is extended in [18] to a complex fuzzy filter that can handle complex channel models and signals. In [19] human expert knowledge and heuristic reasoning are replaced altogether by a Multilayer Perceptron (MLP) preprocessor unit. The MLP unit consists of a 3-layer network, the training of which provides information to the fuzzy logic system. The LMS algorithm, less computationally complex than the RLS algorithm, is then used to update the free parameter of the system.

In [20] a Bayesian equalization architecture has been developed by using a fuzzy adaptive filter construction as in [15]. The adaptive equalization is visualized as a classification problem in which an observation vector is mapped to signal constellations. In contrast to a channel equalizer such as the maximum likelihood sequence estimation (MLSE) there is no need to include a channel estimator, thus making the equalization process less computationally complex. The derived fuzzy filter function in [20] using fuzzy basis functions [21], product inference, a center of gravity (COG) defuzzifier, and Gaussian membership functions, is able to properly represent the Bayesian decision solution. The performance of the fuzzy equalizer is close to the Bayesian, with the advantage of reduced computational complexity.

In [22] a further development of the fuzzy adaptive filters is presented: the type-2 fuzzy adaptive filter, which based on an unnormalized type-2 TSK fuzzy logic system using a training sequence. This is used to implement the Bayesian equalizer with a decision feedback structure, reducing the complexity of the equalizer compared to that of the transversal equalizer (TE). In contrast to a traditional transversal equalizer or filter, the decision feedback equalizer uses previous detector decisions to cancel intersymbol interference. The type-2 fuzzy sets [23] is an extension of ordinary fuzzy sets in that the membership grades are fuzzy as well. It is shown that an unnormalized output type-1 TSK fuzzy logic system is able to implement a Bayesian equalizer for a time-invariant channel, albeit being model free and not based on a Gaussian probability model. This is further developed into a more generalized form with the type-2 fuzzy adaptive filter to accommodate a time-varying channel.

To a lesser extent, work has also been conducted on blind methods for channel equalization. Blind methods are distinguished by only using information contained in the received signal, thus making both channel estimation and training data unnecessary, with the advantage of higher spectral efficiency. However, this also means that they are strongly dependent on the obtained statistical data.

In [24] the fuzzy-C-means (FCM) algorithm is used to perform joint equalization and demodulation of a signal modulated with the Quadrature Amplitude Modulation (QAM) scheme. The receiver mapping the signal onto a set of symbols can be reduced to a classification problem, rendering a clustering analysis useful. The aim of clustering analysis is to classify objects into groups or classes (clusters) with the objects in the same group having similarities. The FCM algorithm is an unsupervised algorithm, i.e., no external information outside the data itself is needed on which the algorithm operates. The membership functions are used as a measure of what degree the data is connected to the clusters, which in this particular application depends on the amplitude and the phase distances between the received symbols, i.e. the signal constellation points. The results, for a test environment with a static channel, indicate that the algorithm converges very quickly, is robust and has lower computational complexity than that of conventional MLSE receivers. One of the problems with the FCM algorithm however is that it forces points seriously degraded by noise to belong to one or more clusters with some degree, instead of giving it low or no membership in any cluster.

The FCM algorithm in [24] is improved on in [25] by introducing what is referred to as a fuzzy possibilistic C-means (FPCM) assisted blind channel equalization scheme for time-varying channels. The FPCM algorithm solves the problem with the FCM algorithm above by making it less sensitive toward the highly noisy symbol samples. The scheme also allows the receiver to take into account cluster center information in previous data, thereby improving the accuracy of the cluster centers with more data samples. However, it is also suggested that a forgetting factor might be taken into consideration to reduce the significance of cluster centers in previous data in a time-varying channel. Due to a rather large coherence time in high speed wireless transmission systems, it is stated in [25], that the accumulation of cluster centers is feasible even under time-variant conditions. The FPCM outperforms the FCM algorithm due to the former's capability of rejecting the interference of data seriously degraded by noise. It is also shown that the performance of both algorithms depend on the amount of data involved, as can be expected.

In [26] and [27] a blind equalization algorithm based on a fuzzy neural network is outlined. Equalization is performed with a combination of channel estimation and a fuzzy neural network classifier. The algorithm first blindly estimates the channel by using the fourth cumulants of the received sequences [28]. Afterward, an approximate deconvolution is carried out. The output from the deconvolution is then fed into a fuzzy neural network classifier. Simulations undertaken in time-invariant channels with 64-QAM, indicate that the convergence speed as well as BER are improved compared to that of a feedforward neural network blind equalization algorithm.

Similarly, in [29] an improvement of a feedforward neural network blind equalization algorithm is proposed by using a fuzzy neural network consisting of an input layer, a fuzzification layer, a rule layer, a normalization layer and a defuzzification layer with the aim of improving the convergence rate. Simulation results for 16-QAM shows that the fuzzy neural network has faster convergence speed and lower BER compared to that of the feedforward neural network blind algorithm.

ISBN: 978-989-95079-6-8

2.3 Decoding and Equalization

In contrast to conventional communication systems where encoding/decoding and channel equalization are performed separately, turbo equalization schemes combine the two mechanisms. This combination is carried out by iterating the equalizer and the channel decoder on the same set of received data. Since turbo equalizers have been implemented with the Bayesian algorithm, and it was shown that a TSK fuzzy logic system is able to implement a Bayesian model in [22], a turbo-fuzzy equalization approach should be feasible. Such an approach was introduced in [30], where a turbo equalizer using fuzzy filters is proposed. The fuzzy turbo equalizer is introduced with the motivation that fuzzy filters could deal with uncertainty characterized by impulse noise, and also has lower computational complexity compared to the Bayesian equalizer. To adapt the parameters of the fuzzy equalizer, the back-propagation algorithm is used.

A critique against the turbo-fuzzy equalizer in [30] can be found in [31] where it is stated that the turbo-fuzzy equalizer in [30] is unable to use the *a priori* information provided by the decoder, hence not having an iterative extrinsic information exchange between the fuzzy system and the decoder. This is improved on in [31] and [32] where fuzzy turbo equalization schemes with low complexity are proposed.

In [32] a turbo equalization scheme, based on the radial basis functions (RBF), is proposed by using an extended FCM algorithm. An emphasis is made on the low computational complexity this scheme provides compared to a turbo equalization scheme based on the Jacobian RBF in the context of binary phase-shift keying (BPSK) modulation in a Rayleigh-fading channel. Simulation results show that the scheme proposed performs closely to that of the Jacobian RBF based turbo equalization scheme but with a significant reduction in computational complexity.

In [31] the Jacobian RBF turbo equalization scheme is modified by using the same Bayesian equalization architecture based on a fuzzy adaptive filter structure as introduced in [20]. The simulations are performed for BPSK and QAM in a Rayleigh fading channel, and indicate that the proposed fuzzy adaptive filter TEQ scheme considerably reduces the computational complexity with only a slight degradation in performance compared to the Jacobian RBF turbo equalization scheme, providing a trade-off for highly low complexity-oriented circuit implementation.

3 Conclusions

In this paper we have traced the research being conducted over the last two decades, leading up to current research, in which the usage of fuzzy logic in wireless communications has yielded successful results. To the best of our knowledge, this is the first such study of its kind. The three areas in wireless communications focused on in this paper have been: channel estimation, channel equalization and decoding.

In channel estimation, the fuzzy based methods to have been applied have ranged from fuzzy tracking based on Kosko's fuzzy associative memory models and the TSK model, as well as fuzzy logic used in combination with adaptive algorithms such as the LMS and RMS algorithms or a neuro-fuzzy inference system. The neural network and adaptive algorithms are commonly used in order to train the parameters of membership functions in a fuzzy inference system.

In channel equalization the research in fuzzy adaptive filters, both type-1 and more recently type-2 TSK fuzzy logic systems, from Wang and Mendel have been highly influential. These fuzzy adaptive filters are able to use input from both human experts and/or training data. Using as a foundation the adaptive fuzzy filters, a Bayesian architecture has been developed which incorporates fuzzy basis functions and Gaussian membership functions, being able to properly represent the Bayesian decision solution. Another category in channel equalization is the blind methods which uses variants of the fuzzy-C-means algorithm or a fuzzy neural network. Turbo equalization has either been based on the Bayesian equalization architecture or the clustering approach with the fuzzy-C-means algorithm.

Conclusions that can be drawn from the research collated and presented in this paper with regards to the main benefits of using fuzzy logic based methods are:

- Fuzzy logic based methods particularly perform well under non-linear and time-variant conditions, where adaptive techniques have to be employed.
- When dealing with complex models that are not completely known and varying with time, fuzzy logic based methods can be used for faster convergence and reduced complexity with a slight degradation in performance compared to that of standard methods.
- When human expert knowledge is available, a fuzzy approach is highly suitable to incorporate this knowledge to complement available numerical data.

Building on the study we have presented in this paper, there are a few research areas that we consider merit further attention. These are: fuzzy adaptive equalization techniques for time-varying MIMO-channels, and fuzzy power control in MIMO-OFDM systems. Furthermore, another interesting area in which research is being conducted, is in cognitive radio [33]. Cognitive radio is an intelligent wireless communication system that adapts to its environment with the purpose of improving the spectrum efficiency. Both signal processing and machine learning techniques are of interest in cognitive radio, with, e.g., game theory being a commonly used method to model the transmit-power control problem. Research has also been conducted in fuzzy based game theory with the aim of application in cognitive radio in [34].

The purpose of this paper has been to give a background to common problems in wireless communication systems. Moreover, a survey of relevant research has been presented in wireless communication systems in which fuzzy methods have been used successfully. The aim has been to give a condensed and clear overview of conducted research as well as highlighting the common features of the problems in which fuzzy logic has been used in order to discern future areas to be investigated further by the ambitious researcher.

References

[1] L-X Wang and Jerry M. Mendel. Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(6):1414–1427, Nov-Dec 1992.

ISBN: 978-989-95079-6-8

[2] Antti Niemi, Jyrki Joutsensalo, and Tapani Ristaniemi. Fuzzy channel estimation in multipath fading CDMA channel. *The 11th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, 2:1131–1135, 2000.

[3] Andrew J Viterbi. *CDMA: Principles of Spread Spectrum Communication*. Prentice Hall PTR., 1995.

[4] Bart Kosko. *Neural Networks and Fuzzy Systems. A Dynamic Systems Approach to Machine Intelligence.* Prentice Hall, 1992.

[5] A. R. S. Bahai, B. R. Saltzberg, and M. Ergen. *Multi Carrier Digital Communications: Theory and Applications of OFDM.* Springer, 2004.

[6] M. Michele and M. Umberto. A comparison of pilot-aided channel estimation methods for OFDM systems. *IEEE Transactions on Signal Processing*, 49(12):3065–3073, December 2001.

[7] Jian Zhang, Zhi ming He, Xue gang Wang, and Yuan yuan Huang. A TSK fuzzy approach to channel estimation for OFDM systems. *Journal of Electronic Science and Technology in China*, 4(2), June 2006.

[8] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modelling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, 15(1):116–132, Jan-Feb 1985.

[9] L-X Wang. Fuzzy systems are universal approximators. *IEEE International Conference on Fuzzy Systems*, pages 1163–1170, March 1992.

[10] Jian Zhang, Zhi ming He, Xue gang Wang, and Yuan yuan Huang. TSK fuzzy approach to channel estimation for MIMO-OFDM systems. *IEEE Signal Processing Letters*, 14(6), June 2007.

[11] D. Gesbert, M. Shafi, Shiu Da-shan, P.J. Smith, and A. Naguib. From theory to practice: an overview of MIMO space-time coded wireless systems. *IEEE Journal on Selected Areas in Communications*, 21(3):281–302, April 2003.

[12] Hsuan-Yu Lin, Chia-Chang Hu, Yu-Fan Chen, and Jyh-Horng Wen. An adaptive robust LMS employing fuzzy step size and partial update. *IEEE Signal Processing Letters*, 12(8):545–548, August 2005.

[13] Jyh-Horng Wen, Cheih-Yao Chang, Gwo-Ruey Lee, and Ching-Yao Huang. Ofdm channel prediction using fuzzy update lms algorithm in time-variant mobile channel. *IEEE 64th Vehicular Technology Conference*, pages 1–5, September 2006.

[14] M. Nuri Seyman and Necmi Taspinar. Channel estimation based on adaptive neuro-fuzzy inference system in OFDM. *IEICE Trans. Commun.*, E91-B(7):2426–2430, July 2008.

[15] L-X Wang and Jerry M. Mendel. Fuzzy adaptive filters, with application to nonlinear channel equalization. *IEEE Transactions on Fuzzy Systems*, 1(3):161–170, August 1993.

[16] L-X Wang and Jerry M. Mendel. An RLS fuzzy adaptive filter with applications to nonlinear channel equalization. *Second IEEE International Conference on Fuzzy Systems*, 2:895–900, Mar-Apr 1993.

[17] P. Sarwal and M. D. Srinath. A fuzzy logic system for channel equalization. *IEEE Transactions on Fuzzy Systems*, 3(2):246–249, May 1995.

[18] K.Y. Lee. Complex RLS fuzzy adaptive filter and its application to channel equalization. *Electronics Letters*, 30(19):1572–1573, September 1994.

[19] Chee Seng Tang and Chin Leonard. Using multilayer perceptron fuzzy adaptive filter in non-linear channel equalization. *IEEE Global Telecommunications Conference, Globecom '95*, 2:884–887, November 1995.

[20] Sarat Kumar Patra and Bernard Mulgrew. Efficient architecture for bayesian equalization using fuzzy filters. *IEEE Transactions on Circuits and Systems*, 45(7):812–820, July 1998.

[21] L-X Wang and Jerry M. Mendel. Fuzzy basis functions, universal approximation, and orthogonal least-squares learning. *IEEE Transactions on Neural Networks*, 3(5):807–814, September 1992.

[22] Qilian Liang and Jerry M. Mendel. Equalization of nonlinear time-varying channels using type-2 fuzzy adaptive filters. *IEEE Transactions on Fuzzy Systems*, 8(5):551–563, October 2000.

[23] L. A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning. *Inform. Sci.*, 8:199–249, 1975.

[24] Eugenio Costamagna, Lorenzo Favalli, and Pietro Savazzi. Blind recovery of M-QAM signals for fading channels using the fuzzy-C-means algorithm. *IEEE 49th Vehicular Technology Conference*, 1:516–520, May 1999.

[25] Jenn-Kaie Lain and Chia-Tai Huang. FPCM-Assisted blind channel equalization of M-QAM signals for time-varying channels. *IEEE International Conference on Networking, Sensing and Control*, pages 787–791, April 2007.

[26] Shen Fang, Sun Yunshan, Zhang Liyi, Li Yanqin, and Li He. Blind equalization algorithm based on fuzzy neural network in QAM system. *2nd IEEE Conference on Industrial Electronics and Applications*, pages 484–487, May 2007.

[27] Sun Yunshan, Zhang Liyi, Li Yanqin, Li He, and Yan Junwei. A QAM blind equalization algorithm based on fuzzy neural network. *IEEE International Conference on Control and Automation*, pages 1420–1423, May-June 2007.

[28] D. Gansawat and T. Stathaki. Blind identification of volterra models using minimax optimization with higher order cumulants. *Proceedings. 7th International Conference on Signal Processing*, 1:164–167, Aug-Sep 2004.

[29] Yongzing Hou and Qinghua Bu. A new blind equalization algorithm based on the fuzzy neural network controller. *11th IEEE International Conference on Communication Technology Proceedings*, pages 466–469, Nov 2008.

[30] Murilo Bellezoni Loiola, Moiss Ribeiro, and Joao Marcos T. Romano. A turbo equalizer using fuzzy filters. *14th IEEE Signal Processing Society Workshop on Machine Learning for Signal Processing*, pages 695–704, Sep-Oct 2004.

[31] Jenn-Kaie Lain, Kun-Huang Kuo, and Yuan-Kai Zhang Liao. Low-complexity fuzzy filter-assisted turbo equalization for dispersive rayleigh-fading channels. *European Transactions on Telecommunications*, 19(2):137–148, 2008.

[32] Kun-Huang Kuo, Jenn-Kaie Lain, and Chia-Tia Huang. Reduced-complexity radial basis function-assisted turbo equalization for dispersive rayleigh-fading channels. *IEEE International Conference on Systems, Man, and Cybernetics*, 5:3691–3696, October 2006.

[33] Simon Haykin. Cognitive radio: brain-empowered wireless communications. *IEEE Journal on Selected Areas in Communications*, 23(2):201–220, February 2005.

[34] Elisabeth Rakus-Andersson, Maria Salomonsson, and Hang Zettervall. Ranking of weighted strategies in the two-player games with fuzzy entries of the payoff matrix. *Eighth International Conference on Hybrid Intelligent Systems*, pages 186–191, September 2008.

ISBN: 978-989-95079-6-8

Learning Fuzzy Rule Based Classifier in High Performance Computing Environment

Vinícius da F. Vieira[1], Alexandre G. Evsukoff[1], Beatriz S. L. P. de Lima[1] and Sylvie Galichet[2]

[1]COPPE/Federal University of Rio de Janeiro, Brazil
Rio de Janeiro, Brazil
[2]LISTIC/Polytech'Savoie, University of Savoie,
BP 80439, 74944 Annecy le Vieux, France

Email: vfvieira@gmail.com, alexandre.evsukoff@coc.ufrj.br, bia@coc.ufrj.br, sylvie.galichet@univ-savoie.fr

Abstract *— An approach to estimate the number of rules by spectral analysis of the training dataset has been recently proposed [1]. This work presents an analysis of such a method in high performance computing environment. Two approaches for parallel implementation of the method were studied considering the structure selection genetic algorithm and the spectral decomposition. The results show that both approaches have allowed to reduce considerably the overall processing time.*

Keywords — Fuzzy rule-based classifier; genetic algorithm; spectral clustering; high performance computing.

1 Introduction

Fuzzy rule-based classifiers have been widely studied and many approaches have been reported in literature [2]. One of the main problems in the design of interpretable fuzzy rule-based models is the determination of an adequate number of rules, which is closely related to the complexity of the problem. An approach to estimate the number of rules by spectral analysis of the training dataset has been recently proposed [1]. The method employs also a genetic algorithm for structure selection, which selects the input variables and defines the number of fuzzy membership functions in the domain of each selected input variable. Although this method has obtained interpretable fuzzy rules, with good accuracy, it is very time consuming due to the numerical complexity of both the spectral decomposition and the genetic algorithm.

This work presents an analysis of the method proposed in [1] in high performance computing environment. Two approaches for parallel implementation of the method were studied considering the structure selection by the genetic algorithm and the spectral decomposition. In the first approach, only the genetic algorithm is implemented in parallel while the spectral decomposition is computed in serial. In the second approach both the genetic algorithm and the spectral decomposition are run in parallel in a two levels parallelization scheme.

This paper is organized as follows. Next section introduces the notation and parameters of the fuzzy modeling approach. In section 3 the fuzzy rule induction algorithm [1] is reviewed. In section 4 the parallel implementation is described. In section 5 the results are discussed and the conclusions are drawn in last section.

2 Fuzzy Symbolic Modeling

Consider a dataset composed of N input-output observed samples $T = \{(\mathbf{x}(t), y(t)),\ t = 1...N\}$, where $\mathbf{x} \in \mathcal{R}^p$ is a vector of input variables and $y \in \mathcal{N}$ is the output variable where a value $y = j$ refers to a class $B_j \in \mathcal{B}$.

The fuzzy symbolic model (FSM) relates a set of input symbols $\mathcal{A} = \{A_i, i = 1...n\}$ to the set of classes $\mathcal{B} = \{B_j, j = 1...m\}$ as rules in the form $A_i \rightarrow B_j$ read as:

$$\textit{if } \mathbf{x}(t) \textit{ is } A_i \textit{ then } y(t) \textit{ is } B_j. \quad (1)$$

As usual, the fuzzy model computes an approximation $\hat{y}(t) = \hat{f}(\mathbf{x}(t))$ in three steps, which are implemented as the operators:

Fuzzification:

$$\hat{\mathbf{u}}(t) = F(\mathbf{x}(t), \Xi). \quad (2)$$

Inference:

$$\hat{\mathbf{v}}(t) = I(\hat{\mathbf{u}}(t), \Phi). \quad (3)$$

Defuzzification:

$$\hat{y}(t) = D(\mathbf{v}(t)). \quad (4)$$

The operators and their parameters, which should be computed by the learning algorithm, are presented in the following [1].

2.1 Fuzzification

The compound symbol A_i in rule (1) represents a region on the input variables domain. The membership function $\mu_{A_i}(\mathbf{x}(t))$ is computed by the conjunction of p elementary symbols $\mu_{A_k^i}(x_k(t)), k = 1...p$ as:

$$u_i(t) = \mu_{A_i}(\mathbf{x}(t)) = \prod_{k=1...p} \mu_{A_k^i}(x_k(t)) \quad (5)$$

The component A_k^i in (5) is associated to an elementary fuzzy set defined in the one dimensional domain of the variable x_k. The fuzzy membership functions are supposed to be equally spaced Gaussian functions, and only the

ISBN: 978-989-95079-6-8

number of fuzzy sets in each input variable fuzzy partition is necessary to define fuzzy membership functions [1].

The table $\Xi = \{\xi_{ik}, i = 1\ldots n, k = 1\ldots p\}$ is the output of the rule induction algorithm, presented in Section 3.2. An element $\xi_{ik} \in \Xi$ defines which symbol in the partition of the variable x_k should be associated to the component A_k^i in the rule i. A value $\xi_{ik} = j$ indicates that the fuzzy set A_{kj} should be used as the component A_k^i in the rule A_i (cf. (5)).

2.2 Inference

Rule weights have been shown to allow more flexibility to the fuzzy model [3][4]. The inference operator is defined by the rule weights matrix $\mathbf{\Phi} \in [0,1]^{n \times m}$, of which each component $\varphi_{ij} = \mu_\Phi(A_i, B_j)$ represents the confidence of the rule $A_i \rightarrow B_j$. The fuzzy inference is computed by the sum-product composition operator:

$$\hat{\mathbf{v}}(t) = \hat{\mathbf{u}}(t).\mathbf{\Phi} \tag{6}$$

The rule weights matrix are computed as the solution of a bounded quadratic optimization problem, as discussed in section 3.3.

2.3 Defuzzification

In classification problems, defuzzification computes the class output index. Generally the maximum rule is used, such that the class index is computed as the component with the greatest membership value:

$$\hat{y}(t) = j : v_j(t) = \max(\hat{\mathbf{v}}(t)) \tag{7}$$

The design of the FSM is presented next.

3 Induction of Fuzzy Rules

The rule induction algorithm is based on the spectral analysis of the training dataset to determine the number of fuzzy rules. The rule-base weights are then computed by solving a bounded quadratic optimization problem.

3.1 Spectral Analysis

Consider a graph associated to the training dataset, of which the set of nodes represent the input variable samples and the set of edges is defined by the $N \times N$ adjacency (or affinity) matrix $\mathbf{A}$. The graph is undirected and weighted, hence the affinity matrix is symmetric, real valued and its elements represent the similarity between two input variable samples. In this work, the similarity metric is computed by the Gaussian function, such that each element of affinity matrix is computed as:

$$a_{ij} = \begin{cases} \exp\left(-\dfrac{\|\mathbf{x}(i) - \mathbf{x}(j)\|^2}{2\sigma^2}\right), & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where σ is a dispersion parameter that controls the spread of the similarity function, fixed in this work at $\sigma = 1$. The $N \times N$ matrix $\mathbf{D}$ is a diagonal matrix whose elements are the degree of the nodes of the graph, computed by the sum of similarities of neighbors of each node:

$$d_{ii} = \sum_{j=1..N} a_{ij} \tag{9}$$

The spectral analysis of the graph is based on the normalized Laplacian matrix, computed as [5][6]:

$$\mathbf{L} = \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2} \tag{10}$$

The rule induction algorithm is based on the spectral decomposition of the Laplacian matrix as:

$$\mathbf{L} = \mathbf{Z}\mathbf{\Lambda}\mathbf{Z}^{\mathrm{T}} \tag{11}$$

The number of positive eigenvalues of the Laplacian is used as an estimation of the number of rules [1].

3.2 The Rule Induction Algorithm

The combination of one-dimensional fuzzy sets is easier to understand than a multi-dimensional fuzzy set. Thus, the idea of the rule induction algorithm is to run a clustering algorithm on the dataset and then to associate a rule to each cluster. For each cluster, the nearest combination to the cluster center, among all possible combinations of fuzzy sets is chosen as the rule to be included in the model.

In the spectral clustering algorithm, the standard k-means algorithm is applied over the matrix $\widehat{\mathbf{Z}}$, computed as the first K columns of the eigenvector matrix $\mathbf{Z}$. Recall that the columns of $\mathbf{Z}$ are ordered according to the corresponding eigenvalues, such that the first columns correspond to the largest eigenvalues.

The rule induction algorithm is run within an outer loop for structure selection, which defines both the variables that should be used by the model and the number of fuzzy sets in each variable's domain. At each iteration, the structure selection algorithm provides the candidate structure represented by the vector $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_p]$, where each component $2 \leq \gamma_k \leq \gamma_{\max}$ is the number of fuzzy sets in the fuzzy partition of variable x_k and a value $\gamma_k = 1$ indicates that the variable x_k should not be considered by the model. The rule induction algorithm is sketched out in Algorithm 1.

Algorithm 1: Rule induction

input: $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_p]$; $T = \{(\mathbf{x}(t), \mathbf{y}(t)), t = 1...N\}$

output: $\Xi = \{\xi_{ik}, i = 1\ldots n, k = 1\ldots p\}$;

```
01 begin
02    compute the matrix A    //Equation (8)
03    compute the matrix D    //Equation (9)
04    compute L = D^{-1/2} A D^{-1/2}
05    compute L = Z Λ Z^T
06    n ← K : λ_i ≥ δ, i = 1...K  //estimate n rules
07    W ← clustering(n)       //assign rules to clusters
08    for i = 1...n
09       for k = 1...p
10          if (γ_k > 1)
11             ξ_ik = j* : α_kj* = arg min(α_kj − w_ik);
12 end
```

ISBN: 978-989-95079-6-8

The set of rules issued by the rule induction algorithm defines the premises of rules (1). The same set of rules is used for all classes, with the classification being assigned by the rule weights, which are computed as described next.

3.3 Rule Weights Optimization

The rule weight matrix can be written as $\mathbf{\Phi} = [\boldsymbol{\varphi}_1 | \ldots | \boldsymbol{\varphi}_m]$, where the vectors $\boldsymbol{\varphi}_j$ are the rule confidence weights that relates the rule premise symbol A_i to the rule consequent (class) B_j. Optimal rule weights $\boldsymbol{\varphi}_j$ can be computed in the least squares sense, by solving the following bounded quadratic programming problem for each class B_j [1][7]:

$$\underset{\boldsymbol{\varphi}_j}{minimize}\ \tfrac{1}{2}\boldsymbol{\varphi}_j^{\mathrm{T}}\mathbf{K}\boldsymbol{\varphi}_j - \mathbf{C}^{\mathrm{T}}\boldsymbol{\varphi}_j \qquad (12)$$

$$subject\ to\ 0 \le \varphi_{ij} \le 1,\ i = 1...n$$

where $\mathbf{K} = \mathbf{U}^{\mathrm{T}}\mathbf{U}$ is a strictly positive definite matrix and $\mathbf{C}^{\mathrm{T}} = \mathbf{V}_j^{\mathrm{T}}\mathbf{U}$. The matrix $\mathbf{U} = [\mathbf{u}(1) | \ldots | \mathbf{u}(N)]^{\mathrm{T}}$ is the matrix of which each line is the fuzzy vector computed for each data sample, whose components are computed as (5). The vector $\mathbf{V}_j = [v_j(1) | \ldots | v_j(N)]^{\mathrm{T}}$ contains the memberships values of class B_j for each sample in the training set, such that $v_j(t) = \mu_{B_j}(y(t))$. The bounds attempt to constrain the weights within the $[0,1]$ interval, such that they can be interpreted as fuzzy rule confidence. The bound constrains also avoid high values for the rule weight parameters, providing the smoothness of the solution.

The quadratic optimization problem defined by (12) is widely known and the solution can be computed by efficient numerical algorithms [8].

For a fixed model structure, the more the solution values $\varphi_{ij} \approx 1$, the more certain is the classifier and also the greater is the norm of the weight vector, defined by:

$$\left\|\boldsymbol{\varphi}_j\right\|^2 = \boldsymbol{\varphi}_j^{\mathrm{T}}.\boldsymbol{\varphi}_j,\ j = 1...m \qquad (13)$$

The sum $\sum_{j=1..m}\left\|\boldsymbol{\varphi}_j\right\|^2$ is thus a measure of the amount of certain rules and also can be interpreted as a measure of the classifier quality.

3.4 Structure Selection

The structure selection is implemented by a genetic algorithm (GA). Each individual in the population represents an estimation of the structure vector $\boldsymbol{\gamma} = [\gamma_1,\ldots,\gamma_p]$, which defines the variables that must be included in the model and also the number of fuzzy sets in the domain of each variable.

The GA fitness function is computed as a trade-off between the model's accuracy and capacity according to the required complexity of the problem [1]:

$$R(\boldsymbol{\gamma}) = \frac{1}{2N}\sum_{j=1...m}\left\|\mathbf{U}(\boldsymbol{\gamma})\boldsymbol{\varphi}_j(\boldsymbol{\gamma}) - \mathbf{V}_j\right\|^2 + \frac{1}{2N}\sum_{j=1...m}\left\|\boldsymbol{\varphi}_j(\boldsymbol{\gamma})\right\|^2 \qquad (14)$$

where the firs term represent the model accuracy and the second accounts for model quality according to (13). The factor $\frac{1}{2N}$ is used to rescale the fitness function to the $[0.1]$ interval.

The GA performs simultaneously the selection of input variables to be considered in the model and the number of fuzzy sets in the domain of each (selected) variable. During the GA execution, it is frequently necessary to evaluate an individual that has been previously evaluated. The number of fuzzy rules is computed from the spectral analysis, which is a function only of the selected input variables. It is thus possible that two different individuals select the same input variables such that they will result in the same number of rules. In order to avoid the unnecessary reevaluation of the spectral analysis for individuals related to the same subset of variables, the number of rules computed by the solution of (11) is stored and related to the corresponding subset of variables. Consequently, the eigendecomposition (11) for a new individual is computed only if the corresponding subset of variables has not been evaluated before. The fitness functions for all visited individuals are also stored such that for a new individual, identical to another already evaluated, the fitness function (14) is not reevaluated.

The GA parameters for all tests described in section 5 were the same. The population was fixed at 32 individuals. An uniform crossover operator with 85% crossover rate was used. The mutation rate was set in 5%. The maximum number of iterations was fixed at 150. A value $\gamma_{\max} = 8$ was used to limit the number of fuzzy sets per variable.

All the implementation was made by free source software available on the web. The GA was implemented in Python programming language. The most computationally expensive tasks (spectral analysis and the rule weights optimization) were implemented in C using standard libraries.

4 Parallel Implementation

The structure selection GA is a very intensive computational task, since each candidate evaluation requires the solution of equation (11) and the GA itself requires the evaluation of a number of individuals during several iterations. The modern multi-core hardware platforms allow the reduction of the global processing time in complex application through parallel implementation.

Depending on the application and the available hardware resources, the parallelism can be exploited in different ways, through the partitioning of the problem with respect to data and/or tasks. In data parallelism, the same task is allocated to different processors, where each one of them works with a different portion of the data. This approach applies to problems whose solution involves a great mass of data. In task parallelism, different tasks are allocated to different

ISBN: 978-989-95079-6-8

processors. This approach applies to problems where there is interdependency among the tasks.

When designing a parallel program, the communication among the processors involved in the execution must be considered carefully. There are two main communication paradigms in parallel applications: memory sharing and message passing. The memory sharing parallel programming model is based in the same principle as the serial programming; essentially, a single memory addressing space is used during the program execution flow. The message passing programming is conceptually independent of the hardware platform, operating system and programming language. As the memory addressing spaces are different in each process the concept of message is used. A message is a set of information that can be sent from process to another through a communication channel.

One of the most known and complete libraries for message passing is the Message Passing Interface (MPI). The MPI is a standardized and portable message passing library designed for distributed memory environments, parallel machines and heterogeneous network. In MPI programs, all the parallelism is explicit, i.e., the programmer must code the synchronization points and the message passing calling the functions. Each process, receives a copy of the entire parallel program code.

In this work two parallel strategies are evaluated: in the first one only the GA is implemented in parallel; in the second one, both the GA and the spectral analysis for each individual are processed in parallel in a two levels strategy.

4.1 Parallel Genetic Algorithm

Several features of GAs allow its parallelization in a very natural way and many approaches have been reported in literature [9]. The main approaches are divided in three types: fine-grained single population GA, coarse-grained multiple population GA and single global population master-slave GA.

Fine-grained parallel genetic algorithms use just one population, but there is a special structure that limits the interaction among the individuals. An individual can only compete and operate with individuals that belong to his neighborhood, which is assigned by a predefined topology.

Coarse-grained parallelization model, also known as island model, uses several populations that occasionally change individuals through migration operator. In island GAs, each sub-population evolves as in traditional GAs, excepting for the migration operator among different islands.

The master-slave parallel genetic algorithms, also known as parallel global population GAs, use a single population and the parallelization is done in the evaluation of the individuals. As in a traditional serial GA, each individual competes and operates with any other individuals in the population. In master-slave genetic algorithms, a master process is responsible by the application of the genetic operators to the population (selection, crossing, mutation) while the slave processors are responsible by the individuals' evaluation [10]. The communication among the individuals occurs when the master process sends the individuals to the slave processes and when the slave processes return their fitness values to the master.

In this work, a master-slave parallel implementation was used for the structure selection GA. The adoption of such GA parallel model is useful, since its implementation is independent of the machine architecture and of the communication model among the processes.

4.2 Two Levels Approach: Parallel GA with Parallel SA

The complexity of a full eigenproblem like (11) is $O(N^3)$ in processing time and $O(N^2)$ in storage requirements. In Algorithm 1, the solution of (11) requires the solution of an eigenproblem of the size of the number of records of the dataset N. The two-level strategy considers thus the parallelization of the spectral analysis (SA) for each individual, within the parallel GA.

The schematic representation of the two-level strategy is shown in Fig. 1. In the higher level there is the GA parallelism, where the master performs the operations over the population and sends the individuals to the slaves. In the lower level is the SA parallelism, where each slave shares the eigenproblem solution with other processors.

The two levels strategy is possible due to the concept of communicators groups in the MPI library that allows the definition of modules that encapsulate the communication operations within a group. The communicator identifies the group of processes and the context in which a given operation must be applied. The MPI library provides a standard communicator, called MPI COMM WORLD, which identifies every processes involved in the computation through global identifiers.

The execution of the GA is performed by the processors that belong to Group 0 (cf. Fig. 1.). Each slave processor in Group 0, also belongs to another group, which is responsible for computing the SA. Thus, new sub-communicators must be created, allowing the communication within each group of slave processors.

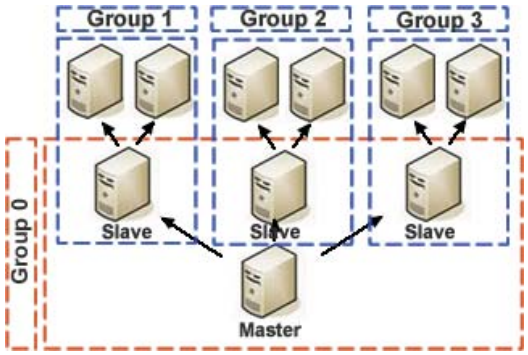


Fig. 1: Two-level Parallel Architecture

There are numbers of software libraries that implements parallel eigenproblem solution using different numerical methods. In this work, SLEPc (Scalable Library for Eingenvalue Problem Computation [11]) was adopted. SLEPc is a software package for solution of large scale problems built over PETSc (Portable, Extensible Toolkit for

ISBN: 978-989-95079-6-8

Scientific Computation), which is a set of libraries of data structure and routines for parallel scientific applications.

The Krylov-Schur solver, which implements the Arnoldi method for eigenvalues problem solution, was adopted in this work. It is probably the most robust solver in SLEPc [11].

5 Results and Discussion

5.1 Performance Results

Four benchmark datasets from the UCI Machine Learning Repository [12] were considered for the performance evaluation. Some synthetic datasets, with 18 variable and different numbers of records, were also considered. These datasets were considered in order to exploit a greater problem size.

Each benchmark evaluation was performed in 10-fold cross validation analysis. The evaluation of each benchmark was repeated 10 times to ensure the GA generalization. The average results for each dataset are shown in Table 1, where Acc. is the average accuracy (computed for the 10-fold cross validation) and Time is the average global time (in seconds) of the serial implementation. For the synthetic datasets just the time results were considered and the time values correspond to the execution time for just one GA generation.

All tests were run an 8-core Linux cluster composed of two machines, each one equipped with two dual-core Intel Xeon processors (1.6GHz) and 4GB DRAM. The serial processing time was computed using only one processor.

Table 1: Performance Results.

Dataset	p	m	N	rules	Acc.	Time
Diabetes	8	2	768	12.5	78.67	1795.72
Ionosphere	32	2	344	26.7	91.47	4752.24
Balance	4	3	625	8.3	81.41	145.69
Cancer	9	2	683	8.4	98.13	1665.62
Synthetic 1	18	-	1000	-	-	99.68
Synthetic 2	18	-	2000	-	-	580.18
Synthetic 3	18	-	3000	-	-	1952.7
Synthetic 4	18	-	4000	-	-	4339.75

As seen in Table 1, the modeling approach has obtained good accuracy with small number of rules results. The accuracy results obtained for the same datasets by other algorithms can be found in [1]. The focus of the analysis in this work is on the processing time and parallelization efficiency.

5.2 Parallel GA Results

The processing time (in seconds) of the parallel GA (with serial SA) implementation are shown in Table 2. The parallel processing time as a percentage of the serial processing time is also shown in Table 2. It can be seen that the parallel GA reduces the global execution time considerably.

Table 2: Execution time for parallel GA (with serial SA).

	2 Proc.		4 Proc.		8 Proc.	
	Time	%	Time	%	Time	%
Diabetes	1194.8	66.5	738.28	41.11	550.75	30.67
Ionosphere	2517.57	52.97	1354.1	28.49	853.56	17.96
Balance	107.29	73.64	88.35	60.64	69.51	47.71
Cancer	1070.35	64.26	696.54	41.81	476.76	28.62
Synthetic1	54.79	54.96	47.66	47.81	54.16	54.33
Synthetic2	317.14	54.66	210.73	36.32	230.89	39.79
Synthetic3	1050.2	53.78	700.89	35.89	620.05	31.75
Synthetic4	2380.84	54.86	1765.83	40.68	1486.28	34.24

The parallel efficiency of the parallel GA algorithm is shown in Fig. 2. The parallel efficiency was computed in the standard way as [13]:

$$\varepsilon_P = 100.\frac{1}{P}\frac{T_S}{T_P} \tag{15}$$

where P is the number of processors, T_S the serial processing time (in seconds) and T_P is the parallel processing time using P processors.

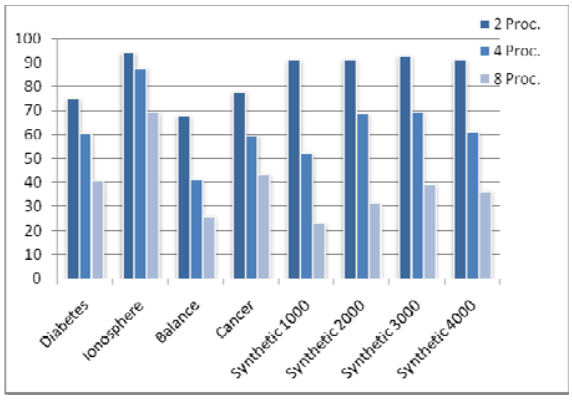


Fig. 2: Parallel Efficiency for Parallel GA With Serial SA.

The parallel efficiency decreases as the number of processors increases, due communication overhead and the load imbalance introduced by the reevaluation avoiding scheme, since the number of evaluated individuals is not the same for all the slave processes. A processor with fewer individual's evaluations will become idle waiting the other processors finish their work to synchronize in a new generation. The resulting load imbalance reflects in a great damage to the overall parallel efficiency.

The results show that this approach is not good for large datasets, since each GA individual evaluation allocates an instance of the Laplacian matrix in memory, making a bad use of the memory resources available.

5.3 Parallel GA with Parallel SA Results

The execution times (in seconds) of the two levels approach for the studied datasets are shown in Table 3 as so as the percentage of the serial processing time.

Table 3: Execution times for the two levels approach.

	2 Proc.		4 Proc.		8 Proc.	
	Time	%	Time	%	Time	%
Diabetes	2185.32	121.6	1303.46	72.58	833.62	46.42
Ionosphere	5704.34	120.0	3401.24	71.57	2949.77	62.07
Balance	185.67	127.4	128.28	88.05	116.71	80.10
Cancer	1628.91	97.79	994.00	59.67	745.43	44.75
Synthetic1	76.01	76.25	40.03	40.15	35.41	35.52
Synthetic2	368.81	63.56	195.59	33.71	160.38	7.64
Synthetic3	1182.08	60.53	615.06	31.49	416.85	21.34
Synthetic4	2537.81	58.47	1300.7	29.97	805.05	18.55

ISBN: 978-989-95079-6-8

The parallel efficiency of the two-level approach for the same datasets are shown in Fig. 3.

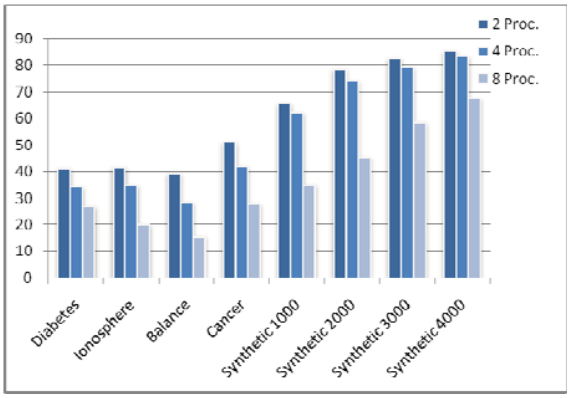


Fig. 3: Parallel Efficiency for the Two-Level Approach.

The results in Table 3 and Figure 3 show that for the four benchmark datasets studied, the global execution time is greater than the results presented for the parallel GA with serial SA. Consequently, the parallel efficiency is lower when compared to the results presented by those approach. This indicates that for small size datasets, the parallelization of the eigenproblem solution is not efficient. For larger datasets, the two levels approach shows more interesting results since the synthetic datasets present better results for the two levels approach when compared to the parallel GA with serial SA. This comparison indicates that, for datasets with a great number of records, the SA parallelization can be a good idea since it reduces the load imbalance introduced by the reevaluation avoiding strategy and makes a better use of the memory available for the method execution, since the Laplacian matrix used in the spectral analysis is sliced for the involved processes.

6 Conclusions

The design of fuzzy symbolic models [1] is focused on the interpretability and the resulting model's structure is selected by a wrapper genetic algorithm that defines the variables that should be included into the model and the number of fuzzy sets in the partition of each variable. The number of rules is determined by a spectral analysis method and rules' weights are optimized.

This work has evaluated two approaches for the parallel implementations of the fuzzy symbolic models. Both approaches have reduced the global execution time of the algorithm when compared to the serial times, which was the main drawback of the previous implementations.

The parallel efficiency results indicate that, even with the load imbalance introduced by the reevaluation avoiding strategy, the GA parallelization is an interesting approach for smaller datasets and the two-level approach is a good approach for larger datasets. Further studies may be done in order to increase the load balance among the processors and evaluate the implementation behavior for larger datasets.

Acknowledgement

The authors are grateful to the Laboratory of Computational Physiology (FISIOCOMP) at Federal University of Juiz de Fora (UFJF), where all the parallel experiments were run. This work has been supported by the Brazilian Research Council CNPq and the Brazilian Agency CAPES under the agreement CAPES/Cofecub.

References

[1] A.G. Evsukoff, S. Galichet, B. S. L. P. de Lima and N. F. F. Ebecken, "Design of interpretable fuzzy rule-based classifiers using spectral analysis with structure and parameters optimization". Fuzzy Sets and Systems (2008), doi: 10.1016/j.fss.2008.08.010.

[2] E. Hüllermeier, "Fuzzy methods in machine learning and data mining: status and prospects", Fuzzy Sets and Systems, vol. 156, no. 3, pp. 387-406, 2005.

[3] H. Ishibuchi and T. Nakashima, "Effect of rule weights in fuzzy rule-based classification systems" IEEE Trans. Fuzzy Systems, vol. 9, no. 4, pp. 506–515, 2001.

[4] H. Ishibuchi and T. Yamamoto, "Rule weight specification in fuzzy rule-based classification systems", IEEE Trans. on Fuzzy Systems, vol. 13, no. 4, pp. 428-435, 2005.

[5] M. Filippone, F. Camastra, F. Masulli and S. Rovetta, "A survey of kernel and spectral methods for clustering". Pattern Recognition, vol. 41, pp. 176-190, 2008.

[6] F. R. K. Chung, Spectral Graph Theory, CBMS Regional Conf. Series in Mathematics., no. 92. American Mathematic Society, 1997.

[7] A. G. Evsukoff, "Learning fuzzy rule-based classifier with rule weights optimization and structure selection by a genetic algorithm", Proceedings of the 2007 IEEE International Conference on Fuzzy Systems, London, 23-26 July, 2007.

[8] B. Schölkopf and A. J. Smola, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, MIT Press, 2001.

[9] E. Cantú-Paz, "A Survey of Parallel Genetic Algorithms", Calculateurs Paralleles, vol. 10, 1998.

[10] C. R. Xavier, V. F. Vieira, D. M. S. Martins, R. W. Santos, "Comparing Two Parallel Genetic Algorithms for the Inverse Problem Associated to the Cardiac Bidomain Equations", IEEE 18th International Symposium on Computer Architecture and High Performance Computing, Ouro Preto, 17-20 October, 2006.

[11] V. Hernandez , J. E. Roman, V. Vidal, "SLEPc: A scalable and flexible toolkit for the solution of eigenvalue problems", ACM Transactions on Mathematical Software, v.31 n.3, p.351-362, September 2005.

[12] A. Asuncion, and D. J. Newman, "UCI Machine Learning Repository" [http://www.ics.uci.edu/~mlearn/MLRepository.html]. Irvine, CA: University of California, School of Information and Computer Science, 2007.

[13] V. Kumar, A. Grama, A. Gupta and G. Karypis, Introduction to Parallel Computing - Design and Analysis of Algorithms, The Benjamin/Cummings Publishing Company, Inc., CA, 1994

ISBN: 978-989-95079-6-8

The Use of Interval Type-2 Fuzzy Logic as a General Method for Edge Detection

Olivia Mendoza[1] Patricia Melin[2] Juan R. Castro[1]

1 School of Engineering, University of Baja California
Tijuana, Mexico
2 Department of Research and Graduate Studies, Tijuana Institute of Technology
Tijuana, Mexico
Email: omendoza@uabc.mx, epmelin@hafsamx.org, jrcastror@uabc.mx

Abstract—*We describe a method for edge detection in digital images based on the morphological gradient and fuzzy logic. The goal is to improve one of the basic methods for edge detection in order to obtain a better result even without applying any filter to the image. The tests were made with a type-1 fuzzy inference system (T1FIS) and with an interval type-2 fuzzy inference system (IT2FIS). We show that the images obtained with fuzzy logic are better than the ones obtained with only the morphological gradient method. In particular the IT2FIS achieved the best results, because of the flexibility to model the uncertainty in the gradient values and the gray ranges for the edge images. In both TIFIS and IT2FIS the membership function parameters were obtained directly from the images; this allows application of the proposed method to images with different gray scales.*

Keywords— digital images, edge detection, image processing, interval type-2 fuzzy logic.

1 Introduction

Edge detection is an intermediate step in image pattern recognition in digital images and has the opposite effect of noise elimination; it consists on emphasizing pixels with gray tones that are different than their neighbours [1]. Some of the existing methods for edge detection have been enhanced with type-1 fuzzy logic [2][3], fractal theory [4][5], neural networks [6][7][8] and recently with interval-valued fuzzy operators [9].

In previous works proposed by the authors, the improvement of the Sobel edge detector using an IT2FIS was achieved [10][11]. The images resulting from these experiments were used as training data for image recognition with modular neural networks, obtaining good results [12][13].

In this paper, an improvement of a traditional edge detection method based on the morphological gradient is presented. This method does not need any filter to obtain the edges, but it is necessary to perform calculations in order to find the relation of each pixel of the image with the eight neighbours pixels around it.

In section two we present a brief description of an edge detection method based in morphological gradient, in section three we explain our contribution consisting on the development and tests with a type-1 fuzzy inference system and an interval type-2 fuzzy inference system; in both cases we apply the same criteria and conditions, for a valid comparison of the results. Section four shows the obtained results with the three methods. The main topic in this paper is the comparison of the results obtained with the IT2FIS, applying different footprint of uncertainty (FOU).

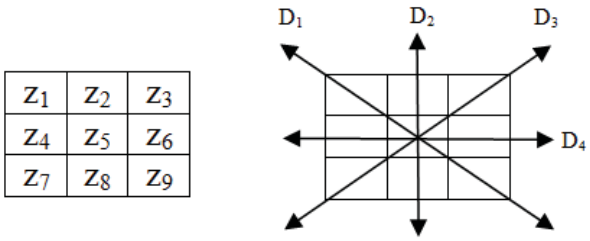
2 Edge detection using morphological gradient

The morphological gradient of a grayscale image can be viewed as the greatest absolute intensity difference between any two pixels within the structuring element, and can be defined with Equation (1) [14].

$$\left| \nabla(f) = \delta_g(f) - \varepsilon_g(f) \right| \tag{1}$$

For the discrete case, we use D_i instead $\nabla(f)$, then if we apply (1) to a 3x3 sample matrix as shown in Fig. 1, we can to obtain coefficients z_i with Equation (2), the possible edge directions D_i with (3). An approximation of (3) can be calculated without the root square using the absolute values of the differences instead the square values (4). The edges E can be calculated with Equation (5) [15][16].

$$\begin{aligned} z_1 &= f(x-1, y-1) & z_2 &= f(x, y-1) \\ z_3 &= f(x+1, y-1) & z_4 &= f(x-1, y) \\ z_5 &= f(x, y) & z_6 &= f(x+1, y) \\ z_7 &= f(x-1, y+1) & z_8 &= f(x, y+1) \\ z_9 &= f(x+1, y+1) \end{aligned} \tag{2}$$



Z1	Z2	Z3
Z4	Z5	Z6
Z7	Z8	Z9

Figure 1: A 3x3 sample matrix with coefficients and edge directions.

ISBN: 978-989-95079-6-8

$$D_1 = \sqrt{(z_5 - z_1)^2 + (z_9 - z_5)^2} \quad (3)$$
$$D_2 = \sqrt{(z_5 - z_2)^2 + (z_8 - z_5)^2}$$
$$D_3 = \sqrt{(z_5 - z_3)^2 + (z_7 - z_5)^2}$$
$$D_4 = \sqrt{(z_5 - z_4)^2 + (z_6 - z_5)^2}$$

$$D_1 = |z_5 - z_1| + |z_9 - z_5| \quad (4)$$
$$D_2 = |z_5 - z_2| + |z_8 - z_5|$$
$$D_3 = |z_5 - z_3| + |z_7 - z_5|$$
$$D_4 = |z_5 - z_4| + |z_6 - z_5|$$

$$E = D_1 + D_2 + D_3 + D_4 \quad (5)$$

The edge image obtained with Equation (5) does not consider any parameters to prevent noise or regions with very low or high contrast. Then some very dark or bright images could be a problem for this edge detector.

3 Edge detection using a fuzzy inference system

The fuzzy approach for edge detection using morphological gradient consists of using Equation (3) or (4) to obtain the gradient in the four directions, and use them as inputs for a fuzzy inference system (FIS) instead the Equation (5). This approach allows us to analyze the gray tones range of any particular image and use them as parameters to the membership functions in the FIS.

3.1 Type-1 fuzzy linguistic variables

The inputs to consider for the T1FIS are the morphological gradients D_i for each pixel with respect to a 3x3 neighbourhood matrix. The membership functions must represent the magnitude of the gradient. In order to adapt the membership functions to the range of gray tones depending on the image, we obtain the maximum, minimum and middle values of D_i, and we use these values for the centre of the membership functions, as shown in the example of Fig. 3.

The σ value for the particular example plotted in Fig. 3 is 20, the Gaussian membership functions for each D input were obtained with (9)(10) and (11), and the center of each function were obtained with (6)(7) and (8). This is the method that we propose to adapt the parameters of the membership functions depending on the contrast level of each image.

$$dlow_i = \min(D_i) \quad (6)$$
$$dhigh_i = \max(D_i) \quad (7)$$
$$dmiddle_i = dlow_i + (dhigh_i - dlow_i)/2 \quad (8)$$

$$\mu(dlow_i) = e^{\frac{-(x-dlow_i)^2}{2(\sigma_i)^2}} \quad (9)$$
$$\mu(dhigh_i) = e^{\frac{-(x-dhigh_i)^2}{2(\sigma_i)^2}} \quad (10)$$
$$\mu(dmiddle_i) = e^{\frac{-(x-dmiddle_i)^2}{2(\sigma_i)^2}} \quad (11)$$

where $\sigma_i = dhigh_i / 8$

For the outputs E (the edges), we select the range [0,255] as the gray tones scale, then we can obtain these membership functions directly with (12)(13) and (14). If some application needs a different gray tones scale, is necessary to adjust the values of *black*, *gray* and *white* of (12), (13) and (14) to the desire scale.

$$\mu_{black} = e^{\frac{-(x-black)^2}{2\sigma^2}} \quad (12)$$

where black=0

$$\mu_{gray} = e^{\frac{-(x-gray)^2}{2\sigma^2}} \quad (13)$$

where gray=255/2

$$\mu_{white} = e^{\frac{-(x-white)^2}{2\sigma^2}} \quad (14)$$

where white=255

where $\sigma = white/8$

In order to plot the membership functions for the inputs D_i of the FIS, as particular example we used the image in Fig. 2. First the image was converted to gray scale, then the following values were calculated, for each variable D.

$dlow_1$=0, $dmiddle_1$=94.7, $dhigh_1$=189.4, σ_1=23.6
$dlow_1$=0, $dmiddle_2$=87.9, $dhigh_2$=175.8, σ_2=21.9
$dlow_1$=0, $dmiddle_3$=89.8, $dhigh_3$=179.6, σ_3=22.4
$dlow_1$=0, $dmiddle_4$=75.6, $dhigh_4$=151.3, σ_4=18.9



Figure 2. Particular image to explain the parameters for the FIS variables.

ISBN: 978-989-95079-6-8

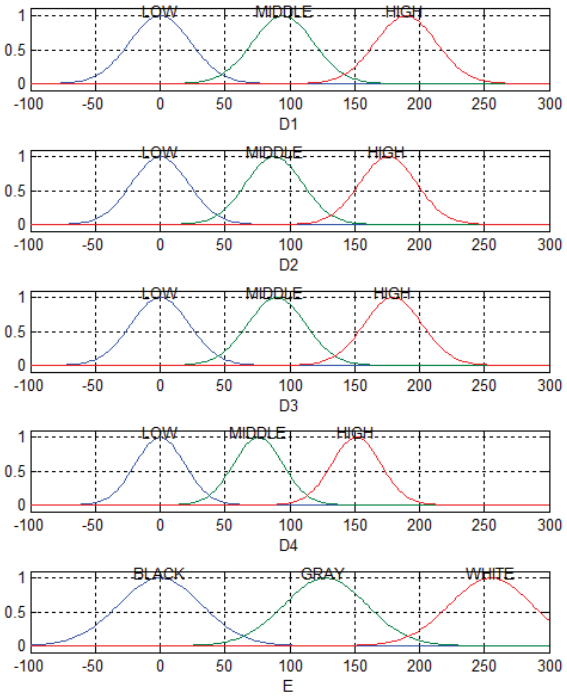


Figure 3: Linguistic variables for the TIFIS corresponding to the image in Fig. 2.

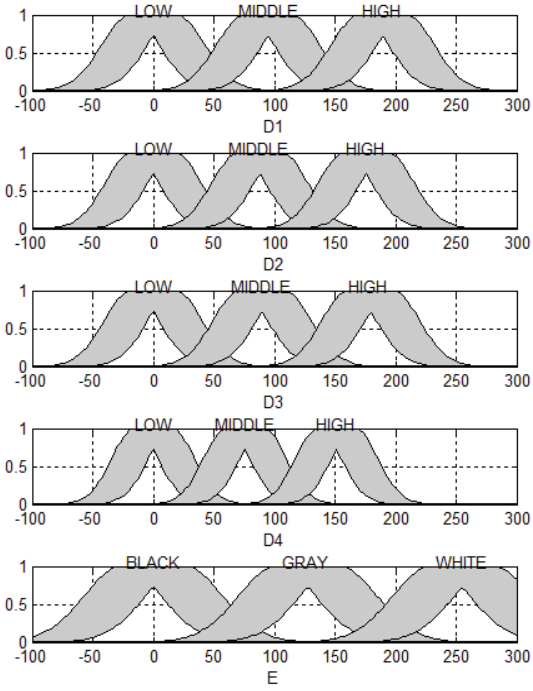


Figure 4: Linguistic variables for the IT2FIS corresponding to the image in Fig. 2, with $FOUd_i=0.4*dmiddle_i$

3.2 *Interval type-2 linguistic variables*

As we previously mentioned, the edge detection method proposed in this paper can be adapted to images with different contrast levels, but particularly the use of interval type-2 fuzzy sets enables us to handle the uncertainty by adding some FOU to the membership functions [17].

To compare the improvement of the edge detector using interval type-2 fuzzy sets, we build an IT2FIS using the same parameters than the TIFS [18]. The membership functions were obtained with Equations (6)-(14), but adding different sizes of FOU.

We made tests using different sizes of the FOU for the input variables D_i, using $dmiddle_i$ as parameter to obtain the values respect to the image contrast, in a proportional manner. The FOU sizes for D_i, then can be obtained using (15).

$$FOUd_i = \varepsilon * dmiddle_i \tag{15}$$

where ε is in (0,1)

The FOU for the output variable E, was calculated in a similar way to the inputs variables. In this case, we obtain the FOU values using *egray* as parameter (16).

$$FOUe = \varepsilon * egray \tag{16}$$

where ε is in (0,1)

The plot in Fig. 4, shows the linguistic variables with interval type-2 membership functions where the value for ε is *0.4*, adding $FOUd_i/2$ for inputs D_i and $FOUe_i/2$ for the output E, at each side of the center of the membership functions; then all the functions are symmetric.

3.3 *The rules*

The task of the fuzzy inference systems is to estimate a value for each pixel, which better describes the relation with its neighborhood.

1. If (D1 is HIGH) or (D2 is HIGH) or (D3 is HIGH) or (D4 is HIGH) then (E is WHITE)
2. If (D1 is MIDDLE) or (D2 is MIDDLE) or (D3 is MIDDLE) or (D4 is MIDDLE) then (E is WHITE)
3. If (D1 is LOW) and (D2 is LOW) and (D3 is LOW) and (D4 is LOW) then (E is BLACK)

The first rule establishes that a high gradient in any direction means an edge, the second rule is similar to the first one, because after many experiments with different images, a medium magnitude of the gradient in any direction results also in an edge, and the third rule is only to confirm the first two, because all derivatives with low magnitude means a homogeneous region in the image, that means there are not edge in this pixel. In Fig. 5 the solution surfaces for the TIFIS and IT2FIS are shown.

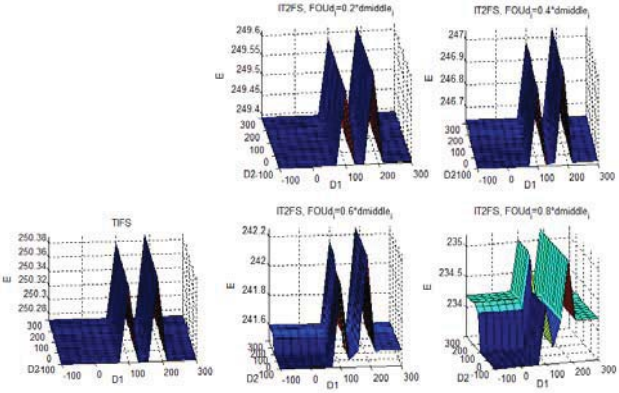


Figure 5: The solution surfaces for the image in Fig. 2, for TIFIS and IT2FIS with different FOU sizes.

ISBN: 978-989-95079-6-8

4 Experimental results

For a better appreciation of the results, we developed some programs in Matlab to obtain the edge image, under the same conditions. In all the images the morphological gradients were obtained with (2) and (3), then using the morphological gradient (MG) the edges were calculated with (5). For the fuzzy methods T1FIS and IT2FIS we use a fuzzy inference system instead of the gradient (5).

We choose Gaussian membership functions for all the variables of the both FIS (TIFIS and ITFIS), only for an easy configuration, because as we explained in section three we used the values obtained with the morphological gradient of each particular image to estimate the parameters for the membership functions.

For the first test, we used a sample of the ORL databases of faces, and selected arbitrary values for the FOU, only to demonstrate that the images obtained using IT2FIS are better that the images obtained with TIFIS. In the example of Fig. 6, we see the changes on the images with different values of the FOU. The best image is the one obtained with FOU=40, while the image obtained with T1FIS looses some details. For all the tests the use of the FIS's improve the results of the MG traditional method.

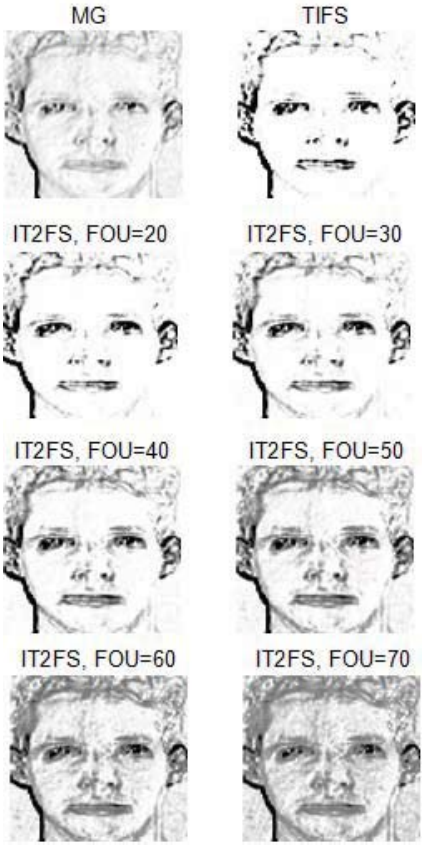


Figure 6: Edges images obtained using Morphological Gradient (MG), and the same method improved with a type-1 fuzzy system (TIFIS), and an interval type-2 fuzzy system (IT2FIS), with arbitrary FOU values.

For a better analysis of the results we plot frequency histograms for each result in Fig. 6. The histograms show that the images obtained with the IT2FIS when the FOU is increased, between 30 and 40, found more pixels corresponding to the edges preserving more details than T1FIS, but when the FOU increases more than 40, the pixels near white decreases, showing more noise in the background.

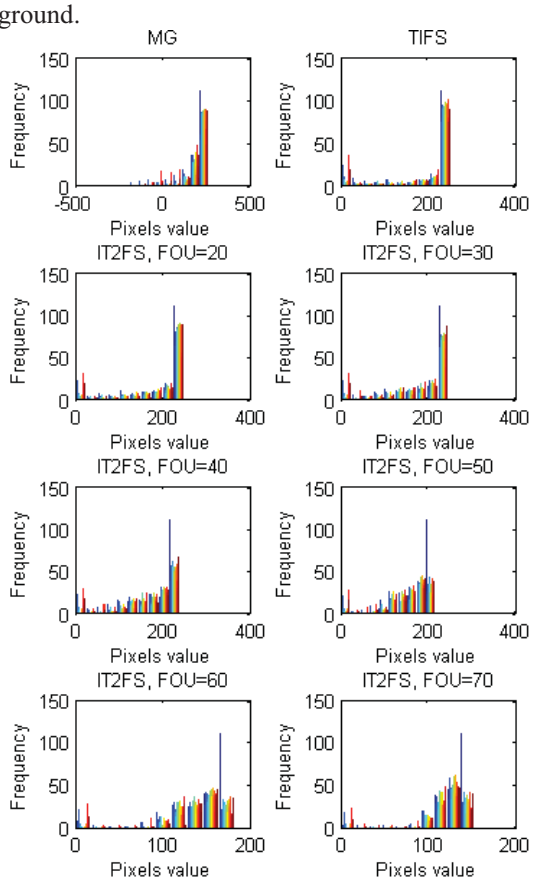


Figure 7: The frequency histograms correspond of the images shown in Fig. 6.

The second test, shown in Fig. 8, was made with an image used as benchmark for edge detection algorithms, and confirms the results obtained in Fig. 6.

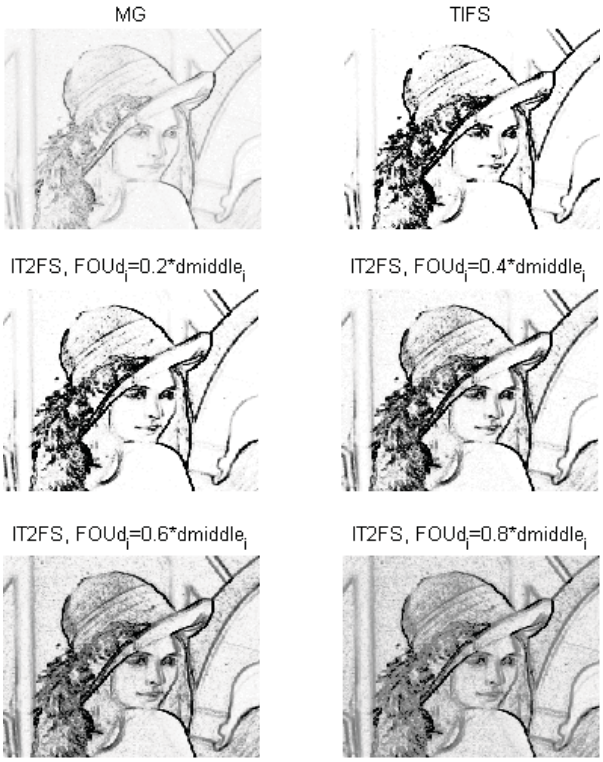


Figure 8: Edges achieved with FOU values obtained directly of the original image with equations (15) and (16).

ISBN: 978-989-95079-6-8

As we can see in Fig. 8 and Fig. 9, the edges can be seen better with the TIFIS method than de morphological gradient method, but more details of the image can be seen with the IT2FIS method when $FOUd_i=0.4*dmiddle_i$. Obviously the images obtained with a FOU near 0 are similar to the images obtained with the T1FIS. The image with $FOUd_i \geq 0.6*dmiddle_i$ show more details, but includes some textures that are not desired for edge detection.

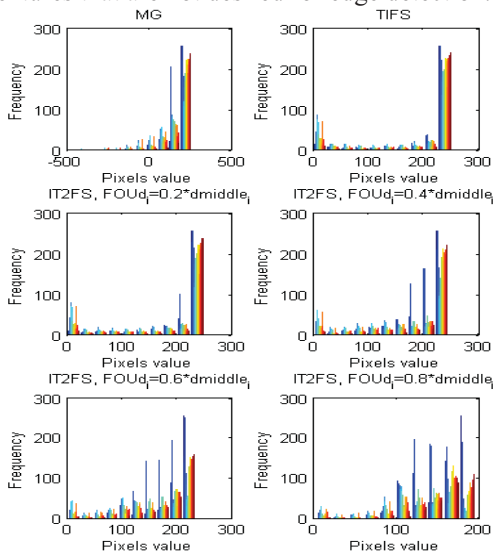


Figure 9: The frequency histograms correspond of the images shown in Fig. 8.

The parameters used for the T1FIS and IT2FIS in the tests of Fig. 8 and Fig. 9, were calculated with equations (6)-(16) and have the following values:

$dlow_1=0,\ dmiddle_1=105.9,\ dhigh_1=211.9,\ \sigma_1=26.4$
$dlow_1=0,\ dmiddle_2=109.6,\ dhigh_2=219.2,\ \sigma_2=27.4$
$dlow_1=0,\ dmiddle_3=107.5,\ dhigh_3=215.0,\ \sigma_3=26.8$
$dlow_1=0,\ dmiddle_4=74.6,\ dhigh_4=149.2,\ \sigma_4=18.6$
$\sigma e=31.8$

In Table I we show the numeric values of the FOUs used in the experiment below.

TABLE I: FOU values for the tests shown in Fig. 8.

FACTOR	$FOUd_1$	$FOUd_2$	$FOUd_3$	$FOUd_4$	$FOUE$
0.2	21.1	21.9	21.5	14.9	25.5
0.4	42.3	43.8	43.0	29.8	51.0
0.6	63.5	65.7	64.5	44.7	76.5
0.8	84.7	87.6	86.0	59.6	102



Figure 10: The original images used for the tests shown in Table II.

To finish the tests in this work, we repeat the same experiment with the images shown in Fig. 10, and compare the results with traditional algorithms [19] like the Canny and Sobel methods, as we can see in Table II.

TABLE II: Tests with Canny and Sobel methods using the Matlab defaults, and the best result using MG+IT2FIS.

Canny	Sobel	MG+IT2FIS
		

For all the tests we used the default parameters of Matlab. One of the images is a sample of the ORL database of faces

ISBN: 978-989-95079-6-8

[20], which is a set of images that are usually considered to compare the behavior of different image processing methods [21].
The other images are some of the typical images used for evaluation of edge detection algorithms, and were obtained them in the USC-SIPI Image Database [21].

In order to compare the results with the Canny and Sobel algorithms, we made modifications on the rules consequents, to obtain white edges and black background as the Matlab algorithms do it.

In all the tests we can observe that the images obtained with our method preserve more details of the original images than the traditional methods; that is good for the human observation. In the other hand, the images obtained with MG+IT2FIS are better to use in an image recognition application, because the training algorithm have more information to learn the image pattern. We can conclude that using interval type-2 fuzzy logic can improve edge detection in benchmark images, and for this reason it can be considered a good alternative for image processing.

5 Conclusions

The experiments presented in this paper show that the use of fuzzy inference systems can improve traditional methods for image processing. In particular, the morphological gradient method for edge detection, considered as one of the basic ones, when enhanced with an interval type-2 fuzzy inference system, results in a combination with satisfactory results. Especially the capability of the IT2FIS to model uncertainty in the morphological gradient and gray tone values for the edges achieved better images than the obtained with the T1FIS, because preserves more details of the original image.

References

[1] M. Heath, S. Sarkar, T. Sanocki, and K.W. Bowyer, 1997, A Robust Visual Method for Assessing the Relative Performance of Edge-Detection Algorithms, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 12, pp. 1338-1359.

[2] L. Hua, H.D. Cheng, M. Zhanga, 2007, A High Performance Edge Detector Based on Fuzzy Inference Rules, Information Sciences: An International Journal, Elsevier Science Inc. New York, NY, USA, Volume 177 , Issue 21, pp. 4768-4784.

[3] Y.-H. Kuo, C.-S. Lee, C.-C. Liu, 1997, A New Fuzzy Edge Detection Method for Image Enhancement , Proceedings of the Sixth IEEE International Conference on Fuzzy Systems, Spain , Vol. 2, Issue 1-5, pp. 1069-1074.

[4] K. Revathy, S. Lekshmi,,S. R. Prabhakaran Nayar, 2005, Fractal-Based Fuzzy Technique for Detection of Active Regions From Solar, Journal of Solar Physics Vol. 228, pp. 43-53.

[5] S. E. El-Khamy, M. Lotfy, and N. El-Yamany, 2000, A Modified Fuzzy Sobel Edge Detector, Radio Science Conference, 2000. 17th NRSC '2000. Seventeenth National, Minufiya, Egypt, pp. C32/1-C32/9.

[6] R. Wang, L.-Q. Gao, S. Yang and Y.-H. Chai, 2006, An Edge Detection Method by Combining Fuzzy Logic and Neural Network, Advances in Machine Learning and Cybernetics , Springer , Berlin, pp. 930-937

[7] T.Q. Pham and L.J. van Vliet, 2005, Blocking artifacts removal by a hybrid filter method, Proceedings of the Eleventh Annual Conference of the Advanced School for Computing and Imaging, E.U.A, p.p. 372-377.

[8] K. Suzuki, I. Horiba, N. Sugie, M. Nanki, 2003, Contour extraction of left ventricular cavity from digital subtraction angiograms using a neural edge detector, Systems and Computers Japan, Wiley Japan, p.p. 55-69

[9] H. Bustince, E. Berrenechea, M. Pagola, J.Fernandez, 2008, Interval-valued fuzzy sets constructed from matrices: Application to edge detection, Fuzzy Sets and Systems, Elsevier, available in: www.sciencedirect.com

[10] O. Mendoza, P. Melin, and G. Licea, 2007, Fuzzy Inference Systems Type-1 And Type-2 for Digital Images Edges Detection, Engineering Letters, International Association of Engineers, E.U.A., Volume 15 Issue 1 (Available on line: http://www.engineeringletters.com/issues_v15/issue_1/EL_15_1_7.pdf)

[11] O. Mendoza, P. Melin, G. Licea, 2007, A New Method for Edge Detection in Image Processing Using Interval Type-2 Fuzzy Logic, IEEE International Conference on Granular Computing (GRC 2007), Silicon Valley. Ca, U.S.A.

[12] O. Mendoza, P. Melin, G. Licea, 2008, A hybrid approach for image recognition combining type-2 fuzzy logic, modular neural networks and the Sugeno integral, Information Sciences, Elsevier, available in: www.sciencedirect.com

[13] O. Mendoza, P. Melin, O. Castillo, G. Licea, 2007, Type-2 Fuzzy Logic For Improving Training Data And Response Integration In Modular Neural Networks For Image Recognition, Foundations of Fuzzy Logic And Soft Computing (LNCC), Springer, Germany, pp. 604-612

[14] A. N. Evans, X.U., Liu, 2006, Morphological gradient operators for colour edges detection, IEEE Transactions on Image Processing, Vol. 15, Issue 6, pp. 1454-1463.

[15] Y. Becerikli, T. M. Karam, 2005, A New Fuzzy Approach for Edge Detection, in Computational Intelligence and Bioinspired Systems, LNCS, Springer, Berlin, pp. 943-951.

[16] F. Russo and G. Ramponi,1994, Edge extraction by FIRE operators Fuzzy Systems, IEEE World Congress on Computational Intelligence., pp, 249 – 253.

[17] J. Mendel, 2001, Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions, Prentice-Hall, U.S.A.

[18] J. R. Castro, O. Castillo, P. Melin, A. Rodriguez-Diaz, 2008, Building fuzzy inference systems with a new interval type-2 fuzzy logic toolbox: In Transactions on Computational Science Book Chapter, Vol. 4750, Springer, p.p. 104-114

[19] The USC-SIPI Image Database, available in http://sipi.usc.edu/database/

[20] O. Mendoza, P. Melin, 2007 The Fuzzy Sugeno Integral as a Decision Operator in the Recognition of Images with Modular Neural Networks, In: Hybrid Intelligent Systems Design and Analysis, Springer-Verlag, Germany, pp. 299-310.

[21] AT&T Laboratories Cambridge, The ORL database of faces, http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.

ISBN: 978-989-95079-6-8

Comparison of two versions of the Ferrers property of fuzzy interval orders

Susana Díaz[1] Bernard De Baets[2] Susana Montes[1]

1.Dept. Statistics and O. R., University of Oviedo
2.Dept. Appl. Math., Biometrics and Process Control; Ghent University
Email: {diazsusana,montes}@uniovi.es; bernard.debaets@ugent.be

Abstract— *We focus on the Ferrers property of fuzzy preference relations. We study the connection between the Ferrers property and fuzzy interval orders. A crisp total interval order is characterized by the Ferrers property of its strict preference relation. Also, a crisp preference structure is a total interval order if and only if its large preference relation satisfies the Ferrers property. For fuzzy relations the Ferrers property admits two non equivalent expressions. Here we compare both conditions by means of completeness. We also study if they characterize a fuzzy total interval order.*

Keywords— completeness, Ferrers property, partial interval order, total interval order, fuzzy relation

1 Introduction

Preference structures are the basis of the preference modelling theory. They formalize the answers of a decision maker over a set of alternatives. Formally, they are triplets of binary relations. For each pair of alternatives, any decision maker provides one (and only one) of the following answers: preference (by one of the alternatives), indifference (between both alternatives), inability to compare (the alternatives). These three situations lead to the three relations that make up a preference structure. Associated to a preference structure we find a relation called large preference relation that characterizes the preference structure from which it comes. Those four relations are the basis of the preference modelling ([9]).

Any preference structure can contain non-coherent relations if the decision maker gives conflicting answers. Different properties were formalized to model coherence. The Ferrers property is one of the most important ones. It has been proved to be a more realist condition than transitivity since the Ferrers property of the large preference relation does not imply the transitivity of the associated indifference relation (a non-realistic condition for some authors). However it does imply the transitivity of the strict preference relation. In a preference structure without incomparability, the large preference relation satisfies the Ferrers property if and only if the associated strict preference relation does.

A class of coherent preference structures are (total) interval orders: the alternatives can be identified with intervals not necessarily of the same length on the real line. A crisp preference structure without incomparabilities is a (total) interval order if and only if its associated large preference relation or its strict preference relation satisfy the Ferrers property.

Since several years ago, crisp relations are considered not flexible enough to model human decisions. In order to model real life situations in a more accurate way, fuzzy relations were introduced [10] and they have received a wide attention (see, e.g., [2, 3, 4, 6]). The notion of fuzzy preference structure was a topic of debate for several years. In this work we present a similar study for the notion of fuzzy interval order. This notion admits many different equivalent definitions in the classical context. We focus on the characterizations based on the Ferrers property. We first discuss the Ferrers property for fuzzy relations. This property admits two different ways to be formalized in our wider context. We compare them and we study their connection to fuzzy interval orders. In this context, a t-norm is needed to define fuzzy interval orders and the Ferrers property. Although we aim at considering any operator in our study, we prove that only two particular families have an appropriate behavior. Thus, we pay special attention to those operators. Since interval order is an important concept for the analysis of temporal events, each of which occurs over some time span (for instance, the time spans over which animal species are found or the occurrence of styles of pottery in archaeological strata), this kind of events in a fuzzy environment could be a potential application area for this study.

This work is organized as follows. In Section 2 we recall crisp relations. In Section 3 we formalize the two versions of the Ferrers property for fuzzy relations and we study their relationship. Section 4 is devoted to the completeness property. For crisp relations every reflexive Ferrers relation is complete. We study the implication for fuzzy relations. Section 5 contains the results on the connection between the first type of Ferrers property and fuzzy interval orders. Section 6 includes an analogous study for the second version of the Ferrers property. Section 7 contains some conclusions.

2 Crisp relations

2.1 Crisp preference structures

Let us consider a set of alternatives A whose elements we want to compare. In preference modeling the comparison is always carried out by pairs. When a decision maker consider two of those alternatives, the answer is always among the following ones: preference for one of the alternatives, indifference among both possibilities, inability to compare them. In the classical setting, each one of the three previous situations is identified with an $A \times A \to \{0,1\}$ application or relation. The first case is identified with the *strict preference relation*, denoted by P. We state $P(a,b) = 1$ if the decision maker prefers a to b and 0 otherwise. The second case is identified with the *indifference relation* denoted I. We assign $I(a,b) = 1$ if a and b are indifferent and $I(a,b) = 0$ if not. The *incomparability relation* is denoted J. If the alternatives a and b cannot be compared by the decision maker, $J(a,b) = 1$. Otherwise, $J(a,b) = 0$. P, I and J make up a preference structure. Let A^2 be the set of ordered pairs

ISBN: 978-989-95079-6-8

((a,b) and (b,a) are different pairs) of alternatives in A. The converse or transpose of Q is defined as $Q^t(a,b) = Q(b,a)$, its complement as $Q^c(a,b) = 1 - Q(a,b)$ and its dual as $Q^d = (Q^t)^c$. Every relation is also identified with a subset of A^2: $Q(a,b) = 1 \Leftrightarrow (a,b) \in Q$. One easily verifies that if we think of P, I and J as subsets of A^2, the quadruplet (P, P^t, I, J) establishes a particular partition of A^2. Formally [9]:

Definition 1 *A preference structure on A is a triplet (P, I, J) of relations on A that satisfy:*

(i) *P is irreflexive, I is reflexive and J is irreflexive;*

(ii) *P is asymmetrical, I and J are symmetrical;*

(iii) *$P \cap I = \emptyset$, $P \cap J = \emptyset$ and $I \cap J = \emptyset$;*

(iv) *$P \cup P^t \cup I \cup J = A^2$.*

Every preference structure is identified with a unique reflexive relation called *large preference relation* $R = P \cup I$. It connects a to b if the decision maker considers a preferred or indifferent to b. This relation leads back to the preference structure in the following way:

$$(P, I, J) = (R \cap R^d, R \cap R^t, R^c \cap R^d). \qquad (1)$$

We say that a relation R is *complete* if aRb or bRa for all a, b in the set of alternatives A. In [8] it was proven that R is complete if and only if the associated J does not connect any pair of alternatives (a,b),

$$R \text{ complete} \quad \Leftrightarrow \quad J = \emptyset.$$

A relation R is a Ferrers relation (see among others [8, 9]) if for any four alternatives a, b, c, d in A, it holds that if aRb and cRd then aRd or cRb. Another way of expressing this property is the following one: if aRb and cRd and $c \not R b$ then aRd. This property is also called biorder by some authors.

Let us remark also that the Ferrers and the completeness properties are connected.

Lemma 1 *Every reflexive Ferrers relation is complete.*

Let us denote the composition of two binary relations Q_1 and Q_2 by $Q_1 \circ Q_2$. That is, $a(Q_1 \circ Q_2)b$ if and only if there exists c such that $aQ_1c \wedge cQ_2b$. Then the equivalent compositional definition of Ferrers property is $R \circ R^d \circ R \subseteq R$.

2.2 Crisp interval orders

Interval orders are one of the most important types of "coherent" preference structures. Formally, we distinguish between partial and total interval order, depending on the possibility or impossibility of incomparable elements.

Definition 2 *A* partial interval order *is a preference structure (P, I, J) such that $P \circ I \circ P \subseteq P$. In the particular case $J = \emptyset$, (P, I, J) is said to be a* total interval order.

As commented in the introduction, this original definition admits different ways of being expressed. The following properties (among others) were shown to be equivalent in [8]:

Proposition 1 *Let $(P, I, \emptyset)$ be a preference structure without incomparability and R its large preference relation. The following conditions are equivalent.*

a) *(P, I, J) is an interval order.*

b) *P satisfies the Ferrers property.*

c) *R satisfies the Ferrers property.*

These definitions are equivalent only if the decision maker is able to compare all the pairs of alternatives, i. e., if $J = \emptyset$, equivalently if R is complete. If we do not impose this condition, we can still consider the previous properties and study their relationship. In what remains of this section we carry out this study.

It is very easy to prove that whenever R satisfies the Ferrers property, P satisfies the Ferrers property and $P \circ I \circ P \subseteq P$. It also holds that for any preference structure (P, I, J), if P satisfies the Ferrers property then $P \circ I \circ P \subseteq P$, but the converse implications do not hold.

The following proposition summarizes the relationship among the conditions of Proposition 1 when we consider any preference structure.

Proposition 2 *Let (P, I, J) be a (crisp) preference structure and let R be its large preference relation. Then,*

i) *(P, I, J) is a total interval order and R satisfies the Ferrers property are equivalent properties.*

ii) *(P, I, J) is a total interval order implies that P satisfies the Ferrers property. In addition to this, if P satisfies the Ferrers property then $P \circ I \circ P \subseteq P$. The converse implications to the ones showed in this item do not hold.*

Propositions 1 and 2 can be outlined as we show in Figure 1, where the missing implications do not hold.

$$\begin{array}{ccc} (P,I,J) \text{ total int. ord.} & \Leftrightarrow & R \text{ Ferrers} \\ & & \Downarrow \ \Updownarrow (J = \emptyset) \\ & & P \text{ Ferrers} \\ & & \Downarrow \ \Updownarrow (J = \emptyset) \\ & & (P,I,J) \text{ partial int. ord.} \end{array}$$

Figure 1: Connection among the characterizations of a crisp interval order.

3 Fuzzy preference structures

All the relations involved in a (crisp) preference structure are crisp. Then they are not always appropriate to model human decisions. They do not include intensity of preference. The lack of flexibility in crisp set theory led to introduce fuzzy relations. They allow the decision maker to encode degrees of preference rather than just its absence or presence. Fuzzy relations were introduced as a natural extension of the concept of crisp binary relation. They have been widely studied. For a complete review about fuzzy relations see [6].

Formally, a fuzzy relation is defined as a function Q from the cartesian product $A \times A$ into the interval $[0, 1]$. For every pair $(a,b) \in A \times A$, the value $Q(a,b)$ shows the degree of truth of the fact that aQb in the crisp sense. The complementary, transpose (or converse) and dual of the binary relation Q are defined for any $(a,b) \in A \times A$ as $Q^c(a,b) = 1 - Q(a,b)$, $Q^t(a,b) = Q(b,a)$ and $Q^d(a,b) = 1 - Q(b,a)$.

ISBN: 978-989-95079-6-8

The intersection of fuzzy relations is usually defined pointwisely based on some *t-norm*. We recall that the binary operator $T : [0,1] \times [0,1] \to [0,1]$ is a *triangular norm* or *t-norm* for short if it is commutative, associative, monotone and has 1 as neutral element. The three most important t-norms are the minimum operator $T_{\mathbf{M}}(x,y) = \min(x,y)$, the algebraic product $T_{\mathbf{P}}(x,y) = xy$ and the Łukasiewicz t-norm $T_{\mathbf{L}}(x,y) = \max(x+y-1,0)$. The minimum operator is the greatest t-norm; the smallest t-norm is the drastic product $T_{\mathbf{D}}$ defined by

$$T_{\mathbf{D}}(x,y) = \begin{cases} 0 & \text{if } \max(x,y) < 1\,, \\ \min(x,y) & \text{otherwise}\,. \end{cases}$$

On the other hand, we say that a value $x \in (0,1)$ is a zero-divisor of a t-norm T if there exists a value $y \in (0,1)$ such that $T(x,y) = 0$. The minimum and the product are t-norms that do not admit zero-divisors, while $T_{\mathbf{L}}$ and $T_{\mathbf{D}}$ do admit them.
An important family of t-norms that admit zero-divisors are the rotation invariant t-norms. A *rotation invariant t-norm* is a t-norm T that verifies:

$$T(x,y) \le z \;\Leftrightarrow\; T(x,1-z) \le 1-y, \quad \forall x,y,z \in [0,1].$$

The rotation invariance property for T means that the part of the space $[0,1]^3$ which is exactly below the graph of T remains invariant under an order three transformation. This transformation is indeed a rotation of $[0,1]^3$ with angle $2\pi/3$ around the axis that passes through the points $(0,0,1)$ and $(1,1,0)$. Rotation invariant t-norms satisfy in particular that $T(x,y) > 0 \Leftrightarrow x + y > 1$, i.e. the lower-left triangle of the unit square constitutes the zero-divisors of T. This excludes t-norms without zero-divisors (such as $T_{\mathbf{P}}$), nilpotent t-norms other than $T_{\mathbf{L}}$ and the drastic product $T_{\mathbf{D}}$.
As extreme members of the family we find the Łukasiewicz t-norm $T_{\mathbf{L}}$ and the nilpotent minimum $T_{\mathbf{nM}}$. This t-norm is defined as:

$$T_{\mathbf{nM}}(x,y) = \begin{cases} 0 & \text{if } x+y \le 1\,, \\ \min(x,y) & \text{otherwise}\,. \end{cases}$$

Similarly to the intersection, the union of fuzzy relations is based on a *t-conorm*, *i.e.* an increasing, commutative and associative binary operation on $[0,1]$ with neutral element 0. t-norms and t-conorms come in dual pairs: to any t-norm T there corresponds a t-conorm S through the relationship

$$S(x,y) = 1 - T(1-x, 1-y).$$

For the main three t-norms this yields the maximum operator $S_{\mathbf{M}}(x,y) = \max(x,y)$, the probabilistic sum $S_{\mathbf{P}}(x,y) = x + y - xy$ and the Łukasiewicz t-conorm (bounded sum) $S_{\mathbf{L}}(x,y) = \min(x+y,1)$. For more information on t-norms and t-conorms, we refer to [7].

The definition of a fuzzy preference structure has been discussed for many years [6]. The *assignment principle*, expressing that for any pair of alternatives (a,b) the decision maker is allowed to assign at least one of the degrees $P(a,b)$, $P(b,a)$, $I(a,b)$ and $J(a,b)$ freely in the unit interval, leads to a graduation of Definition 1 with intersection based on the Łukasiewicz t-norm and union based on the Łukasiewicz t-conorm. This definition admits the same short formulation as the classical one: a triplet (P,I,J) of fuzzy relations on A is an *additive fuzzy preference structure* on A if and only if I is reflexive ($I(a,a) = 1$ for any $a \in A$) and symmetrical, and for any $(a,b) \in A^2$ it holds that $P(a,b) + P^t(a,b) + I(a,b) + J(a,b) = 1$. This expression justifies the adjective *additive*. Note that P is irreflexive, and that J is irreflexive and symmetrical.

Another difficult point has been how to construct such a structure from a reflexive fuzzy relation. The most recent and most successful approach is that of De Baets and Fodor based on (indifference) *generators* [1].

Definition 3 *A generator i is a commutative $[0,1]^2 \to [0,1]$ mapping that satisfies $T_{\mathbf{L}} \le i \le T_{\mathbf{M}}$.*

Note that a generator always has neutral element 1. For any reflexive fuzzy relation R on A it holds that the triplet (P,I,J) of fuzzy relations on A defined by:

$$\begin{aligned} P(a,b) &= R(a,b) - i(R(a,b),R(b,a))\,, \\ I(a,b) &= i(R(a,b),R(b,a))\,, \\ J(a,b) &= i(R(a,b),R(b,a)) - (R(a,b)+R(b,a)-1)\,, \end{aligned}$$

is an additive fuzzy preference structure on A such that R is the union by the Łukasiewicz t-conorm of P and I, this is, $R(a,b) = P(a,b) + I(a,b)$. The fuzzy relation R is again called the *large preference relation*.
A family of generators that present important properties is the family of Frank t-norms (see, e.g., [6]). In the following section we will see that a particular Frank t-norm is a specially good generator: the Łukasiewicz t-norm.

Definition 4 *Given a t-norm T and a t-conorm S, a fuzzy relation Q is a type 1 T-S-*Ferrers *relation if*

$$T(Q(a,b),Q(c,d)) \le S(Q(a,d),Q(c,b)),\ \forall\, a,b,c,d, \in A.$$

The t-norm T and the t-conorm S are not necessarily connected. However, given T, the most usual t-conorm considered is the dual of T. In this case, the t-conorm is not specified and we just talk about type 1 T-Ferrers property. Along this work, this will be the setting: we will always consider as t-conorm the dual of the fixed t-norm.

Definition 5 *Given a t-norm T, a fuzzy relation Q is a* type 2 T-Ferrers *relation if*

$$T(Q(a,b),Q^d(b,c),Q(c,d)) \le Q(a,d),\ \forall\, a,b,c,d \in A.$$

The composition of two fuzzy relations Q_1 and Q_2 defined on A also involves t-norms. Given a t-norm T, the T-composition of Q_1 and Q_2 is the fuzzy relation $Q_1 \circ_T Q_2$ defined as follows

$$(Q_1 \circ_T Q_2)(a,b) = \sup\{T(Q_1(a,c),Q_2(c,b)) | c \in A\}.$$

With this notation we can also say that a fuzzy relation Q is a type 2 T-Ferrers relation if it holds that $Q \circ_T Q^d \circ_T Q \subseteq Q$.

Definition 6 [2] *An additive fuzzy preference structure is a T-*partial interval order *if it holds that*

$$P \circ_T I \circ_T P \subseteq P\,,$$

*or, equivalently, $T(T(P(a,b),I(b,c)),P(c,d)) \le P(a,d)$, $\forall\, a,b,c,d \in A$. It is a T-*total interval order *if it is a T-partial interval order and $J = \emptyset$.*

ISBN: 978-989-95079-6-8

Clearly, if $T_1 \leq T_2$, every type 1 T_2-Ferrers relation is also a type 1 T_1-Ferrers relation, every type 2 T_2-Ferrers relation is also a type 2 T_1-Ferrers relation and every T_2-interval order is also a T_1-interval order.

The equivalence between the notions of type 1 and type 2 T-Ferrers relations has already been studied in some particular cases ([5]). The following theorem further generalizes these results and characterizes all t-norms for which both notions coincide.

Theorem 1 *Consider a t-norm T. Then the following statements are equivalent:*

(i) *Any type 1 T-Ferrers relation is type 2 T-Ferrers.*

(ii) *Any type 2 T-Ferrers relation is type 1 T-Ferrers.*

(iii) *T is rotation-invariant.*

It follows from this theorem that in particular, for any reflexive relation R and for any strict preference relation P, the equivalence between the type 1 T-Ferrers property and the type 2 T-Ferrers property holds if and only if T is a rotation invariant t-norm. This justifies the importance we give to this family of t-norms in this work.

The theorem above does not apply to the minimum t-norm. However it is an important operator and the greatest t-norm. What can we say about it? Since $T_{\mathbf{nM}} \leq T_{\mathbf{M}}$, it follows immediately that any type 1 $T_{\mathbf{M}}$-Ferrers relation is a type 2 $T_{\mathbf{nM}}$-Ferrers relation and that any type 2 $T_{\mathbf{M}}$-Ferrers relation satisfies the type 1 $T_{\mathbf{nM}}$-Ferrers property. This result can be strengthened considerably.

Proposition 3

(i) *$T_{\mathbf{nM}}$ is the largest t-norm T such that any type 1 $T_{\mathbf{M}}$-Ferrers relation is also type 2 T-Ferrers.*

(ii) *$T_{\mathbf{nM}}$ is the largest t-norm T such that any type 2 $T_{\mathbf{M}}$-Ferrers relation is also type 1 T-Ferrers.*

4 Completeness of the large preference relation

We have seen that for crisp relations the completeness of the large preference relation follows from the Ferrers property. In this section we consider both types of the Ferrers property. We investigate the kind of completeness exhibited by a fuzzy reflexive relation satisfying one of the two properties.

Definition 7 *Let us consider a t-conorm S and a fuzzy relation Q defined on a set of alternatives A. Q is S-complete if it satisfies*

$$S(Q(a,b), Q(b,a)) = 1, \qquad \forall\, a, b \in A.$$

The two most important notions are based on the minimum and the Łukasiewicz t-norms. Given a relation Q we say that it is

- *Strongly complete:* if $Q(a,b) = 1$ or $Q(b,a) = 1, \forall a, b \in A$.
- *weakly complete:* if $Q(a,b) + Q(b,a) \geq 1, \forall a, b \in A$.

The weakest S-completeness condition is based on $S_{\mathbf{D}}$. Given Q, it is $S_{\mathbf{D}}$-complete if $\max(Q(a,b), Q(b,a)) > 0$ for all a, b in A.

No one of the "fuzzy notions" of completeness of R is equivalent to the condition $J = \emptyset$, but we can establish the following equivalence.

Lemma 2 [4] *Let R be a reflexive fuzzy relation and let J be the incomparability relation associated to R by means of any generator i. Then the following equivalence holds*

$$J = \emptyset \Leftrightarrow \begin{cases} R \text{ is weakly complete} \\ i|_S = T_{\mathbf{L}} \end{cases}$$

where $S = \{(u,v) \in [0,1]^2 : \exists (x,y) \in A^2$ with $R(x,y) = u, R(y,x) = v\}$.

Let us recall that when $i = T_{\mathbf{L}}$, the fuzzy preference structure obtained from a weakly complete reflexive relation R is the triplet $(P_{\mathbf{L}}, I_{\mathbf{L}}, J_{\mathbf{L}}) = (R^d, \min(R, R^t), \emptyset)$. We would like to stress the fact that the strict preference relation becomes the dual of the large preference relation.

We next focus on the completeness assured for reflexive type 1 and type 2 T-Ferrers relations.

Proposition 4 *Let T be a t-norm and S its dual t-conorm. Every type 1 T-Ferrers reflexive relation is S-complete.*

As an immediate consequence we get the following: Every type 1 T-Ferrers reflexive relation satisfies the weakest kind of completeness ($S_{\mathbf{D}}$-completeness).

Since strong and weak completeness are in fact $S_{\mathbf{M}}$-completeness and $S_{\mathbf{L}}$-completeness, it also follows from Proposition 4 that every type 1 $T_{\mathbf{L}}$-Ferrers relation is weakly complete and every type 1 $T_{\mathbf{M}}$-Ferrers relation is strongly complete. We can say more.

Theorem 2 *Consider a t-norm T. Every reflexive type 1 T-Ferrers relation is weakly complete if and only if $T(x,y) > 0$ for any pair $(x,y) \in [0,1]^2$ such that $x + y > 1$.*

This theorem applies to all rotation-invariant t-norms. In particular, it applies to $T_{\mathbf{L}}$, and hence to any $T \geq T_{\mathbf{L}}$. Theorem 2 also applies to t-norms without zero-divisors, although in that case the following much stronger result holds.

Theorem 3 *Any reflexive type 1 T-Ferrers relation is strongly complete if and only if the t-norm T has no zero-divisors.*

This theorem applies in particular to $T_{\mathbf{M}}$.

Next we repeat the same study for reflexive type 2 T-Ferrers relations.

Theorem 4 *Any reflexive type 2 T-Ferrers relation is weakly complete.*

A first comparison between types 1 and 2 of the T-Ferrers property is in favour of the type 2 definition as the previous theorem holds for any t-norm T. However, our enthusiasm is tempered by the following negative result.

Proposition 5 *Not all reflexive type 2 $T_{\mathbf{M}}$-Ferrers relations are strongly complete.*

ISBN: 978-989-95079-6-8

Despite this negative result, t-norms without zero-divisors have good properties.

Theorem 5 *Let $(P, I, \emptyset)$ be an additive fuzzy preference structure without incomparability and let R be its associated large preference relation. Let T be a t-norm. The following assertions are equivalent.*
(i) *If $(P, I, \emptyset)$ is a T-total interval order, then R is strongly complete.*
(ii) *If P is type 1 T-Ferrers, then R is strongly complete.*
(iii) *If P is type 2 T-Ferrers, then R is strongly complete.*
(iv) *If R is type 1 T-Ferrers, then R is strongly complete.*
(v) *T does not admit zero-divisors.*

The results presented in this section show that there are two important families of t-norms in the study of the Ferrers property: rotation invariant t-norms and t-norms without zero-divisors. In the two following sections we pay special attention to those two families [8].

5 Fuzzy interval orders and the type 1 Ferrers property

In this section we study the implications of Figure 1 that remain true for fuzzy relations when the Ferrers property of a crisp relation is generalized by means of the type 1 Ferrers property.

Lemma 3 [2] *Not every $T_{\mathbf{M}}$-total interval order verifies that its strict preference relation is type 1 $T_{\mathbf{M}}$-Ferrers. Not every $T_{\mathbf{L}}$-total interval order verifies that its strict preference relation is type 1 $T_{\mathbf{L}}$-Ferrers.*

Therefore, we cannot assure that for every total interval order its strict preference relation satisfies the type 1 Ferrers property.

Let us focus on the connection between the large and the strict preference relations.

Let us recall first of all that the results we present in this contribution concern a fixed generator $i = T_{\mathbf{L}}$, as it has shown an appropriate behavior (see Lemma 2).

Theorem 6 *The following statements are equivalent:*

(i) The strict preference relation obtained from a type 1 T-Ferrers reflexive relation also satisfies the type 1 T-Ferrers property.

(ii) The t-norm T satisfies $T(x, y) > 0$ for any pair $(x, y) \in [0, 1]^2$ such that $x + y > 1$.

In particular, the previous result can be applied to rotation-invariant t-norms and t-norms without zero-divisors.

The result presented in Theorem 6 can be strengthened under completeness conditions.

Theorem 7 *Consider a fuzzy reflexive relation R with corresponding fuzzy preference structure $(P_{\mathbf{L}}, I_{\mathbf{L}}, J_{\mathbf{L}})$ generated by means of $i = T_{\mathbf{L}}$. For any t-norm T, the following statements are equivalent:*

- *R weakly complete and type 1 T-Ferrers*
- *$J_{\mathbf{L}} = \emptyset$ and $P_{\mathbf{L}}$ type 1 T-Ferrers*

Proposition 6 *Let (P, I, J) be an additive fuzzy preference structure and let T be a rotation invariant t-norm. If P is a type 1 T-Ferrers relation, then (P, I, J) is a T-partial interval order.*

The previous implication cannot be extended to t-norms without zero-divisors. In [2] it was proven that not every type 1 $T_{\mathbf{M}}$-Ferrers reflexive relation satisfies that the fuzzy preference structure obtained from it (by the Łukasiewicz t-norm) is a $T_{\mathbf{M}}$-total interval order. It follows from here and Theorem 6 that the implication of Proposition 6 does not hold for the family of t-norms without zero-divisors.

Now we focus on the connection between a total interval order and the type 1 T-Ferrers property of the associated large preference relation.

Corollary 1 *Let T denote the minimum or the Łukasiewicz t-norms. Not every T-total interval order verifies that its associated large preference relation satisfies the type 1 T-Ferrers property.*

The converse implication neither holds for $T_{\mathbf{M}}$ [2]. However the implication does hold for rotation invariant t-norms.

Proposition 7 *Let T be a rotation invariant t-norm. The additive fuzzy preference structure obtained (by the Łukasiewicz t-norm) from a type 1 T-Ferrers reflexive relation is a T-total interval order.*

In Figures 2 and 3 we summarize the results obtained for the type 1 Ferrers property both for the family of rotation invariant t-norms and for the family of t-norms without zero divisors. We have proven that the implications missing do not hold.

$$\begin{array}{ccccc} (P,I,J) & T\text{-total int. ord.} & \Leftarrow & R & \text{type 1 } T\text{-Ferrers} \\ & (J=\emptyset)\nwarrow & & \not\swarrow \quad \not\nearrow & (J=\emptyset) \\ & & & P & \text{type 1 } T\text{-Ferrers} \\ & & & \Downarrow & \\ & & & (P,I,J) & T\text{-partial int. ord.} \end{array}$$

Figure 2: Connection among the type 1 T-Ferrers property and the notion of interval order for T a rotation invariant t-norm.

$$\begin{array}{ccccc} (P,I,J) & T\text{-total int. ord.} & & R & \text{type 1 } T\text{-Ferrers} \\ & (J=\emptyset) & & \not\swarrow \quad \not\nearrow & (J=\emptyset) \\ & & & P & \text{type 1 } T\text{-Ferrers} \\ & & & & \\ & & & (P,I,J) & T\text{-partial int. ord.} \end{array}$$

Figure 3: Connection among the type 1 T-Ferrers property and the notion of interval order for T a t-norm without zero-divisors.

6 Fuzzy interval orders and the type 2 Ferrers property

We already know the behaviour of the type 1 Ferrers property when dealing with interval orders. Now we focus on the type 2 Ferrers property.

ISBN: 978-989-95079-6-8

Proposition 8 *Let T be a t-norm without zero-divisors and let $(P, I, \emptyset)$ be a T-total interval order, then P satisfies the type 2 T-Ferrers property.*

However the implication cannot be extended to any t-norm.

Proposition 9 *Let T be a t-norm such that $T(x, 1-x) = 0$ for some $x \in (0,1)$. Not every T-total interval order satisfies that its strict preference relation is type 2 T-Ferrers.*

A particular type of the t-norms described in the previous proposition are the rotation invariant t-norms.

Next we study how the type 2 Ferrers property propagates between R and P.

Theorem 8 *Consider a t-norm T. Then the following statements are equivalent:*

(i) *For any reflexive relation R, if R satisfies the type 2 T-Ferrers property, then $J_{\mathbf{L}} = \emptyset$ and $P_{\mathbf{L}}$ satisfies the type 2 T-Ferrers property.*

(ii) *For any reflexive relation R, if $J_{\mathbf{L}} = \emptyset$ and $P_{\mathbf{L}}$ satisfies the type 2 T-Ferrers property, then R satisfies the type 2 T-Ferrers property.*

(iii) *T is rotation-invariant.*

We now study the relationship between the type 2 T-Ferrers property of P and a T-partial interval order.

Proposition 10 *Let (P, I, J) be an additive fuzzy preference structure and let T be a t-norm. If P is type 2 T-Ferrers then (P, I, J) is a T-partial interval order.*

It follows from the crisp case that the converse implication does not hold for any t-norm.

The characterization of total interval orders based on the Ferrers property of the associated large preference relation gets lost for fuzzy relations when considering the type 2 T-Ferrers condition.

Proposition 11 *It does not hold that for every $T_{\mathbf{M}}$-total interval order the associated large preference relation is type 2 $T_{\mathbf{M}}$-Ferrers. It neither holds that for any type 2 $T_{\mathbf{M}}$-Ferrers reflexive relation the associated by the Łukasiewicz t-norm preference structure is a $T_{\mathbf{M}}$-total interval order.*

Proposition 12 *For any type 2 $T_{\mathbf{L}}$-Ferrers reflexive relation R the associated fuzzy preference structure $(P_{\mathbf{L}}, I_{\mathbf{L}}, J_{\mathbf{L}})$ is a $T_{\mathbf{L}}$-total interval order.*
The converse implication cannot be assured for any type 2 $T_{\mathbf{L}}$-total interval order.

We draw in Figures 4 and 5 the results obtained for the type 2 Ferrers property.

7 Conclusions

We have studied the connection between the fuzzy Ferrers property and the definition of fuzzy interval orders. Since the Ferrers property admits two different ways to be formalized for fuzzy relations, we have compared those two versions by its connection to fuzzy interval orders. We have seen that we cannot identify a best definition since each property behaves better in different cases. We have seen that both properties are equivalent only when dealing with rotation invariant t-norms. For these operators, Ferrers property is a stronger condition than the notion of total interval order. For t-norms without zero-divisors, no one of them is good. The type 2 condition seems to assure the equivalence between total interval orders and the Ferrers property of the associated strict preference relation. However, type 1 behaves better if we want to propagate the Ferrers property between the large and the strict preference relations.

$$\begin{array}{ccc} (P,I,J)\ T\text{-total int. ord.} & \Leftarrow & R \text{ type 2 } T\text{-Ferrers} \\ (J=\emptyset)\nwarrow & & \swarrow\ \nearrow (J=\emptyset) \\ & P \text{ type 2 } T\text{-Ferrers} & \\ & \Downarrow & \\ & (P,I,J)\quad T\text{-partial int. ord.} & \end{array}$$

Figure 4: Connection between the type 2 T-Ferrers property and interval orders for T a rotation invariant t-norm.

$$\begin{array}{ccc} (P,I,J)\ T\text{-total int. ord.} & & R \text{ type 2 } T\text{-Ferrers} \\ (J=\emptyset)\swarrow\ \searrow & & (J=\emptyset) \\ & P \text{ type 2 } T\text{-Ferrers} & \\ & \Downarrow & \\ & (P,I,J)\quad T\text{-partial int. ord.} & \end{array}$$

Figure 5: Connection between the type 2 T-Ferrers property and interval orders for T a t-norm without zero-divisors.

Acknowledgment

The research reported on in this paper has been partially supported by Project FEDER-MEC-MTM2007-61193.

References

[1] B. De Baets and J. Fodor. Generator triplets of additive fuzzy preference structures. In *Proc. RelMiCS6*.

[2] B. De Baets and B. Van de Walle. Weak and strong fuzzy interval orders. *Fuzzy Sets and Systems*, 79:213–225, 1996.

[3] S. Díaz, B. De Baets, and S. Montes. Additive decomposition of fuzzy pre-orders. *Fuzzy Sets and Systems*, 158:830–842, 2007.

[4] S. Díaz, S. Montes, and B. De Baets. Transitivity bounds in additive fuzzy preference structures. *IEEE Transactions on fuzzy systems*, 15(2):275–286, 2007.

[5] J. Fodor. Traces of binary fuzzy relations. *Fuzzy Sets and Systems*, 50:331–341, 1992.

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

[7] E. P. Klement, R. Mesiar, and E. Pap. *Triangular Norms*. Kluwer Academic Publishers, Dordrecht, 2000.

[8] B. Monjardet. Axiomatiques et propriétés des quasi-ordres. *Math. Sci. Hum.*, 63:51–82, 1978.

[9] M. Roubens and Ph. Vincke. Preference modelling. In *Lecture Notes in Economics and Mathematical Systems*, volume 250. Springer, Berlin, 1985.

[10] L. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

ISBN: 978-989-95079-6-8

Decomposition of the transitivity for additive fuzzy preference structures

Susana Díaz, Susana Montes[1] Bernard De Baets[2]

1.Dept. of Statistics and O.R., University of Oviedo
Oviedo, Spain
2.Dept. of Applied Mathematics, Biometrics and Process Control, Ghent University
Gent, Belgium
Email: {diazsusana,montes}@uniovi.es, Bernard.DeBaets@UGent.be

Abstract— *Transitivity is a very important property in order to provide coherence to a preference relation. Usually, t-norms are considered to define the transitivity of fuzzy relations. In this paper we deal with conjunctors, a wider family than t-norm, to define the transitivity. This more general definition allows to impove the results found in the literature. We characterize the behaviour with respect to transitivity of the strict preference and indifference relations of any fuzzy preference structure associated to any large preference relation. Those two characterizations provide very general expressions. We also obtain easier expressions in some particular cases.*

Keywords— Fuzzy preference relation, transitivity, conjunctor, indifference generator.

1 Introduction

Decision making is present in many situations in life. Preference models are an essential part of the design phase of a decision making process. The departure point in preference modelling is the comparison by pairs of possible alternatives. If this point of the process lacks of coherence, the whole process makes no sense. One of the most important property introduced to ensure coherence in the two-by-two comparison is transitivity. In this paper we focus on this property.

Given its relevance in preference modelling, transitivity has been widely studied (see [3, 13, 14], among others). In classical or crisp preference modelling there exists a basic relation R, called *large preference relation* that allows to compare a pair of possibilities: given the alternatives a and b, $(a, b) \in R$ expresses that alternative a is considered to be at least as good as alternative b. From R, three relations can be defined: the strict preference relation P, the indifference relation I and the incomparability relation J. It is well known that the transitivity of R is completely characterized by the transitivity of I and P and two additional relational inequalities involving I and P; in case of completeness, the transitivity of R is only characterized by the transitivity of I and P [2].

All those relations are crisp and therefore, they are not always appropriate to model human decisions. The lack of flexibility in crisp set theory lead to introduce fuzzy sets in preference modeling and to study the concept of fuzzy preference structure [8, 18, 24]; see [6] for a historical account of its development. Transitivity is traditionally defined for fuzzy relations by means of t-norms. In this context, the transitivity of the (fuzzy) large preference relation R has been characterized in a similar way as for crisp relations, in a particular case: when R is strongly complete [12]. But strong completeness is a quite restrictive condition. Subsequent works have treated more general preference relations [3, 4]. Those authors center their study on the translation of a fixed transitivity from a more general (fuzzy) large preference relation R to different associated indifference and strict preference relations (I and P respectively). We have continued along this line but we have treated the problem from a totally different viewpoint. On the one hand, we work with conjunctors, a wider family of operators than t-norms, so we handle a very general notion of transitivity. On the other hand, we do not study the preservation of the transitivity when decomposing R, but we explore the strongest transitivity we can assure I and P satisfy. In a previous work [15], we have fenced in the transitivity of I and P according to the transitivity R satisfies. In this work we go further: we provide an explicit expression for the transitivity that any I (respectively any P) satisfies. We prove that no stronger transitivity is fulfilled by all I (respectively for all P). Transitivity is a fundamental assumption in some decision-making models, like rational model (founded in quantitative disciplines like economics, mathematics and statistics) or political model (primarily based on the disciplines of political science, philosophy, psychology and sociology). Without this property, it is not possible to obtain a coherent order of the alternatives and therefore the choice of the best alternative is more complicate or even impossible. Thus, the study made in this work is a first step to develop a consistent theory to order the alternatives and therefore to provide a decision, when the comparison between alternatives can be graded.

The work is structured in six sections. In Section 2 we recall the most relevant concepts concerning additive fuzzy preference structures. In Section 3 we introduce conjunctors and we discuss some properties they satisfy and that will be useful in next sections. In Section 4 we include the general result we have obtained for the transitivity of the indifference relation I and we discuss the appearance of the general expression for the three most important particular cases. In Section 5 we characterize the transitivity of the strict preference relation P in a very general context. We also prove that the general expression gets much simplified for the most important particular cases. In Section 6 we briefly address some conclusions and future work.

2 Preference structures

2.1 Crisp preference structures

Let us consider a decision maker who is given a set of alternatives A. Let us suppose that this person compares the alternatives two by two. Given two alternatives, the decision maker

ISBN: 978-989-95079-6-8

can act in one of the following three ways: (i) he/she clearly prefers one to the other; (ii) the two alternatives are indifferent to him/her; (iii) he/she is unable to compare the two alternatives. According to these cases, three binary relations can be defined on A: the strict preference relation P, the indifference relation I and the incomparability relation J. Thus, for any $(a,b) \in A^2$, we classify:

$$\begin{array}{rcl}(a,b) \in P & \Leftrightarrow & \text{he/she prefers } a \text{ to } b;\\ (a,b) \in I & \Leftrightarrow & a \text{ and } b \text{ are indifferent to him/her;}\\ (a,b) \in J & \Leftrightarrow & \text{he/she is unable to compare } a \text{ and } b.\end{array}$$

We recall that for a binary relation Q on A, its converse is defined as $Q^t = \{(b,a) \mid (a,b) \in Q\}$, its complement as $Q^c = \{(a,b) \mid (a,b) \notin Q\}$ and its dual as $Q^d = (Q^t)^c$. If we consider the set A^2 ordered, *i.e.* assuming (a,b) and (b,a) as different pairs, one easily verifies that the triplet (P,I,J) and P^t establish a particular partition of A^2 [21].

Definition 1 *A preference structure on A is a triplet (P,I,J) of binary relations on A that satisfy:*

(i) *P is irreflexive, I is reflexive and J is irreflexive;*

(ii) *P is asymmetrical, I and J are symmetrical;*

(iii) *$P \cap I = \emptyset$, $P \cap J = \emptyset$ and $I \cap J = \emptyset$;*

(iv) *$P \cup P^t \cup I \cup J = A^2$.*

Every preference structure has associated a reflexive relation that completely characterizes this structure. A preference structure (P,I,J) on A is characterized by the reflexive binary relation $R = P \cup I$, its large preference relation, in the following way:

$$(P,I,J) = (R \cap R^d, R \cap R^t, R^c \cap R^d). \qquad (1)$$

Conversely, for any reflexive binary relation R on A, the triplet (P,I,J) constructed in this way from R is a preference structure on A such that $R = P \cup I$. As R is the union of the strict preference and the indifference, $(a,b) \in R$ means that a is at least as good as b.

Given a binary relation Q on A, we say that Q is transitive if $(aQb \wedge bQc) \Rightarrow aQc$, for any $(a,b,c) \in A^3$. Given two binary relations Q_1 and Q_2 on A, the composition is a binary relation denoted $Q_1 \circ Q_2$ such that for any $(a,b) \in A^2$ $a\,(Q_1 \circ Q_2)\,b \Leftrightarrow \exists c / aQ_1c \wedge cQ_2b$. Then, it is clear that Q is transitive if and only if $Q \circ Q \subseteq Q$. The transitivity of the large preference relation R can be characterized as follows [2].

Theorem 1 *For any reflexive relation R with corresponding preference structure (P,I,J) it holds that*

$$R \circ R \subseteq R \Leftrightarrow (P \circ P \subseteq P \wedge I \circ I \subseteq I \wedge P \circ I \subseteq P \wedge I \circ P \subseteq P).$$

In case R is complete, *i.e.* $R \cup R^t = A^2$, this characterization can be simplified as follows. Note that the completeness of R is equivalent to establish that any two elements are comparable, that is, $J = \emptyset$.

Theorem 2 *For any complete reflexive relation R with corresponding preference structure $(P,I,\emptyset)$ it holds that*

$$R \circ R \subseteq R \Leftrightarrow (P \circ P \subseteq P \ \wedge \ I \circ I \subseteq I).$$

Next we recall an important characterization of a preference structure. Let us consider for every relation its characteristic mapping *i.e.* $Q(a,b) = 1 \Leftrightarrow aQb$. Definition 1 can be written in the following minimal way [11]: I is reflexive and symmetrical, and for any $(a,b) \in A^2$:

$$P(a,b) + P^t(a,b) + I(a,b) + J(a,b) = 1.$$

Classical preference structures can therefore also be considered as Boolean preference structures, employing 1 and 0 for describing presence or absence of strict preferences, indifferences and incomparabilities.

2.2 Additive fuzzy preference structures

In the classical model, relations only express presence or absence of relationship, while fuzzy relations capture the nuances of human choices. In fuzzy preference modelling, strict preference, indifference and incomparability are a matter of degree. These degrees can take any value between 0 and 1 and fuzzy relations are used for capturing them (for a complete review about fuzzy relations see [18]).

The intersection of fuzzy relations is usually defined pointwisely based on some t-norm, *i.e.* an increasing, commutative and associative binary operation on $[0,1]$ with neutral element 1. The three most important t-norms are the minimum operator $T_{\mathbf{M}}(x,y) = \min\{x,y\}$, the algebraic product $T_{\mathbf{P}}(x,y) = xy$ and the Łukasiewicz t-norm $T_{\mathbf{L}}(x,y) = \max\{x+y-1,0\}$. The minimum operator is the greatest t-norm; the smallest t-norm is the drastic product defined by

$$T_{\mathbf{D}}(x,y) = \begin{cases} \min\{x,y\} & \text{, if } \max\{x,y\} = 1\,,\\ 0 & \text{, otherwise.}\end{cases}$$

The above t-norms can be ordered (usual ordering of functions) as follows: $T_{\mathbf{D}} \leq T_{\mathbf{L}} \leq T_{\mathbf{P}} \leq T_{\mathbf{M}}$. Similarly, the union of fuzzy relations is based on a t-conorm, *i.e.* a non-decreasing, commutative and associative binary operation on $[0,1]$ with neutral element 0. T-norms and t-conorms come in dual pairs: to any t-norm T there corresponds a t-conorm S through the relationship $S(x,y) = 1 - T(1-x, 1-y)$. For the above three t-norms this yields the maximum operator $S_{\mathbf{M}}(x,y) = \max\{x,y\}$, the probabilistic sum $S_{\mathbf{P}}(x,y) = x + y - xy$ and the Łukasiewicz t-conorm (bounded sum) $S_{\mathbf{L}}(x,y) = \min\{x+y,1\}$. For more information on t-norms and t-conorms, we refer to [19]. Along this paper we use the notations for t-norms presented there.

The definition of a fuzzy preference structure has been a topic of debate during several years (see e.g. [18, 23, 24]). Accepting the *assignment principle* — for any pair of alternatives (a,b) the decision maker is allowed to assign at least one of the degrees $P(a,b)$, $P(b,a)$, $I(a,b)$ and $J(a,b)$ freely in the unit interval — has finally led to a graduation of Definition 1 with the intersection based on the Łukasiewicz t-norm and the union based on the Łukasiewicz t-conorm. Interestingly, a minimal definition is identical to the classical one if we replace ordinary by fuzzy binary relations: a triplet (P,I,J) of fuzzy binary relations on A is a fuzzy preference structure on A if and only if I is reflexive and symmetrical, and for any $(a,b) \in A^2$:

$$P(a,b) + P^t(a,b) + I(a,b) + J(a,b) = 1,$$

ISBN: 978-989-95079-6-8

where $P^t(a,b) = P(b,a)$. This identity explains the name *additive fuzzy preference structures*.

Another topic of controversy has been how to construct such a structure from a reflexive fuzzy relation. Alsina [1] proved a kind of impossibility theorem showing that a construction based on a single t-norm is unfeasible. As a reaction, Fodor and Roubens adopted an axiomatic approach [18]. The most recent and most successful approach is that of De Baets and Fodor based on (indifference) generators [8].

Definition 2 *A generator i is a commutative $[0,1]^2 \to [0,1]$ mapping bounded by the Łukasiewicz t-norm $T_{\mathbf{L}}$, and the minimum operator $T_{\mathbf{M}}$: $T_{\mathbf{L}} \leq i \leq T_{\mathbf{M}}$.*

Note that the definition of a generator does not speak of monotonicity and therefore they are not necessarily t-norms, albeit having neutral element 1. For any reflexive fuzzy relation R on A it holds that the triplet (P,I,J) of fuzzy binary relations on A defined by:

$$\begin{aligned} P(a,b) &= R(a,b) - i(R(a,b),R(b,a)), \\ I(a,b) &= i(R(a,b),R(b,a)), \\ J(a,b) &= i(R(a,b),R(b,a)) - (R(a,b)+R(b,a)-1), \end{aligned}$$

is an additive fuzzy preference structure on A such that $R = P \cup_{S_{\mathbf{L}}} I$ *i.e.* $R(a,b) = P(a,b) + I(a,b)$.

Popular generators (see e.g. [18]) are the Frank t-norms. For the sake of completeness, we recall that the Frank t-norms are given by

$$T_\lambda^{\mathbf{F}}(x,y) = \begin{cases} T_{\mathbf{M}}(x,y) & \text{, if } \lambda = 0, \\ T_{\mathbf{P}}(x,y) & \text{, if } \lambda = 1, \\ T_{\mathbf{L}}(x,y) & \text{, if } \lambda = \infty, \\ \log_\lambda(1 + \frac{(\lambda^x-1)(\lambda^y-1)}{\lambda-1}) & \text{, otherwise.} \end{cases}$$

We also recall that for any $\lambda \in [0,\infty]$ and for any $(x,y) \in [0,1]^2$,

$$T_{1/\lambda}^{\mathbf{F}}(x,y) = x - T_\lambda^{\mathbf{F}}(x,1-y).$$

3 Conjunctors

3.1 Generalizing T-transitivity

The usual way of defining the transitivity of a fuzzy relation is by means of a t-norm. Recall that a fuzzy relation Q on A is T-transitive if $T(Q(a,b),Q(b,c)) \leq Q(a,c)$ for any $(a,b,c) \in A^3$. However, such a framework is too restrictive in the setting of fuzzy preference modelling. On the one hand, even when the large preference relation R is T-transitive w.r.t. a t-norm T, the transitivity of the generated P and I cannot always be expressed w.r.t. a t-norm [14, 16, 15]. On the other hand, the results we will present in the following sections also hold when R itself is transitive w.r.t. a more general operator. From a fuzzy preference modelling point of view, it is not that surprising that the class of t-norms is too restrictive, as a similar conclusion was drawn when identifying suitable generators, as briefly explained in the previous section.

As it was discussed in [15, 16], suitable operators for defining transitivity for fuzzy relations are conjunctors:

Definition 3 *A binary operation $f : [0,1]^2 \to [0,1]$ is called a conjunctor if*

1. *it coincides with the Boolean conjunction on $\{0,1\}^2$;*
2. *it is increasing in each variable.*

Given a conjunctor f, we say that a fuzzy relation Q defined on A is f-transitive if for any $(a,b,c) \in A^3$,

$$f(Q(a,b),Q(b,c)) \leq Q(a,c).$$

It is clear that the definition of conjunctor largely extends the notion of t-norm. However, conjunctors and generators are not connected. The smallest conjunctor c_S and greatest conjunctor c_G are given by

$$c_S(x,y) = \begin{cases} 0 & \text{, if } \min\{x,y\} < 1, \\ 1 & \text{, otherwise,} \end{cases}$$

and

$$c_G(x,y) = \begin{cases} 0 & \text{, if } \min\{x,y\} = 0, \\ 1 & \text{, otherwise.} \end{cases}$$

As it is logical, $c_S \leq T_{\mathbf{D}} \leq T_{\mathbf{M}} \leq c_G$. Moreover, for two conjunctors $f \leq g$, it obviously holds that g-transitivity implies f-transitivity.

Defining the composition $Q_1 \circ_f Q_2$ of two fuzzy relations Q_1 and Q_2 w.r.t. a conjunctor f by

$$Q_1 \circ_f Q_2(a,c) = \sup_b f(Q_1(a,b),Q_2(b,c)),$$

still allows us to use the shorthand $Q \circ_f Q \subseteq Q$ to denote f-transitivity.

If we restrict our study for reflexive fuzzy relations, such as a large preference relations, the set of all the conjunctors that allow us to define transitivity is a proper subset of the set of all the conjunctors. Thus, the upper bound of this subset is $T_{\mathbf{M}}$ instead of c_G.

3.2 Dominance and bisymmetry

Dominance is a well-known relation for t-norms (see e.g. [19]) and its usefulness has been demonstrated several times (see e.g. [9, 22]). It can be generalized to conjunctors without any problem.

Definition 4 *A conjunctor f_1 is said to dominate a conjunctor f_2, denoted $f_1 \gg f_2$, if for any $(x,y,z,t) \in [0,1]^4$ it holds that*

$$f_1(f_2(x,y),f_2(z,t)) \geq f_2(f_1(x,z),f_1(y,t)).$$

But conjunctors do not verify the same properties as t-norms do with respect to this property. For instance, not every conjunctor dominates itself (see [15]). Dominance and the classical order of binary operators are not related for conjunctors neither. The minimum for example (which is not the greatest conjunctor), dominates any other conjunctor (see [15]).

The notion of self-dominance of conjunctors is obviously equivalent to another well-known property: bisymmetry (see e.g. [19]).

Definition 5 *A conjunctor f is said to be bisymmetric if for any $(x,y,z,t) \in [0,1]^4$ it holds that*

$$f(f(x,y),f(z,t)) = f(f(x,z),f(y,t)).$$

ISBN: 978-989-95079-6-8

3.3 Implications

Given a t-norm T, an implication (also called R-implication or T-residuum) based on T is defined (see e.g. [18, 19]) as follows:

$$I_T(x,y) = \sup\{z \mid T(x,z) \leq y\}.$$

This concept aims at generalizing the notion of Boolean implication. The definition usually concerns only left continuous t-norms. When T is left-continuous it holds that $T(x,z) \leq y \Leftrightarrow z \leq I_T(x,y)$.

Given a t-conorm S, a co-implication based on S is defined (see again e.g. [18, 19]) as follows:

$$J_S(x,y) = \inf\{z \mid S(x,z) \geq y\}.$$

When S is right-continuous it holds that $S(x,z) \geq y \Leftrightarrow z \geq J_S(x,y)$.

Following these ideas we can introduce two operators associated not only for a t-norm or a t-conorm, but to any commutative conjunctor.

Definition 6 *Given a commutative conjunctor $f : [0,1]^2 \to [0,1]$, we define the following operator from $[0,1]^2$ into $[0,1]$:*

$$I_f(x,y) = \sup\{z \mid f(x,z) \leq y\},$$

and

$$J_f(x,y) = \inf\{z \mid f(x,z) \geq y\}.$$

Let us notice that I_f is an implication, but J_f is not a co-implication.

Lemma 1 *Given a commutative conjunctor f and its associated operators I_f and J_f, we have that*

1. *I_f and J_f are decreasing in their first arguments and increasing in their second arguments.*
2. *If f is left continuous then*

$$f(x,z) \leq y \quad \Leftrightarrow \quad z \leq I_f(x,y).$$

3. *If f is upper bounded by the minimum t-norm, then*

$$x \leq y \quad \Rightarrow \quad I_f(x,y) = 1.$$

Moreover, if f is left-continuous and has 1 as neutral element, then we have the equivalence:

$$x \leq y \quad \Leftrightarrow \quad I_f(x,y) = 1.$$

4 Transitivity of I

In this section we consider indifference relations. As it was shown in Subsection 2.2, for a generator i, the symmetric component of a large preference relation is obtained as $I = i(R, R^t)$. Here we study the transitivity we can derive for this relation I when we fix the transitivity of R defined by a conjunctor h.

We begin recalling some upper and lower bounds for the transitivity of I. We know from [15] that I is at least c_S-transitive when R is h-transitive for any conjunctor h. We also showed in that paper that the transitivity we can assure for I is defined by a conjunctor upper bounded both by the conjunctor h and by the generator i that builds I from R. These are upper and lower bounds. We next provide a characterization for the transitivity of I.

Theorem 3 *Let i be an increasing generator and h a conjunctor. For any reflexive fuzzy relation R with corresponding indifference relation I generated by means of i, it holds that*

$$R \text{ is } h\text{-transitive} \quad \Rightarrow \quad I \text{ is } f_h^i\text{-transitive,}$$

where the conjunctor f_h^i is:

$$f_h^i(x,y) = \inf_{\substack{1 \geq u \geq x \\ 1 \geq v \geq y}} \left(i\left(h(u,v), h\left(J_i(v,y), J_i(u,x)\right)\right)\right).$$

Moreover, if i is continuous, the previous one is the strongest possible implication.

An interesting problem is to know when the transitivity of R is inherited by I, that is, to know when, departing from an h-transitive R we can assure that I is also h-transitive. We have already answered that question in [15]:

Theorem 4 *Let i be an increasing generator and let h be a commutative conjunctor upper bounded by the minimum t-norm. The associated conjunctor f_h^i is equal to h if and only if i dominates h.*

In particular this result could be applied to $i = T_{\mathbf{M}}$. Thus, when the indifference relation is obtained from the reflexive relation R by the minimum t-norm, $I = \min\{R, R^t\}$, it satisfies the same transitivity as R does.

Not only dominance allows us to obtain some general result, also the usual order among conjunctor, as it is showed in the following corollary.

Corollary 1 *For any increasing and bisymmetric generator i and any conjunctor h upper bounded by the minimum t-norm, if $i \leq h$, then $f_h^i = i$.*

Since every t-norm is increasing and bisymmetric, this result can be applied in particular to any generator that is a t-norm. In that case, transitivity of R w.r.t. that t-norm is preserved.

We will now consider Theorem 4 and Corollary 1 to study the particular case of a t-norm of the Frank family as generator.

Corollary 2 *Let $T_\lambda^{\mathbf{F}}$ be a t-norm of the Frank family, for any $\lambda \in [0,\infty]$. It holds that $f_{T_{\mathbf{P}}}^{T_{\mathbf{M}}} = T_{\mathbf{P}}$, $f_{T_{\mathbf{L}}}^{T_\lambda^{\mathbf{F}}} = T_{\mathbf{L}}$ and $f_h^{T_\lambda^{\mathbf{F}}} = T_\lambda^{\mathbf{F}}$ for any $h \geq T_\lambda^{\mathbf{F}}$.*

Remark 1 *Combining the results in Corollary 2 leads to the following table, where the entries are the conjunctors f_h^i.*

$h \setminus i$	$T_{\mathbf{L}}$	$T_{\mathbf{P}}$	$T_{\mathbf{M}}$
$T_{\mathbf{L}}$	$T_{\mathbf{L}}$	$T_{\mathbf{L}}$	$T_{\mathbf{L}}$
$T_{\mathbf{P}}$	$T_{\mathbf{L}}$	$T_{\mathbf{P}}$	$T_{\mathbf{P}}$
$T_{\mathbf{M}}$	$T_{\mathbf{L}}$	$T_{\mathbf{P}}$	$T_{\mathbf{M}}$

5 Transitivity of P

In this section we focus on the strict preference relation P and the transitivity it can satisfy once the transitivity of R is fixed and defined by a conjunctor h.

Let us recall that the strict preference relation P is obtained from R by means of the generator i as follows: $P = R - i(R, R^t)$.

ISBN: 978-989-95079-6-8

We showed in [15] that the transitivity that can be assured for the strict preference relation associated to any h-transitive large preference relation R is upper bounded by the conjunctor h, i.e., we cannot assure for P a stronger transitivity than h-transitivity. Next we provide not a bound but the explicit expression for the transitivity of P.

Theorem 5 *Let i be an increasing 1-Lipschitz generator and h a commutative conjunctor with neutral element 1. For any reflexive fuzzy relation R with corresponding strict preference relation P generated by means of i, it holds that*

$$R \text{ is } h\text{-transitive} \quad \Rightarrow \quad P \text{ is } g_h^i\text{-transitive}\,,$$

where the conjunctor g_h^i is:

$$g_h^i(x,y) = \inf_{\substack{1 \geq u \geq x \\ 1 \geq v \geq y}} (h(u,v) - i(h(u,v), \min\{I_h(v, I_i(u, u-x)), I_h(u, I_i(v, v-y))\}))\,.$$

Moreover, the previous one is the strongest possible implication.

According to our result from [15], g_h^i should always be not greater than h. Again, we will focus our attention on the cases the transitivity of R is totally inherited by P.

Proposition 1 *Let h be a rotation-invariant conjunctor, i.e., for all $(x,y,z) \in [0,1]^3$ it holds that*

$$h(x,y) \leq z \quad \Leftrightarrow \quad h(y, 1-z) \leq 1-x.$$

If $i = T_{\mathbf{L}}$, then $g_h^{T_{\mathbf{L}}} = h$.

Two well-known rotation-invariant t-norms are $T_{\mathbf{L}}$ and the minimum nilpotent $T_{\mathbf{nM}}$ defined by $T_{\mathbf{nM}}(x,y) = \min\{x,y\}\chi_{\{(x,y)|x+y>1\}}$, where χ_B denotes the characteristic function of any set B. Therefore, from Proposition 1, $g_{T_{\mathbf{L}}}^{T_{\mathbf{L}}} = T_{\mathbf{L}}$ and $g_{T_{\mathbf{nM}}}^{T_{\mathbf{L}}} = T_{\mathbf{nM}}$. This last conjunctor was directly obtained in [16].

If $i = T_{\mathbf{M}}$, only some special conjuntors satisfy the equality $g_h^{T_{\mathbf{M}}} = h$. The minimum t-norm is one of those t-norms. The following theorem characterizes all the conjunctors that verify the previous equality.

Proposition 2 *Let h be a commutative conjunctor with neutral element 1. Then h satisfies the equality $g_h^{T_{\mathbf{M}}} = h$ if and only if h is of the form*

$$h_k^d(x,y) = \begin{cases} 0 & \text{if } max(x,y) < d \\ k \cdot min(x,y) & \text{if } max(x,y) = d \\ min(x,y) & \text{otherwise} \end{cases}$$

where $k \in \{0,1\}$ and $d \in [0,1) \cup \{k\}$.

For the particular case when the conjunctor defining the transitivity of R and the indifference generator i are the same continuous t-norm the thorny general expression obtained in Theorem 5 gets much simpler.

Theorem 6 *Let T be a continuous t-norm. For any reflexive fuzzy relation R with corresponding strict preference relation P generated by means of T, it holds that*

$$R \text{ is } T\text{-transitive} \quad \Rightarrow \quad P \text{ is } g_T^T\text{-transitive}\,,$$

where

$$g_T^T(x,y) = \inf_{0 \leq \alpha \leq \min\{1-x, 1-y\}} \max\{T(x+\alpha, y+\alpha) - \alpha, 0\}$$

and this is the strongest possible implication.

In [5] the operator g_T^T was studied in depth not only for a t-norm T but in general for any binary aggregation operator A. In that general case, it was denoted as $\mathcal{D}[A]$. We can find there that for the Frank family of t-norms the expression can be simplified:

Proposition 3 *In case $i = T_\lambda^{\mathbf{F}} = h$ for $\lambda \in [0,\infty]$, it holds that*

$$g_{T_\lambda^{\mathbf{F}}}^{T_\lambda^{\mathbf{F}}}(x,y) = \begin{cases} T_\lambda^{\mathbf{F}}(x,y), & \text{if } x+y > 1\,, \\ S_M(T_\lambda^{\mathbf{F}}\left(\frac{1+x-y}{2}, \frac{1+y-x}{2}\right) - \frac{1-x-y}{2}\,,\, 0), & \text{otherwise}. \end{cases}$$

In particular, it holds that $g_{T_{\mathbf{L}}}^{T_{\mathbf{L}}} = T_{\mathbf{L}}$ (also obtained as a consequence of Proposition 1) , $g_{T_{\mathbf{M}}}^{T_{\mathbf{M}}} = T_{\mathbf{M}}$ (directly obtained in [13]) and

$$g_{T_{\mathbf{P}}}^{T_{\mathbf{P}}}(x,y) = \begin{cases} T_{\mathbf{P}}(x,y) - \left(\frac{T_{\mathbf{L}}(1-x,1-y)}{2}\right)^2, \text{ if } \sqrt{x} + \sqrt{y} > 1 \\ 0, \text{ otherwise.} \end{cases}$$

Other particular cases involving the three most important t-norms, i.e. concerning $h \in \{T_{\mathbf{L}}, T_{\mathbf{P}}, T_{\mathbf{M}}\}$ and $i \in \{T_{\mathbf{L}}, T_{\mathbf{P}}, T_{\mathbf{M}}\}$ are presented in the following propositions.

Proposition 4 *If $i = T_\lambda^{\mathbf{F}}$ for $\lambda \in [0,\infty]$ and $h = T_{\mathbf{L}}$, it holds that*

$$g_{T_{\mathbf{L}}}^{T_\lambda^{\mathbf{F}}}(x,y) = T_{1/\lambda}^{\mathbf{F}}(T_{\mathbf{L}}(x,y), S_{\mathbf{M}}(x,y)).$$

Proposition 5 *If $i = T_\lambda^{\mathbf{F}}$ and $h = T_{\mathbf{M}}$, it holds, for any $\lambda \in [0,\infty]$, that*

$$g_{T_{\mathbf{M}}}^{T_\lambda^{\mathbf{F}}}(x,y) = T_{\mathbf{nM}}^{\varphi_{1/\lambda}}(x,y)$$

where φ_λ is the automorphism of the interval $[0,\infty]$ defined as follows

$$\varphi_\lambda(x) = \begin{cases} x & \text{if } \lambda = 0, \\ \sqrt{x} & \text{if } \lambda = 1, \\ \frac{x+1}{2}\chi_{(0,1]}(x) & \text{if } \lambda = \infty, \\ \log_\lambda\left(\sqrt{\frac{\lambda^x - 1}{\lambda - 1}}\,(\lambda - 1) + 1\right) & \text{otherwise}, \end{cases}$$

and $T_{\mathbf{nM}}^{\varphi_{1/\lambda}}$ is the transformation by the automorphism $\varphi_{1/\lambda}$ of the t-norm $T_{\mathbf{nM}}$. Moreover, it holds that $g_{T_{\mathbf{M}}}^{T_\lambda^{\mathbf{F}}}$ is a t-norm.

As a consequence of this proposition, we obtain that $g_{T_{\mathbf{M}}}^{T_{\mathbf{L}}} = T_{\mathbf{nM}}$ and

$$g_{T_{\mathbf{M}}}^{T_{\mathbf{P}}}(x,y) = T_{\mathbf{nM}}^{\varphi_1}(x,y) = \begin{cases} \min\{x,y\} & \text{if } \sqrt{x} + \sqrt{y} > 1, \\ 0 & \text{otherwise}\,. \end{cases}$$

ISBN: 978-989-95079-6-8

In order to complete the study of all the combinations of the three most important t-norms, two cases are missing, namely $g_{T_{\mathbf{P}}}^{T_{\mathbf{L}}}$ and $g_{T_{\mathbf{P}}}^{T_{\mathbf{M}}}$, which will be revisited in Propositions 6 and 7.

Proposition 6 *It holds that*

$$g_{T_{\mathbf{P}}}^{T_{\mathbf{L}}}(x,y) = T_{\mathbf{M}}\left(T_{\mathbf{P}}(x,y), \frac{T_{\mathbf{L}}(x,y)}{T_{\mathbf{M}}(x,y)}\right) \cdot \chi_{(0,1]}(T_{\mathbf{M}}(x,y)).$$

When $T_{\mathbf{L}}$ is replaced by $T_{\mathbf{M}}$ as generator, the expression of $g_{T_{\mathbf{P}}}^{i}$ gets much more complicated.

Proposition 7 *It holds that*

$$g_{T_{\mathbf{P}}}^{T_{\mathbf{M}}}(x,y) = \max\{\min_{\substack{1 \geq u \geq x \\ 1 \geq v \geq y}} \left[uv - \min\left\{\tfrac{u-x}{v}, \tfrac{v-y}{u}\right\}\right], 0\} \cdot \chi_{(0,1]^2}(x,y).$$

Remark 2 *Combining the results in Propositions 1, 3, 4, 5, 6 and 7 leads to the following table, where the entries are the conjunctors g_h^i.*

$h \setminus i$	$T_{\mathbf{L}}$	$T_{\mathbf{P}}$	$T_{\mathbf{M}}$
$T_{\mathbf{L}}$	$T_{\mathbf{L}}$	$T_{\mathbf{P}}(T_{\mathbf{L}}, S_{\mathbf{M}})$	$T_{\mathbf{L}}(T_{\mathbf{L}}, S_{\mathbf{M}})$
$T_{\mathbf{P}}$	$g_{T_{\mathbf{P}}}^{T_{\mathbf{L}}}$	$g_{T_{\mathbf{P}}}^{T_{\mathbf{P}}}$	$g_{T_{\mathbf{P}}}^{T_{\mathbf{M}}}$
$T_{\mathbf{M}}$	$T_{\mathbf{nM}}$	$T_{\mathbf{nM}}^{\varphi_1}$	$T_{\mathbf{M}}$

6 Conclusions

This paper combines very general results with propositions concerning the most relevant particular cases. The general results can be applied to any conjunctor employed to define the transitivity of a large preference relation and any generator used to decompose that relation. The specific results concern the most important particular cases: those for which the conjunctors and the generators are the most important t-norms. The most general theorems we have introduced close the study of the transitivity that can satisfy the symmetric and asymmetric components of a (large) preference relation since those theorems involve (almost) any conjunctor and any generator we can use. The results concerning more particular operators provide more easy-to-use expressions for the most usual conjunctors and generators, i.e. for some t-norms. Only the last example leads to an unwieldy expression.

Despite the ugly general formulae obtained in Theorems 3 and 5, we have already proven that those conjunctors satisfy interesting properties for some particular cases. In future works we would like to study in depth the general expressions and the properties they satisfy.

Acknowledgment

The research reported on in this paper has been partially supported by Project FEDER-MEC-MTM2007-61193.

References

[1] C. Alsina. On a family of connectives for fuzzy sets. In *Fuzzy Sets and Systems*, 16:231–235, 1985.

[2] K. J. Arrow Social Choice and individual values. Wiley, 1951.

[3] M. Dasgupta and R. Deb. Transitivity and fuzzy preferences. In *Social Choice and Welfare*, 13:305–318, 1996.

[4] M. Dasgupta and R. Deb. Factoring fuzzy transitivity. In *Fuzzy Sets and Systems*, 118:489–502, 2001.

[5] B. De Baets, H. De Meyer, and S. Díaz. On an idempotent transformation of aggregation functions and its application on absolutely continuous archimedean copulas. In *Fuzzy Sets and Systems*, 160:733–751, 2009.

[6] B. De Baets and J. Fodor. Twenty years of fuzzy preference structures (1978-1997). *Belg. J. Oper. Res. Statist. Comput. Sci.*, 37:61–82, 1997.

[7] B. De Baets and J. Fodor. Generator triplets of additive fuzzy preference structures. In *Proc. RelMiCS6*.

[8] B. De Baets and J. Fodor. Additive fuzzy preference structures: the next generation. In *Principles of Fuzzy Preference Modelling and Decision Making*, 15–25, Academia Press, 2003.

[9] B. De Baets and R. Mesiar. t-partitions. In *Fuzzy Sets and Systems*, 97:211–223, 1998.

[10] B. De Baets and B. Van de Walle. Weak and strong fuzzy interval orders. *Fuzzy Sets and Systems*, 79:213–225, 1996.

[11] B. De Baets and B. Van de Walle. Minimal definitions of classical and fuzzy preference structures. In *Proceedings of the Annual Meeting of the Horth Amercan Fuzzy Information Precessing Society (Syracuse, New York, USA*, 299–304, 1997.

[12] B. De Baets, B. Van de Walle, and E. Kerre Fuzzy preference structures without incomparability. *Fuzzy Sets and Systems*, 76:333-348, 1995.

[13] S. Díaz, B. De Baets, and S. Montes. Additive decomposition of fuzzy pre-orders. *Fuzzy Sets and Systems*, 158:830–842, 2007.

[14] S. Díaz, B. De Baets, and S. Montes. On the compositional characterization of complete fuzzy pre-orders. *Fuzzy Sets and Systems*, 159:2221–2239, 2008.

[15] S. Díaz, S. Montes, and B. De Baets. Transitivity bounds in additive fuzzy preference structures. *IEEE Transactions on fuzzy systems*, 15(2):275–286, 2007.

[16] S. Díaz, S. Montes, and B. De Baets. Transitive decomposition of fuzzy preference relations: the case of nilpotent minimum. *Kybernetika*, 40:71–88, 2004.

[17] J. Fodor. Traces of binary fuzzy relations. *Fuzzy Sets and Systems*, 50:331–341, 1992.

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

[19] E. P. Klement, R. Mesiar, and E. Pap. *Triangular Norms*. Kluwer Academic Publishers, Dordrecht, 2000.

[20] B. Monjardet. Axiomatiques et propriétés des quasi-ordres. *Math. Sci. Hum.*, 63:51–82, 1978.

[21] M. Roubens and Ph. Vincke. Preference modelling. In *Lecture Notes in Economics and Mathematical Systems*, volume 250. Springer, Berlin, 1985.

[22] S. Saminger, R. Mesiar, and U. Bodenhofer. Domination of aggregation operators and preservation of transitivity. *Int. J. Uncertainty Fuz. and Knowledge-Based Syst.*, 10:11–35, 2002.

[23] B. Van de Walle, B. De Baets, and E. Kerre A plea for the use of Łukasiewicz triplets in the definition of fuzzy preference structures. Part 1: General argumentation. *Fuzzy Sets and Systems*, 97:349–359, 1998.

[24] B. Van de Walle, B. De Baets, and E. Kerre Characterizable fuzzy preference structures. *Ann. Oper. Res.*, 80:105–136, 1998.

[25] L. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

ISBN: 978-989-95079-6-8

Fuzzy Tendency based Time Series Model for Forecasting Server Traffic

Nadezhda G. Yarushkina[1], Tagir R.Unusov[2], Tatyana V.Afanasyeva[3]

1 Ulyanovsk State Technical University
32, Severny Venets street, Ulyanovsk, 432027, Russia
2 Ulyanovsk State Technical University
32, Severny Venets street, Ulyanovsk, 432027, Russia
3 Ulyanovsk State Technical University
32, Severny Venets street, Ulyanovsk, 432027, Russia
Email: jng@ulstu.ru, tagir@ulstu.ru, tv.afanaseva@mail.ru

Abstract *- For modeling of change of terminal server load, the approach including representation of time series of server parameters in the form of fuzzy time series is used. Further in the article the analysis of fuzzy time series is considered. The model of fuzzy tendencies is offered for terminal server traffic modeling. This model reflects change of the volume of terminal server traffic expressed linguistically, and it is used for its forecasting. The results of the forecast permit to determine linguistically expressed fuzzy tendencies on the basis of the offered model.*

Keywords— Fuzzy Time Series, Fuzzy Tendency, Fuzzy Neural Network, Forecasting, Terminal Server.

1 Introduction

During the last decade development of computing networks (CN) of different levels (from local area networks to global area networks) has resulted in the fact that the quality of production processes depends on the quality of network services more and more. Correspondence of scales of trade processes to the computer network can be achieved due to constant analysis of quality of CN functioning (including the application layer) and forecasting its development. Such analysis of existent network often rests upon results of time series (TS) analysis of server parameters.

Time series of server parameters have a number of peculiarities: non-stationarity, non-homogeneity, the complex form of dynamics, which limit the applicability of classical statistic methods of TS analysis such as regression and spectral analysis, exponential smoothing. In this connection, at present promising technologies of time series analysis are intelligent methods which permit to use expert knowledge and have no rigid requirements to initial data.

Modeling of experts' knowledge in the form of linguistic terms (evaluations) formalized as fuzzy sets became a prerequisite to TS modeling in the form of fuzzy time series.

Since 1993 a number of works have been dedicated to forecasting of fuzzy TS [1 - 4] constructed on the basis of numerical TS. In these investigations let us note the following problems which can be solved: increase of accuracy of numerical forecast owing to specification of the quantity and form of membership functions, decrease of computing owing to decrease of the number of fuzzy rules, increase of the level of automation when solving forecast problems. The indicated problems can be solved owing to application of methods of intelligent data analysis such as a fuzzy cluster analysis, genetic algorithms and neural networks.

The study [5] proposes the fuzzy neural network with initial weights generated by genetic algorithm (GFNN) for the sake of learning fuzzy IF–THEN rules. The result from GFNN is further integrated with an ANN artificial neural networks (ANN) forecast using the time series data from another ANN. Model evaluation results for a convenience store (CVS) company indicate that the proposed system can perform more accurately than the conventional statistical method and a single ANN.

In the work [6] the method of forecasting of TS inflation on the basis of fuzzy systems and a neural network is offered. To determine the fuzzy relation of the first-order fuzzy time series model, the neural network with SCL clustering technique was used to derive the rules directly from the database. The method may be of real usefulness in practical applications, where usually the expert can not explain linguistically, what control actions the process takes or there is no knowledge of the process.

To forecast fuzzy time series, in study [7] a backpropagation neural network is applied because of its nonlinear structures. Authors propose a hybrid model consisting of a neural network approach to forecast the known patterns as well as a simple method to forecast the unknown patterns.

Then the authors published the work [8] in which they proposed bivariate models in order to improve forecasting. The stock index and its corresponding index futures were taken as the inputs to forecast the stock index for the next day.

In this work there is offered a new hybrid forecasting model of TS in the form of a time series of fuzzy tendencies representing context-dependent interval linguistic evaluations of the traffic of the terminal server of a complex network. For the time series of FT, the model of forecasting of the first order is defined in terms of fuzzy tendencies to make more reasonable decisions on planning of network structure change.

In the part 2 the model and base operations of FT processing are defined.

In the part 3 a fuzzy neural network for generation of fuzzy rules of forecasting fuzzy tendencies is offered. The developed mathematical simulation model of the terminal server as an element of a computing network on the basis of time series of fuzzy tendencies presented in part 4 permits to forecast the traffic of the terminal server both in terms of fuzzy tendencies and in terms of the initial TS.

ISBN: 978-989-95079-6-8

2 Models and base operations of fuzzy tendencies processing

In the work [8] the evaluation of TS is analyzed which is expressed in linguistic terms, the concepts fuzzy time series (FTS) and fuzzy tendency (FT) are introduces for that. The fuzzy tendency of a fuzzy time series is the fuzzy label which expresses the dynamics (systematic motion) of a fuzzy time series. The concept "tendency", being modeled by the fuzzy set, can be described by the set of constructed empiric rules. To find the fuzzy tendencies, it is offered to use Fuzzy Neurons Network (FNN). The analysis of the FT permits to forecast the behavior of the time series on the basis of process dynamics by means of prediction of tendencies of development of the object being studied. With it forecasting of numerical values of TS is not foreseen, and it can be considered as a shortcoming of the offered methods.

The base operations of fuzzy tendencies processing are algorithmic operations of fuzzy tendencies processing, namely the operation of forming the time series of fuzzy tendencies by the initial time series and the inverse operation of generation of the series – a representative of the fuzzy tendency.

Definition 1. A fuzzy time series $\widetilde{Y}$ is the ordered sequence of observations of some phenomenon the character of which changes in the course of time if values which some quantity takes on at the instant of time are expressed by the fuzzy label:

$$\widetilde{y}_t^i = \{(\widetilde{y}^i, U_y, \mu_i(y_t)), t\},$$

where $\widetilde{y}^i$ is the *i*-th element of the term-set of the linguistic variable $(\widetilde{Y}, T_y, U_y, G_y, M_y)$.

Thus, a FTS is the vector time series of values of all fuzzy variables:

$$\widetilde{Y} = \{\widetilde{y}_t\},$$

where $\widetilde{y}_t = \{\widetilde{y}_t^1, ..., \widetilde{y}_t^n\}$, *n* is the number of terms.

Definition 2. Fuzzy tendency. Let $\widetilde{y}_\Delta = \{\widetilde{y}_1, ..., \widetilde{y}_m\}$ be the fuzzy time series of the linguistic variable $(\widetilde{Y}, T_y, U_y, G_y, M_y)$, $\widetilde{Y}_\Delta = \{\widetilde{y}_\Delta\}$ is the set of fuzzy time series of the same length. Then the fuzzy tendency τ, determined on the $\widetilde{Y}_\Delta$, is the aggregate of ordered pairs:

$$\tau = \{\widetilde{y}_\Delta, \mu_\tau(\widetilde{y}_\Delta)\},$$

where $\mu_\tau(\widetilde{y}_\Delta)$ is the degree of membership of $\widetilde{y}_\Delta$ to FT.

Determining the fuzzy tendency on all intervals [*t-m+1, t*] of the time series and positioning the beginning and the end of the interval to the time scale, we receive the time series of the fuzzy tendency.

Definition 3. The time series of a fuzzy tendency (TSFT). Let $\{\widetilde{y}_{\Delta t}\}$ be the set of fuzzy time series which length is equal to *m*, where $\widetilde{y}_{\Delta t} = \{\widetilde{y}_{t-m+1}, ..., \widetilde{y}_t\}$, $\widetilde{y}_t \in \widetilde{Y}$ t. Then the time series of a fuzzy tendency is the ordered in time fuzzy set:

$$\tau_t = \{t, \mu_\tau(\widetilde{y}_{\Delta t})\}.$$

For the linguistic variable FT, the time series of fuzzy tendencies is determined as the vector TS of values of all kinds of fuzzy tendencies:

$$\tau_t = \{t, \tau_t^1, ..., \tau_t^p\}.$$

Let us assume that there is dependence between fuzzy tendencies observed at different instants of time. Using the scheme of the differed equation (1) let us represent the model of the time on the basis of FT for one variable:

$$\tau_t = f(\tau_{t-1}, ...\tau_{t-l}), \quad (1)$$

where *l* is the time interval. For construction of the model of TS on the basis of FT, it is necessary to solve the following problems:

-to determine the procedure of fuzzyfication and defuzzyfication of the TS;

-to determine the procedure of description and identification of the FT;

-to determine the procedure of receiving the crisp time series from values of FT;

-to identify the functional dependence.

Receiving a FTS from a crisp TS is based on the operation of fuzzyfication, i.e. transition from quantitative values to linguistic evaluations. The procedure of fuzzyfication is widely enough described in literature. Let us denote the process of finding the value of membership function of elements *Uy* to the terminal set T_y by the functional Fuzzy:

$$\widetilde{Y} = Fuzzy[Y].$$

The functional supposes solving the problems of linguistic terms description and rules of use, choosing the corresponding membership functions. When determining the linguistic terms the following methods can be used: subjective description of the universe U_y on the basis of expert knowledge about the system and descriptive terminology accepted in the field being investigated; uniform partition of the universe (into 3-7 intervals, as a rule); clusterisation of values of the time series for determination of the most significant intervals on the universe.

Defuzzyfication is the problem inverse to fuzzyfication, i.e. receiving the crisp TS from a fuzzy one, is given by the functional deFuzzy:

$$Y' = deFuzzy[\widetilde{Y}].$$

The problem of determination of the FT is fundamental in the construction of the system of data analysis, decision of which will permit to start disclosing the connections between parameters of the object being investigated. This problem consists of the following sub-problems:

-determination of linguistic variables of the FT;

-construction of the time series of the FT.

Let us denote solving the problem by the functional Tend:

$$\tau = Tend[\widetilde{Y}_\Delta]$$

For determination of the linguistic variable expressing the fuzzy tendency, it is necessary to single out the typical behavior of the fuzzy time series, that is to make up the terminal set. The analysis of the initial series for evaluation of the dynamics can be done by the following methods:

-using the evaluations of the dynamics (increase, decrease, stability);

ISBN: 978-989-95079-6-8

-subjective description on the basis of expert knowledge about the systematic behavior of system parameters;

-clusterisation of values of the fuzzy time series for determination of the most significant tendencies on the universe.

For construction of TSFT, it is necessary not only to single them out (to denote linguistic terms), but to give a description and be able to find the tendency on the FTS. Thus, the functional Tend includes the tool of FT description, the algorithm of finding the correspondence of fuzzy time series to the chosen description, that is the algorithm of fuzzy evaluation.

For the analysis and construction of crisp TS by the fuzzy model, the operation of receiving the FTS from tendency evaluation is necessary. Let us denote solving of this problem by the functional *deTend*, which is inverse to the functional Tend:

$$\widetilde{Y}`_{\Delta} = deTend[\tau],\ \widetilde{Y}`_{\Delta} \in \widetilde{Y}_{\Delta},$$

where $\widetilde{Y}`_{\Delta} = \{\widetilde{y}`_{\Delta}\}$ is the set of typical (characteristic) fuzzy time series for FT which have the maximum value of membership function $\mu_{\tau}(\widetilde{y}`_{\Delta}) = \max(\mu_{\tau}[\widetilde{Y}_{\Delta}])$.

Using the functional of defuzzyfication *deFuzzy*, receiving the fuzzy time series from tendencies *deTend* solves the problem of crisp TS forecasting.

Let us denote the aggregate of components and equations:

$$\widetilde{y}_t^i = Fuzzy[y_t],\ y_t = deFuzzy[\widetilde{y}_t^1,...\widetilde{y}_t^n], i = 1..n,$$

$$\tau_t^j = Tend[\widetilde{y}_{t-m_j+1},...,\widetilde{y}_t],$$

$$\widetilde{y}_t^i = deTend[\tau_t,...,\tau_{t+m-1}], m = \max(m_j),\ j = 1..p,$$

$$\tau_t = f(\tau_{t-1},...\tau_{t-l}),$$

where n is the number of terms of FTS, p is the number of terms of FT, m_j is the interval of definition of FT, l is the time log, by the *model of fuzzy tendencies* (MFT) with characterizing parameters (n, p, m, l). In more detailed form these parameters can be denoted as (n, $\{p_k\}$, $\{m_k\}$, l), where p_k is the number of kinds of tendencies which have the interval of definition m_k.

For the analysis and construction of the crisp time series by the fuzzy model (the functional *deTend)* let us make the fuzzy time series which has the maximum value of membership function in the form of the rule $\widetilde{y}_{\Delta t} = (\tau_t) \circ R_2$ correspond to each kind of fuzzy tendency. For a example, the procedure *deTend* can be represented in the form of the system of rules:

if τ_t is jump, then y_{t-2} is low,

if τ_t is jump, then y_{t-1} is high,

if τ_t is jump, then y_t is low.

Let us represent the equation (1) also by the fuzzy relation:

$$\tau_t = (\tau_{t-1},...\tau_{t-l}) \circ R_3.$$

The MFT is realized by multilevel system of logical relations:

$$R_1 \Rightarrow R_3 \Rightarrow R_2,$$

where outputs in the form of fuzzy variables of one set of rules are applied to inputs of the next set of rules without defuzzyfication and fuzzyfication. Transformation into fuzzy and crisp values happens only in relations R_1 and R_2 correspondingly.

3 Fuzzy neural network

Let us use classical fuzzy neurons in which operations of addition and multiplication are replaced by triangular norms:

AND-neuron $\beta = T(S(\tau_1, w_1), S(\tau_2, w_2))$,

OR-neuron $\tau^0 = S(T(\beta_1, z_1), T(\beta_2, z_2))$.

It can be interpreted in the linguistic form correspondingly as:

if ($\tau 1$ or $w1$) and (τ_2 or w_2) then β,

if (β_1 *and* z_1) or (β_2 *and* z_2) then τ_0.

On the basis of such neurons it is possible to construct the network of logical inference by Mamdani for finding the fuzzy tendency τ by adding weight coefficients. In a formalized way such network can be expressed in the following form:

$$\tau^0 = \underset{i=1}{\overset{k}{S}}[T(\beta_i, z_i)],\ \beta_i = \underset{j=1}{\overset{m}{T}}[S(\tau_j, w_{j,i})],$$

where T is the operator of conjunction, S is the operator of disjunction, k is the number of rules, m is the number of inputs. Weight coefficients are interpreted in the following way: z_i is the degree of influence of i-th rule on the general result (0 – has no influence, 1 – has influence); $w_{j,i}$ are degrees of no influence of j-th input on i-th rule (0 – has influence, 1 – has no influence). The basic idea of FNN learning consists in the iterative procedure of weights optimization (z_i, $w_{j,i}$) and removal of insignificant connections (network reduction), as a result of which the necessary composition and number of rules are formed.

The order of search fuzzy dependences on the basis of FNN is the following:

-Initialization of the fuzzy neural network.

-Learning: optimization of network weights.

-Analysis of the network: network reduction.

At the initialization stage it is necessary to determine input variables: kinds of tendencies and the time log. It is necessary to generate an excessive number of "complete" rules, which include all inputs with arbitrary weights in the interval (0,1).

The stage of learning is the process of FNN weights change on the basis of the learning sample. Network learning is possible by the method of back error propagation. For this purpose let us determine the function of error $E = \frac{1}{2}(\tau^0 - \tau)^2$ which is to be minimized by the method of gradient lowering. For decrease of the squared error E, it is necessary to change weights (z_i, $w_{j,i}$) in the direction of antigradient of the function E:

$$w_{j,i}^{+1} = w_{i,j} - \eta \frac{\partial E}{\partial w_{j,i}},\ z_i^{+1} = z_i - \eta \frac{\partial E}{\partial z_i},$$

where η is the speed of learning with the limitation $w_{j,i}^{+1} \in [0,1]$, $z_i^{+1} \in [0,1]$. For calculation of the derivative error, it is necessary to choose the corresponding functions of t-norms and t-conorms. The indication of completion of learning is reaching the error level.

ISBN: 978-989-95079-6-8

At the stage of analysis processing of FNN is carried out with the purpose of its simplification. The principle of removal of insignificant connections and neurons from the network forms the basis of reduction algorithms. One of simplest reduction methods is the method of projections which is realized in the following way. The synoptic weight is nulled if its value has fallen into the given range

$$w_{j,i} = \begin{cases} 1, w_{j,i} \geq (1-\varepsilon) \\ w_{j,i}, w_{j,i} < \varepsilon \end{cases}, \; z_i = \begin{cases} 0, z_i \leq \varepsilon \\ z_i, z_i > \varepsilon \end{cases},$$

where ε is some constant. On the basis of the chosen level of weights ε connections "input – AND-neuron", when $w_{j,i}=1$, and "AND-neuron – OR-neuron", when $z_i = 0$, are removed. Not used inputs and rules are also removed.

As a result the system of logical inference is got, which is the explaining function in the model of time series. Fuzzy rules are easily interpreted for an expert since they are expressed in terms inherent to the investigated field.

4 Server traffic modeling on the basis of the time series of fuzzy tendencies

For time series analysis on the basis of the offered model of a fuzzy time series, the software *FuzzyTendNet* is created.

Thus, for each time series a user assigns the set of linguistic terms with corresponding membership functions (trapezoid function form is used) and the set of fuzzy tendencies. The possibility of using several rules of identification and inverse rules for each tendency permits to describe various interval expert evaluations flexibly enough.

The equation of the model of a time series is the fuzzy neural network which has subnetworks for each output tendency. Each subnetwork is the system of logical deduction which consists of the set of rules. These rules can be formed manually and in the automatic mode with optimization on the basis of the algorithm of back error propagation. The method of projections is used for cancellation of the network. The following logical connectives (functions of triangular norms) are used: minimum and maximum — for rules of identification and inverse rules, product and probabilistic sum — for FNN.

For modeling of the traffic of the terminal server on the basis of the offered approach, seven parameters of the work of the terminal server were chosen, and time series were formed from them.

So, the server parameters of work were chosen (Table 1) and the statistics with 15 seconds interval during one day was collected.

Table 1. Description of variables

Parameter
X1 - Memory\ Pages exchange per second
X2 - Record accesses to disc per second
X3 - Read accesses from disc per second
X4 - % of processor load
X5 - Reading operations
X6 - Recording operations
Y - Traffic

For all parameters fuzzy variables of FTS describing the values "high" and "low" are determined. Fuzzy tendencies "load", "idle" are determined. Additional fuzzy tendencies "increase" and "decrease" are also determined for traffic. In Table 2 description of functionals of the MFT is presented.

Let us note the following:

-the functional deTend is described only for the dependent variable (traffic);

-FT determined on the unit interval, which have the semantic load analogous to variables of FTS, are used in the model.

Table 2. Description of tendencies of terminal-server model parameters change

Parameter	FTS	TSFT	
	$\tilde{y}$	τ	Tend
X1	low	idle	low(1) &low(2)
	high	load	high(1) & high(2)
X2	low	idle	low(1) & low(2)
	high	load	high(1) & high(2)
X3	low	idle	low(1) & low(2)
	high	load	high(1) & high(2)
X4	low	idle	low(1) & low(2)
	high	load	high(1) & high(2)
X5	low	idle	low(1) & low(2)
	high	load	high(1) & high(2)
X6	low	idle	low(1) & low(2)
	high	load	high(1) & high(2)
Y	low	idle	low(1) & low(2)
	high	load	high(1) & high(2)
		increase	low(1) & high(2)
		decrease	high(1) & low(2)

For comparison of models on the basis of fuzzy time series and fuzzy tendencies, a number of experiments were conducted.

The conclusion can be made that the model based on fuzzy tendencies describes the modeled process more successfully. On Fig. 1 the forecast for one step forward is presented for two models.

ISBN: 978-989-95079-6-8

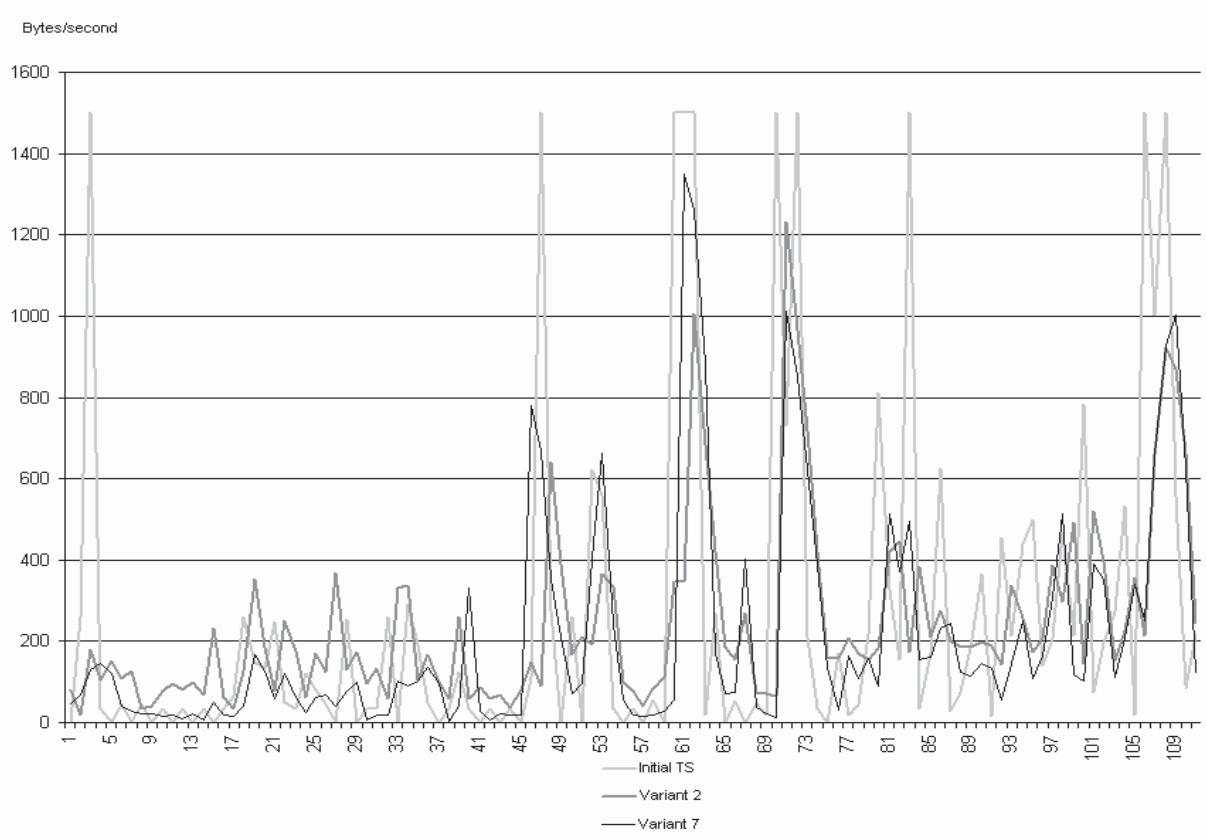


Figure 1:The graph of tested and forecasted TS

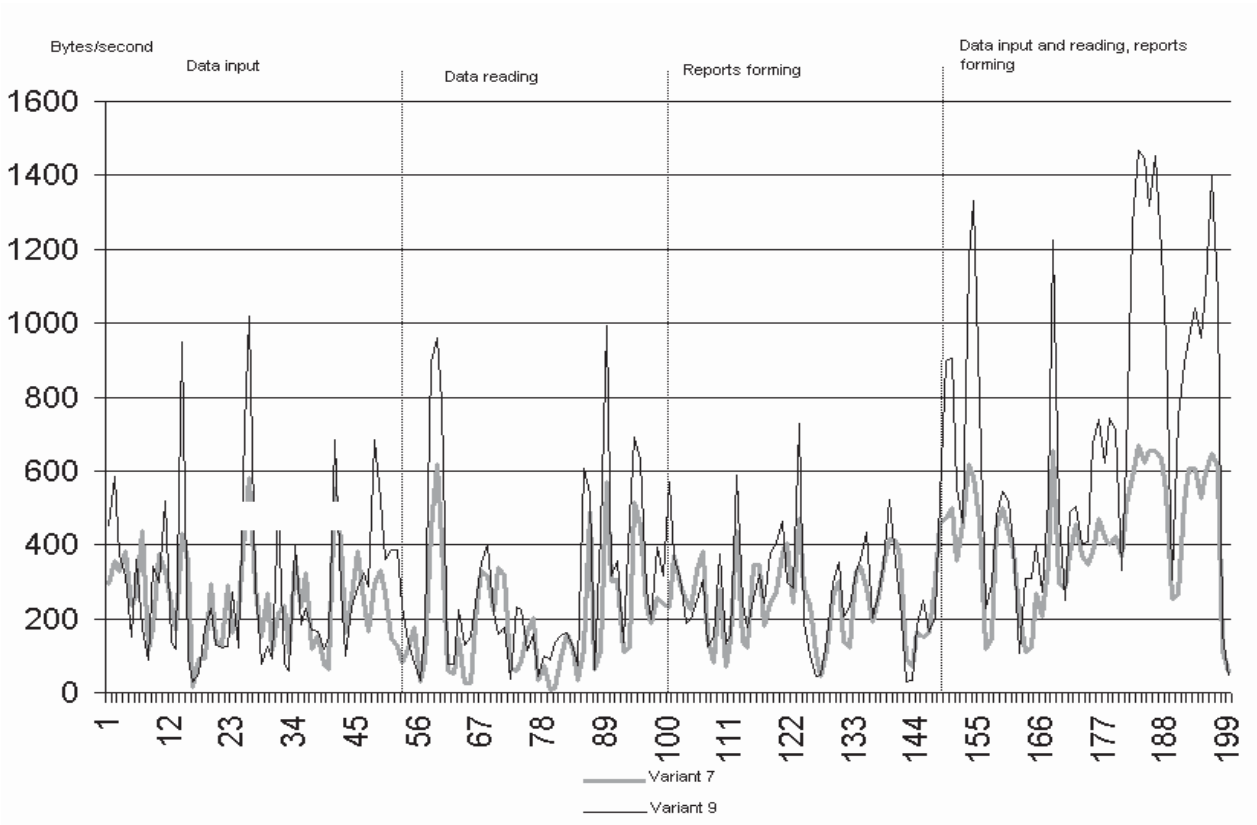


Figure 2:Load modeling

For server traffic volume modeling, the following modes of server work were described by the expert: data input, data reading, reports forming — with high values of corresponding groups of parameters. On Fig 2 results of modeling are presented.

By fuzzy tendencies low and high values of the traffic and its crisp values have been forecasted in different modes of work. The analysis of linguistic series of fuzzy tendencies showed increase of the traffic of the terminal server in the mode of data input and reports forming.

ISBN: 978-989-95079-6-8

Thus the conclusion can be made that increase of users performing the mentioned operations will be accompanied by increase of the traffic. When performing all operations on the server simultaneously, we can simultaneously observe both high and low traffic of the network, which can be explained by instability in this mode of work.

5 Conclusions

Thus, the time series of fuzzy tendencies is an effective indicator of development of a modeled process in a complex technical system described qualitatively by linguistic terms. Basic operations of fuzzy tendencies processing are algorithmic operations of fuzzy tendencies processing, namely the operation of forming the time series of fuzzy tendencies by initial time series and the inverse operation of generation of the series — a representative of the fuzzy tendency. The fuzzy neural network of the offered architecture is an effective generator of rules of fuzzy tendencies identification. The developed mathematical simulation model of the terminal-server as the element of the computing network on the basis of time series of fuzzy tendencies permits to forecast processor load, the outgoing and incoming traffic of the server.

References

[1] Q. Song, B. Chissom, Fuzzy time series and its models, Fuzzy Sets and Systems 54 (1993) 269–277.

[2] K. Huarng. Effective lengths of intervals to improve forecasting in fuzzy time series. Fuzzy Sets and Systems, 2001, 123: 387–394.

[3] K. Huarng, Heuristic models of fuzzy time series for forecasting, Fuzzy Sets and Systems 123 (2001) 369–386.

[4] Şah, M., Degtiarev, K.Y. : Forecasting Enrollment Model Based on First-Order Fuzzy Time Series. Proc. Int. Conf. Computational Intelligence (ICCI) (2004) 375–378

[5] R. J. Kuo . A sales forecasting system based on fuzzy neural network with initial weights generated by genetic algorithm// European Journal of Operational Research Volume 129, Issue 3, 16 March 2001, Pages 496-517

[6] Dušan Marček, Modelling of inflation fuzzy time series using the competitive neural network, Proc. International conference on softcomputing applied in computer and economic environments (ISCS) (2004) 107-111.

[7] Kun-Huang Huarng, Tiffany Hui-Kuang Yu. The application of neural networks to forecast fuzzy time series \\ Physica A: Statistical Mechanics and its Applications, Volume 363, Issue 2, 1 May 2006, Pages 481-491

[8] Tiffany Hui-Kuang Yu, Kun-Huang Huarng. A bivariate fuzzy time series model to forecast the TAIEX \\ Expert Systems with Applications Volume 34, Issue 4, May 2008, Pages 2945-2952

[9] N.G. Yarushkina Principles of the theory of fuzzy and hybrid systems: Training aid. - Moscow: Finances and statistics, 2004.

ISBN: 978-989-95079-6-8

Continuity and additivity of the trapezoidal approximation preserving the expected interval operator

Adrian I. Ban Lucian C. Coroianu
Department of Mathematics and Informatics, University of Oradea,
Universităţii 1, 410087 Oradea, Romania
Email: aiban@uoradea.ro, luccoro@yahoo.com

Abstract— *The nearest trapezoidal fuzzy number to a fuzzy number, with respect to a well-known metric and preserving the expected interval, was determined in recent articles. In the present paper the properties of additivity and continuity of the trapezoidal approximation operator are studied.*

Keywords— Additivity, Approximation, Continuity, Fuzzy number, Trapezoidal fuzzy number.

1 Introduction

In the papers [1] and [10] the problem to find the nearest (with respect to a well-known metric) trapezoidal fuzzy number $T(A)$ to a fuzzy number A such that $EI(T(A)) = EI(A)$, where $EI(B)$ denotes the expected interval of a fuzzy number B, was completely solved. Algorithms for computing the proper approximations are proposed in [6]. The properties of translation and scale invariance, identity, nearness criterion, expected interval invariance, order invariance, correlation invariance, uncertainty invariance, value and ambiguity invariance were considered in [1]. The properties of additivity and continuity are studied in the present paper.

In Section 3 we prove, by a simple example, that the additivity is not satisfied. The main result in this section contains a property of partial additivity. In the paper [5] the property of continuity of the approximations of fuzzy numbers is considered of extreme importance in applications, especially in fuzzy control systems, where it is sometimes indicated as robustness. The authors of the paper [2] pointed out that the method to study the continuity of the trapezoidal approximation without condition in [11] is not applicable in the case of trapezoidal approximation preserving the expected interval. In Section 4 we prove that the discussed approximation operator has the property to be Lipschitz, such that it is continuous.

2 Preliminaries

A fuzzy number A is a fuzzy subset of the real line $\mathbb{R}$ with the membership function μ_A which is normal, fuzzy convex, upper semicontinuous and supp A is bounded, where supp $A = cl\{x \in \mathbb{R} : \mu_A(x) > 0\}$ and cl is the closure operator.

Every λ-cut, $\lambda \in]0,1]$ of a fuzzy number A is a closed interval $A_\lambda = [A^-(\lambda), A^+(\lambda)]$, where

$$A^-(\lambda) = \inf\{x \in \mathbb{R} : \mu_A(x) \geq \lambda\},$$
$$A^+(\lambda) = \sup\{x \in \mathbb{R} : \mu_A(x) \geq \lambda\}.$$

We denote

$$A_0 = \left[A^-(0), A^+(0)\right] = \text{supp } A.$$

For two arbitrary fuzzy numbers A, B

$$A_\lambda = \left[A^-(\lambda), A^+(\lambda)\right]$$

and

$$B_\lambda = \left[B^-(\lambda), B^+(\lambda)\right]$$

the addition $A + B$ is introduced by

$$(A+B)_\lambda = \left[A^-(\lambda) + B^-(\lambda), A^+(\lambda) + B^+(\lambda)\right]$$

and the quantity

$$D^2(A,B) = \int_0^1 \left(A^-(\lambda) - B^-(\lambda)\right)^2 d\lambda + \int_0^1 \left(A^+(\lambda) - B^+(\lambda)\right)^2 d\lambda \quad (1)$$

gives a distance between A and B (see, e.g., [4]). The expected interval $EI(A)$ of a fuzzy number $A, A_\lambda = [A^-(\lambda), A^+(\lambda)]$, is defined by (see [3], [7])

$$EI(A) = \left[\int_0^1 A^-(\lambda)\, d\lambda, \int_0^1 A^+(\lambda)\, d\lambda\right].$$

An often used fuzzy number is the trapezoidal fuzzy number, completely characterized by four real numbers $t_1 \leq t_2 \leq t_3 \leq t_4$, denoted by $T = (t_1, t_2, t_3, t_4)$, with

$$T^-(\lambda) = t_1 + (t_2 - t_1)\lambda,$$
$$T^+(\lambda) = t_4 - (t_4 - t_3)\lambda, \lambda \in [0,1].$$

Sometimes (see [10]) a trapezoidal fuzzy numbers is denoted by $T = (l, u; x, y)$, with $l, u, x, y \in \mathbb{R}$ such that $x, y \geq 0, x + y \leq 2(u - l)$,

$$T^-(\lambda) = l + x\left(\lambda - \frac{1}{2}\right),$$
$$T^+(\lambda) = u - y\left(\lambda - \frac{1}{2}\right), \lambda \in [0,1].$$

It is immediate that

$$l = \frac{t_1 + t_2}{2}, \quad (2)$$
$$u = \frac{t_3 + t_4}{2}, \quad (3)$$
$$x = t_2 - t_1, \quad (4)$$
$$y = t_4 - t_3, \quad (5)$$

ISBN: 978-989-95079-6-8

the expected interval of a trapezoidal fuzzy number represented in this way is the real interval $[l,u]$ and the distance between $T=(l,u;x,y)$ and $T'=(l',u';x',y')$ becomes ([10])

$$D^2(T,T') = (l-l')^2 + (u-u')^2 + \frac{1}{12}(x-x')^2 + \frac{1}{12}(y-y')^2. \quad (6)$$

We denote by $F(\mathbb{R})$ the set of all fuzzy numbers, by $F^T(\mathbb{R})$ the set of all trapezoidal fuzzy numbers and by $T(A)$ the nearest (with respect to the metric D) trapezoidal fuzzy number to fuzzy number A preserving the expected interval of A. To express in a simplified form the approximation operator $T: F(\mathbb{R}) \to F^T(\mathbb{R})$ we denote

$$I(A) = I = \int_0^1 A^-(\lambda)d\lambda, \quad (7)$$

$$S(A) = S = \int_0^1 A^+(\lambda)d\lambda, \quad (8)$$

$$L(A) = L = \int_0^1 \lambda A^-(\lambda)d\lambda \quad (9)$$

and

$$U(A) = U = \int_0^1 \lambda A^+(\lambda)d\lambda, \quad (10)$$

for every fuzzy number $A, A_\lambda = [A^-(\lambda), A^+(\lambda)], \lambda \in [0,1]$ and we consider the following sets of fuzzy numbers

$$\Omega_1 = \{A \in F(\mathbb{R}) : 2I+S-3L-3U>0\},$$
$$\Omega_2 = \{A \in F(\mathbb{R}) : -I-2S+3L+3U>0\},$$
$$\Omega_3 = \{A \in F(\mathbb{R}) : -I+S+3L-3U\le 0\},$$
$$\Omega_4 = \Omega_1^c \cap \Omega_2^c \cap \Omega_3^c.$$

The set $\{\Omega_1,\Omega_2,\Omega_3,\Omega_4\}$ is a partition of $F(\mathbb{R})$ (see the proof of Theorem 7 in [1]).

The following result was proved in [1] and [10].

Theorem 1 (i) *If* $A \in \Omega_1$ *then*

$$T(A) = T_1(A) = (I, I, I, 2S-I);$$

(ii) *If* $A \in \Omega_2$ *then*

$$T(A) = T_2(A) = (2I-S, S, S, S);$$

(iii) *If* $A \in \Omega_3$ *then*

$$T(A) = T_3(A) = (4I-6L, -2I+6L, -2S+6U, 4S-6U);$$

(iv) *If* $A \in \Omega_4$ *then*

$$T(A) = T_4(A) = (3I+S-3L-3U, -I-S+3L+3U, -I-S+3L+3U, I+3S-3L-3U).$$

Four different operators $T_i, i \in \{1,2,3,4\}$ give us the nearest trapezoidal approximation preserving the expected interval: T_1 and T_2 lead to triangular fuzzy numbers with right side only or left side only, respectively, T_3 produces proper trapezoidal fuzzy numbers and T_4 produces triangular fuzzy numbers.

3 Additivity

The approximation operator $T: F(\mathbb{R}) \to F^T(\mathbb{R})$ given in Theorem 1 is scale invariant (see [1], Theorem 12, (ii)) and invariant to translations (see [1], Theorem 12, (i)), that is

$$T(\alpha A) = \alpha T(A)$$

and

$$T(A+z) = T(A) + z,$$

for every $A \in F(\mathbb{R}), z \in \mathbb{R}$ and $\alpha \in \mathbb{R} \setminus \{0\}$. In this section we point out a partial result of additivity and prove, be a simple example, that the operator T is not additive.

Example 2 *Let us consider the fuzzy numbers* A *and* B *given by*

$$A^-(\lambda) = 1+\sqrt{\lambda},$$
$$A^+(\lambda) = 30-27\sqrt{\lambda},$$
$$B^-(\lambda) = -1+\sqrt{\lambda},$$
$$B^+(\lambda) = 1-\sqrt{\lambda}, \lambda \in [0,1].$$

Then (see [1], Example 10 and Example 11) $A \in \Omega_1$,

$$T(A) = \left(\frac{5}{3}, \frac{5}{3}, \frac{5}{3}, \frac{67}{3}\right)$$

$B \in \Omega_4$,

$$T(B) = \left(-\frac{2}{3}, 0, 0, \frac{2}{3}\right),$$

therefore

$$T(A)+T(B) = \left(1, \frac{5}{3}, \frac{5}{3}, 23\right).$$

Because

$$(A+B)^-(\lambda) = 2\sqrt{\lambda},$$
$$(A+B)^+(\lambda) = 31-28\sqrt{\lambda}, \lambda \in [0,1],$$

we obtain

$$I(A+B) = \frac{4}{3},$$
$$S(A+B) = \frac{37}{3},$$
$$L(A+B) = \frac{4}{5}$$

and

$$U(A+B) = \frac{43}{10}.$$

We get $A+B \in \Omega_4$ *and*

$$T(A+B) = \left(\frac{31}{30}, \frac{49}{30}, \frac{49}{30}, \frac{691}{30}\right)$$

which implies

$$T(A+B) \neq T(A)+T(B).$$

Theorem 3 *If* $A, B \in \Omega_i$ *(*$i \in \{1,2,3,4\}$*) then*

$$T(A+B) = T(A)+T(B).$$

ISBN: 978-989-95079-6-8

Proof. The equalities

$$I(A+B) = I(A) + I(B),$$
$$S(A+B) = S(A) + S(B),$$
$$L(A+B) = L(A) + L(B),$$
$$U(A+B) = U(A) + U(B), A, B \in F(\mathbb{R})$$

are immediate. Then $A, B \in \Omega_i$ ($i \in \{1,2,3,4\}$) implies $A+B \in \Omega_i$ and $T(A)+T(B) = T(A+B)$ in every case $(i)-(iv)$ in Theorem 1. ■

4 Continuity

An extended trapezoidal fuzzy number ([10]) is an order pair of polynomial functions of degree less than or equal to 1. An extended trapezoidal fuzzy number may be not a fuzzy number, but the distance between two extended trapezoidal fuzzy numbers is similarly defined as in (1) or (6). The extended trapezoidal approximation $T_e(A) = (l_e, u_e; x_e, y_e)$ of a fuzzy number A is defined ([10]) as the extended trapezoidal fuzzy number which minimizes the distance $D(A,B)$, where B is an extended trapezoidal fuzzy number, and preserves the expected interval of A. The real numbers x_e and y_e are non-negative (see [11]). In [5], the authors proved that $T_e(A) = T_3(A)$.

Let us denote by $T(A) = (l_0, u_0; x_0, y_0)$ the nearest (with respect to the metric D) trapezoidal fuzzy number to fuzzy number A, preserving the expected interval, given in Theorem 1, in the form proposed in [10] (see also Section 2). We have ([10])

$$l_0 = l_e = \int_0^1 A^-(\lambda)d\lambda \tag{11}$$

$$u_0 = u_e = \int_0^1 A^+(\lambda)d\lambda. \tag{12}$$

Let us consider in the Euclidean space $\mathbb{R}^2$ (it is a finite dimensional Hilbert space) the points $A_e(x_e, y_e)$, $A_0(x_0, y_0)$ and the set

$$M = \{(x,y) \in \mathbb{R}^2 : x \geq 0, y \geq 0, x + y \leq 2u_e - 2l_e\}$$

which is a closed convex subset of $\mathbb{R}^2$. Then $P_M(A_e)$ is the unique element in M (see e.g. [8], Theorem 4.10, p. 79) which minimizes the Euclidean distance in $\mathbb{R}^2$, $D_E(A_e, P)$, where $P \in M$. Taking into account the remarks in [10] we get

$$A_0 = P_M(A_e). \tag{13}$$

We present the main result of this section.

Theorem 4 *The nearest trapezoidal approximation operator preserving the expected interval $T : F(\mathbb{R}) \to F^T(\mathbb{R})$ is continuous.*

Proof. Let us consider two fuzzy numbers A and B,

$$A_\lambda = \left[A^-(\lambda), A^+(\lambda)\right],$$
$$B_\lambda = \left[B^-(\lambda), B^+(\lambda)\right], \lambda \in [0,1],$$

$T_e(A) = (l_e, u_e; x_e, y_e)$, $T_e(B) = (l'_e, u'_e; x'_e, y'_e)$ the extended trapezoidal approximations of A and B and $T(A) = (l_0, u_0; x_0, y_0)$, $T(B) = (l'_0, u'_0; x'_0, y'_0)$ the trapezoidal approximations preserving the expected interval of A and B. The relations (11) and (12) imply

$$D^2(T(A), T(B)) = (l_0 - l'_0)^2 + (u_0 - u'_0)^2 + \frac{1}{12}(x_0 - x'_0)^2 + \frac{1}{12}(y_0 - y'_0)^2$$
$$= (l_e - l'_e)^2 + (u_e - u'_e)^2 + \frac{1}{12}(x_0 - x'_0)^2 + \frac{1}{12}(y_0 - y'_0)^2$$

Because (see [9], Proposition 4.4)

$$D(T_e(A), T_e(B)) \leq D(A,B) \tag{14}$$

we obtain

$$D^2(T(A), T(B)) \leq D^2(A,B) + \frac{1}{12}(x_0 - x'_0)^2 + \frac{1}{12}(y_0 - y'_0)^2$$

or

$$D^2(T(A), T(B)) \leq D^2(A,B) + \frac{1}{12}D_E^2(A_0, B_0), \tag{15}$$

where $D_E(A_0, B_0)$ denotes the Euclidean distance between $A_0(x_0, y_0)$ and $B_0(x'_0, y'_0)$. Let us assume (contrariwise the proof is similar)

$$2u'_e - 2l'_e \geq 2u_e - 2l_e.$$

We consider

$$M_A = \{(x,y) \in \mathbb{R}^2 : x \geq 0, y \geq 0, x + y \leq 2u_e - 2l_e\},$$
$$M_B = \{(x,y) \in \mathbb{R}^2 : x \geq 0, y \geq 0, x + y \leq 2u'_e - 2l'_e\}$$

and

$$C(2u_e - 2l_e, 0),$$
$$C'(0, 2u_e - 2l_e),$$
$$G(2u'_e - 2l'_e, 0),$$
$$G'(0, 2u'_e - 2l'_e),$$

the points which define the closed convex sets M_A and M_B in the Euclidean space $\mathbb{R}^2$. The pairs (l_e, u_e), (l'_e, u'_e) being already fixed ((11), (12)), we denote $A_e(x_e, y_e)$ and $B_e(x'_e, y'_e)$ the points in $\mathbb{R}^2$ which represent the others components of the extended trapezoidal approximations of A and B, respectively. It is known that $x_e, y_e, x'_e, y'_e \geq 0$.

According to (13) we get

$$A_0 = P_{M_A}(A_e)$$
$$B_0 = P_{M_B}(B_e)$$

We denote by B_1 the projection of B_0 on the convex set M_A, that is the unique element in M_A which minimizes $D_E(B_0, Q)$, where $Q \in M_A$. We prove that B_1 is the

ISBN: 978-989-95079-6-8

projection of B_e on the set M_A, that is $B_1 \in M_A$ and $\min_{R \in M_A} D_E(B_e, R) = D_E(B_e, B_1)$, and

$$D_E(B_1, B_0) \leq D_E(C, G) = D_E(C', G').$$

Indeed, if $x'_e + y'_e \leq 2(u_e - l_e)$ then $B_e \in M_A$. Because $M_A \subseteq M_B$ we get

$$B_e \in M_B,$$
$$B_0 = B_1 = B_e$$

and

$$D_E(B_1, B_0) = 0 \leq D_E(C, G).$$

Contrariwise, if $x'_e + y'_e > 2(u_e - l_e)$ then the following situations are possible:

(i) $x'_e - y'_e > 2(u'_e - l'_e)$. Then

$$B_0 = G,$$
$$P_{M_A}(B_e) = C$$

and

$$B_1 = P_{M_A}(B_0) = P_{M_A}(G) = C.$$

In addition,

$$D_E(B_1, B_0) = D_E(C, G).$$

(ii) $2(u_e - l_e) \leq x'_e - y'_e \leq 2(u'_e - l'_e)$. Then

$$B_1 = P_{M_A}(B_e) = P_{M_A}(B_0) = C$$

and

$$D_E(B_1, B_0) = D_E(C, B_0) \leq D_E(C, G).$$

(iii) $-2(u_e - l_e) < x'_e - y'_e < 2(u_e - l_e)$. Because $B_e B_0$ is perpendicular on GG', $B_e B_1$ is perpendicular on CC' and GG', CC' are parallel we get

$$B_1 = P_{M_A}(B_e).$$

In addition,

$$D_E(B_1, B_0) < D_E(C, G).$$

(iv) $-2(u'_e - l'_e) \leq x'_e - y'_e \leq -2(u_e - l_e)$. Then

$$B_1 = P_{M_A}(B_e) = P_{M_A}(B_0) = C'$$

and

$$D_E(B_1, B_0) = D_E(C', B_0) \leq D_E(C', G').$$

(v) $x'_e - y'_e < -2(u'_e - l'_e)$. Then

$$B_0 = G',$$
$$P_{M_A}(B_e) = C'$$

and

$$B_1 = P_{M_A}(B_0) = P_{M_A}(G') = C'.$$

In addition,

$$D_E(B_1, B_0) = D_E(C', G').$$

We have

$$\begin{aligned} & D_E^2(C, G) \\ &= (2u'_e - 2l'_e - 2u_e + 2l_e)^2 \\ &= 4\left[(u'_e - u_e) - (l'_e - l_e)\right]^2 \\ &= 4\left[\int_0^1 \left(B^+(\lambda) - A^+(\lambda)\right) d\lambda \right. \\ &\quad \left. - \int_0^1 \left(B^-(\lambda) - A^-(\lambda)\right) d\lambda\right]^2 \\ &\leq 8\left[\int_0^1 \left(B^+(\lambda) - A^+(\lambda)\right) d\lambda\right]^2 \qquad (16) \\ &\quad + 8\left[\int_0^1 \left(B^-(\lambda) - A^-(\lambda)\right) d\lambda\right]^2 \\ &\leq 8\int_0^1 \left(B^+(\lambda) - A^+(\lambda)\right)^2 d\lambda \qquad (17) \\ &\quad + 8\int_0^1 \left(B^-(\lambda) - A^-(\lambda)\right)^2 d\lambda \end{aligned}$$

therefore

$$D_E^2(B_1, B_0) \leq 8D^2(A, B).$$

Because M_A is a closed convex subset of R^2 we obtain (see [11], Appendix C)

$$D_E(P_{M_A}(A_e), P_{M_A}(B_e)) \leq D_E(A_e, B_e)$$

that is

$$D_E(A_0, B_1) \leq D_E(A_e, B_e).$$

We get

$$\begin{aligned} D_E(A_0, B_0) &\leq D_E(A_0, B_1) + D_E(B_1, B_0) \\ &\leq D_E(A_e, B_e) + 2\sqrt{2}D(A, B). \end{aligned}$$

From (14) we obtain

$$D_E^2(A_e, B_e) \leq 12D^2(A, B), \qquad (18)$$

therefore

$$D_E(A_0, B_0) \leq 2(\sqrt{2} + \sqrt{3})D(A, B). \qquad (19)$$

The inequalities (15) and (19) imply

$$D(T(A), T(B)) \leq \sqrt{1 + \frac{(\sqrt{2} + \sqrt{3})^2}{3}} D(A, B) \qquad (20)$$

and the proof is complete. ■

If $D(A, B) = 0$ then (20) becomes equality.

If

$$D(T(A), T(B)) = \sqrt{1 + \frac{(\sqrt{2} + \sqrt{3})^2}{3}} D(A, B)$$

ISBN: 978-989-95079-6-8

then (16)-(18) must be equalities (contrariwise, inequality (20) becomes strict). Equality in (16) implies

$$\int_0^1 \left(B^+(\lambda) - A^+(\lambda)\right) d\lambda + \int_0^1 \left(B^-(\lambda) - A^-(\lambda)\right) d\lambda = 0 \quad (21)$$

and equality in (17) implies

$$B^+(\lambda) - A^+(\lambda) = k^+, \text{ a.e. } \lambda \in [0,1]$$

and

$$B^-(\lambda) - A^-(\lambda) = k^-, \text{ a.e. } \lambda \in [0,1],$$

where $k^+, k^- \in \mathbb{R}$. Substituting in (21) we get $k^+ + k^- = 0$, therefore

$$B^+(\lambda) = A^+(\lambda) + k, \text{ a.e. } \lambda \in [0,1]$$

and

$$B^-(\lambda) = A^-(\lambda) - k, \text{ a.e. } \lambda \in [0,1],$$

where $k \in \mathbb{R}$. According to (7)-(10) we obtain

$$I(B) = I(A) - k,$$
$$S(B) = S(A) + k,$$
$$L(B) = L(A) - \frac{k}{2},$$

and

$$U(B) = U(A) + \frac{k}{2}.$$

Taking into account Theorem 1 (iii) and (4), (5),

$$x_e^{'} = -6I(B) + 12L(B) = -6I(A) + 12L(A) = x_e,$$

$$y_e^{'} = 6S(B) - 12U(B) = 6S(A) - 12U(A) = y_e.$$

We get $A_e = B_e$ and equality in (18) implies $D(A,B) = 0$.

We conclude that equality in (20) holds if and only if $D(A,B) = 0$.

5 Conclusion

Many approximation methods for fuzzy numbers were proposed in the last years. Because the quality of approximation is important, a list of criteria that the approximation operator should or just can possess has been given in [5]. Continuity and additivity of nearest trapezoidal approximation operator which preserves the expected interval are studied in the present paper. Because the property of scale invariance was already proved in [1], the property of partial additivity, given in Theorem 3, assures the linearity of the trapezoidal approximation operator, when the fuzzy numbers under study are all with the same kind of asymmetry of membership functions (see [6]). The property of continuity, given in Theorem 4, means that if two fuzzy numbers are close, with respect to metric D, then their approximations are also close, with respect to metric D. Continuity is of extreme importance in applications, especially in fuzzy control systems, where it is sometimes indicated as robustness and discontinuous approximation operators seem unnatural.

References

[1] A. Ban. Approximation of fuzzy numbers preserving the expected interval. *Fuzzy Sets and Systems*, 159:1327-1344, 2008.

[2] A. Ban and L. Coroianu. Continuity of trapezoidal approximation operators, submitted.

[3] D. Dubois and H. Prade. The mean value of a fuzzy number. *Fuzzy Sets and Systems*, 24:279-300, 1987.

[4] P. Grzegorzewski. Metrics and orders of fuzzy numbers. *Fuzzy Sets and Systems*, 97:83-94, 1998.

[5] P. Grzegorzewski and E. Mrówka. Trapezoidal approximations of fuzzy numbers. *Fuzzy Sets and Systems*, 153:115-135, 2005.

[6] P. Grzegorzewski. Trapezoidal approximations of fuzzy numbers preserving the expected interval-Algorithms and properties. *Fuzzy Sets and Systems*, 159:1354-1364, 2008.

[7] S. Heilpern. The expected value of a fuzzy number. *Fuzzy Sets and Systems*, 47:81-86, 1992.

[8] W. Rudin. *Real and Complex Analysis*. McGraw-Hill, New York, 1986.

[9] C.-T. Yeh. A note on trapezoidal approximations of fuzzy numbers. *Fuzzy Sets and Systems*, 158:747-754, 2007.

[10] C.-T. Yeh. Trapezoidal and triangular approximations preserving the expected interval. *Fuzzy Sets and Systems*, 159:1345-1353, 2008.

[11] C.-T. Yeh. On improving trapezoidal and triangular approximations of fuzzy numbers. *International Journal of Approximate Reasoning*, 48:297-313, 2008

ISBN: 978-989-95079-6-8

A baseline learning genetic fuzzy classifier based on low quality data

Ana M. Palacios[1] Luciano Sánchez[1] Inés Couso[2]

1. Departamento de Informática, Universidad de Oviedo, Gijón, Asturias, Spain
2. Departamento de Estadística e I.O. y D.M, Universidad de Oviedo, Gijón, Asturias Spain
Email: apalaciosjimenez@hotmail.com, luciano@uniovi.es, couso@uniovi.es

***Abstract*— Obtaining fuzzy rules from low quality data is a topic that has been recently formalized. This paper contains the first application of these principles to classification problems. We intend that the classifier proposed here serves as a baseline for future developments in the field. For that reason, we have extended a simple crisp genetic fuzzy classifier to imprecise data, paying special attention to the computational details. In particular, we will discuss some issues about the fuzzy-valued fitness function that is used in our formalism. A synthetic problem, plus two real-world datasets of low and medium complexities are also proposed, and used to evaluate the algorithm.**

1 Introduction

The term "low quality data" refers to datasets where some or all of the features are imprecisely perceived. It embodies different concepts: censored data, binning, missing values, significant digits, confidence intervals, contradictory information from different sources, fuzzy numbers, linguistic information, and others.

We are interested in those cases where the imprecision in the perception of the data is defined by a family of confidence intervals. This includes most of the preceding situations as particular cases. For instance, we can model few significant digits, censored or missing data by means of a single interval that spans the range of the unknown measure: "the weight of an object is in $[1.1, 1.2]$" or "the pressure is lower than 3". We can also reconcile different measurements of the same feature by means of a set of confidence intervals [8]. Certain (but not all) kinds of fuzzy data can also be regarded as imprecise perceptions of crisp values, known through a family of α-cuts [1].

To make clear the extent of this concept, let us recall the statistical framework of a standard Artificial Intelligence-based classifier. The purpose of a classifier is to predict the class of an object, given the values of other properties of the same object. To that end, a probability distribution is defined on the set of objects, and the mentioned properties are assumed to be random variables. The classifier, in turn, is a decision rule that depends on the posterior probability of each class, given a vector of properties.

The algorithms in this paper differ from the standard case because we do not suppose that we can accurately observe all the properties of the object. Otherwise, the same assumptions hold. In the most simple case (interval-valued data) we will perceive sets that contain these values. In the general case, we will be given a nested family of sets, each one of them containing the true value with certain probability. From a theoretical point of view, we hence understand that low quality data is a kind of data for which we can not achieve a precise knowledge about the posterior probability distributions induced by the mentioned random variables, but we can obtain families of probability distributions that are compatible with them. Accordingly, the mentioned decision rule can be inconclusive if the input data is not specific enough.

Recent works in fuzzy statistics suggest using a fuzzy representation when the data is known through a family of confidence intervals [1]. This representation assumes that a fuzzy set can be interpreted as a possibility distribution (which, in turn, is a family of probability distributions) and each α-cut of a fuzzy feature is a random set that contains the unknown crisp value of the feature with probability $1 - \alpha$. [7, 8].

1.1 GFS and Fuzzy fitness functions

Our particular fuzzy representation of low quality data is tied to the use of a fuzzy or interval-valued measurement of the accuracy of a classifier or model. Let us use an example: we have a classification system, defined by these rules:

$$\begin{array}{l}\text{if } x < 1 \text{ then class is } A\\ \text{if } x \in [1,2] \text{ then class is } B\\ \text{if } x > 2 \text{ then class is } C\end{array} \quad (1)$$

and the input that follows:

$$x < 1.8 \quad (2)$$

The output of the classifier is the set of classes $\{A, B\}$. If the object being classified is of class C, we know that the classifier has failed. Otherwise, we cannot know. Nonetheless, we can use a set-valued variable "number of errors", and state that the error of the classifier in that example is the set $\{0, 1\}$. The number of errors of the whole classifier can be obtained by adding these individual errors with interval arithmetic operators. If the output of the classifier is a fuzzy set, the number of errors is a fuzzy number too, and we must use fuzzy arithmetic [6].

Roughly speaking, estimating a classifier from data requires a numerical technique that finds the minimum of the classification error with respect to the free parameters of the classifying system. In our case, this function is interval-valued or fuzzy. But there are not many techniques for optimizing interval-valued or fuzzy valued functions. In the genetic algorithms field, the solutions are related to precedence operators between imprecise values [3, 4, 11]. We have previous works where we have jointly optimized a mix of crisp and fuzzy objectives with genetic algorithms [7]. We have also proposed a number of different algorithms for learning regression models from low quality data and the fuzzy representation mentioned before [5, 10, 8]. However, to the best of our knowledge there

ISBN: 978-989-95079-6-8

```
function GFS
1    Initialize population
2    for iter in {1, ..., Iterations}
3      for sub in {1, ..., subPop}
4        Select parents
5        Crossover and mutation
6        assignConsequent(offspring)
7      end for sub
8      Replace the worst subPop individuals
9      assignFitness(population,dataset)
10   end for iter
11   Purge unused rules
return population
```

Figure 1: Outline of the GFS that will be generalized [2]. Each chromosome codifies one rule. The fitness of the classifier is distributed among the rules at each generation.

```
function assignConsequent(rule)
1    for example in {1, ..., N}
2      m = membership(Antecedent,example)
3      weight[class[example]] = weight[class[example]] + m
4    end for example
5    mostFrequent = 0
6    for c in {1, ..., N_c}
7      if (weight[c]>weight[mostFrequent]) then
8        mostFrequent = c
9      end if
10   end for c
11   Consequent = mostFrequent
return rule
```

Figure 2: The consequent of a rule is not codified in the GA, but it is assigned the most frequent class label, between those compatible with the antecedent of the rule [2].

have not been previous GFSs where those principles have been applied to learn classification problems.

1.2 Summary

The structure of this paper is as follows: in the next section we generalize the crisp GFS defined in [2] to low quality data. In Section 3 we evaluate the generalized algorithm in both crisp and imprecise datasets, and propose two real-world datasets of classification with imprecise data. The paper finishes with the concluding remarks, in Section 4.

2 Generalizing a Genetic Fuzzy Classifier to imprecise data

Generalizing a GFS to imprecise data involves, at the very least, changes in the inference mechanism and the fitness function, as we have discussed in [6]. In this section we will generalize the GFS outlined in Figure 1, which was introduced in [2]. This is a very compact algorithm that allows us to focus in the subject of this paper (extending Genetic Fuzzy Classifiers to imprecise data) without getting lost in the details.

Observe that this algorithm depends on two functions: "assignConsequent" (line 6) and "assignFitness" (line 9). These functions are also listed in Figures 2 and 3. This algorithm does not codify the consequent of the fuzzy rules in the genetic individual. Instead, the function "assignConsequent"

```
function assignFitness(population,dataset)
1    for example in {1, ..., N}
2      winnerRule = 0
3      bestMatch = 0
4      for rule in {1, ..., M}
5        m = membership(Antecedent[rule],example)
6        if (m>bestMatch) then
7          winnerRule = rule
8          bestMatch = m
9        end if
10     end for rule
11     if (consequent(winnerRule)==class(example)) then
12       fitness[winnerRule] = fitness[winnerRule] + 1
13     end if
14   end for example
return fitness
```

Figure 3: The fitness of an individual is the number of examples that it classifies correctly. Single-winner inference is used, thus at most one rule changes its fitness when the rule base is evaluated in an example [2].

determines the class label that matches an antecedent with a maximum confidence. The function "assignFitness," in turn, determines the winner rule for each object in the training set and increments the fitness of the corresponding individual if its consequent matches the class of the object. In the remainder of this section, we will study the impact of the imprecise knowledge about the independent and dependent variables in the structure of the rules, and how to extend these two functions. That is to say: we analyze the reasoning method, the assignment of the consequents, the computation of a set-valued fitness and the genetic selection and replacement of the worst individuals.

2.1 Analysis of the reasoning method

The objective of the extended GFS is to obtain a fuzzy rule base from objects, when there is imprecise knowledge about some or all of the attributes of these objects. The set of classes that is produced when the input value is a fuzzy set can be computed in some different ways, but not all of them are consistent with our representation of an imprecise value.

Having an imprecise knowledge about the input variables differs from the standard case because:

1. The output of the FRBS will not be completely determined.

2. The number of errors of the FRBS in the training data will be partially known. The same happens if any other quality function is used instead of the number of errors, i.e. likelihood, logistic loss functions, etc.

Both issues have been introduced with the crisp classifier commented in the introduction: if the input is a set X that contains a range of inputs $x \in X$, the output is not a class but a set of classes:

$$\text{class(X)} = \{\text{class}(x) \mid x \in X\}. \tag{3}$$

The first difference is apparently a trivial issue: after all, one of the advantages of fuzzy rules is dealing with imprecision. Nevertheless, the standard reasoning method does not produce

ISBN: 978-989-95079-6-8

the set of classes that we need. To make our point clearer, let us study a fuzzy classifier comprising M rules:

$$\textbf{if } (x \textbf{ is } \widetilde{A}_i \,) \textbf{ then class is } C_i, \tag{4}$$

and let us use the single-winner inference mechanism:

$$\text{class(x)} = C_{\arg\max_i \{\widetilde{A}_i(x)\}}. \tag{5}$$

Observe that, classifying the set X by means of the standard mechanism produces the class label that follows:

$$\text{class'(X)} = C_{\arg\max_i \{\min\{\widetilde{A}_i(x) | x \in X\}\}} \tag{6}$$

while we need this set of labels:

$$\text{class(X)} = \{C_{\arg\max_i \{\widetilde{A}_i(x)\}} \mid x \in X\} \tag{7}$$

which is different than 6. Furthermore, the fuzzy set of classes obtained when aggregating the rules, before the defuzzification stage, is not what we need either, because our set class(X) does not coincide with the fuzzy set

$$\sum C / \max\{\min\{\widetilde{A}_i(x) \mid x \in X\} | C_i = C\}. \tag{8}$$

The code we propose to use is included in lines 2–23 in Figure 5, that we will explain later.

It is remarked that, in this paper, each rule will contain a single consequent. This is a result of our interpretation of "low quality" as a family of confidence intervals. In words: in this paper, if a point is labeled as "class {A,C}" we are not stating that it belongs to both categories at the same time (which is not an imprecise assert). We are expressing that we are not sure about the class of the object, i.e. we only know that it is not in class "B". Therefore, it makes not sense in this context to produce a rule with a double consequent, since this rule necessarily will have non-zero error at any example.

2.2 Assignment of consequents

```
function assignImpreciseConsequent(rule)
1    for example in {1,...,N}
2      m̃ = fuzMembership(Antecedent,example)
3      weight[class[example]] = weight[class[example]] ⊕ m̃
4    end for example
5    mostFrequent = {1,...,Nc}
6    for c in {1,...,Nc}
7      for c1 in {c+1,...,Nc}
8        if (weight[c] dominates weight[c1]) then
9          mostFrequent = mostFrequent - { c1}
10       end if
11     end for c1
12   end for c
13   Consequent = select(mostFrequent)
return rule
```

Figure 4: If the examples are imprecise, we might not know the most frequent class label –lines 5 to 12–. In this paper we have used the dominance proposed in [4] to reduce this set to one element.

The assignment of consequents seen in Figure 2 is extended in Figure 4. The original assigment consists in computing the confidences of the rules "if (x is $\widetilde{A}$) then class is C" for all the values of "C", then selecting the alternative with maximum confidence. In this case, the confidence of a rule is a set of values. The operation "dominates" used in line 8 can have different meanings, ranging from the strict dominance (A dominates B iff $a < b$ for all $a \in A, b \in B$) [11] to other definitions that induce a total order in the set of confidences. Generally speaking, we have to select one of the values in the set of nondominated confidences and use its corresponding consequent. In this paper, we have used the uniform dominance defined in [4], that induces a total order and thus the set of nondominated consequents has size 1.

2.3 Computation of fitness

```
function assignImpreciseFitnessApprox(population,dataset)
1    for example in {1,...,N}
2      setWinnerRule = ∅
3      for r in {1,...,M}
4        dominated = FALSE
5        rule.m̃ = fuzMembership(Antecedent[r],example)
6        for sRule in setWinnerRule
7          if (sRule dominates rule) then
8            dominated = TRUE
9          end if
10       end for sRule
11       if (not dominated) then
12         for sRule in setWinnerRule
13           if (m̃ dominates sRule) then
14             setWinnerRule = setWinnerRule −{ sRule }
15           end if
16         end for sRule
17         setWinnerRule = setWinnerRule ∪{ rule }
18       end if
19     end for r
20     setOfCons= ∅
21     for sRule in setWinnerRule
22       setOfCons= setOfCons ∪{ consequent(rule) }
23     end for sRule
24     deltaFit= 0
25     if ({class(example)} == setOfCons and
           size(setOfCons)==1) then
26       deltaFit = {1}
27     else
28       if ({class(example)}∩ setOfCons ≠ ∅) then
29         deltaFit = {0, 1}
30       end if
31     end if
32     Select winnerRule ∈ setWinnerRule
33     fitness[winnerRule] = fitness[winnerRule] ⊕ deltaFit
34   end for example
return fitness
```

Figure 5: Generalization of the function "assignFitness" to imprecise data. If the example is imprecisely perceived, there are three ambiguities that must be resolved: (a) some different crisp values compatible the same example might correspond to different winner rules –lines 3 to 19—, (b) these rules might have different consequent, thus we do not know if the rule base fails in the example –lines 20 to 31– and (c) we must assign credit to just one of these rules –lines 32 and 33–.

The output of the FRBS at the i-th object of the training set

ISBN: 978-989-95079-6-8

is a set of classes:

$$C_{\text{FRBS}}(X_i) = \{C_{\arg\max_j\{\widetilde{A}_j(x)\}} \mid x \in X_i\}. \quad (9)$$

The theoretical expression of the fitness function of the FRBS is:

$$\text{fitness} = \bigoplus e_i \quad (10)$$

where

$$e_i = \begin{cases} 1 & C_{\text{FRBS}}(X_i) = C_i \text{ and } \#(C_i) = 1 \\ 0 & C_{\text{FRBS}}(X_i) \cap C_i = \emptyset \\ \{0,1\} & \text{else} \end{cases} \quad (11)$$

In words, if the output of the FRBS is a single class label that matches the class label of the example, this point scores 1. If the set of classes emitted by the FRBS does not intersect with that of the object, this point scores 0. Otherwise, it scores the set $\{0,1\}$.

The evaluation of this function is computationally very expensive, and we will use an approximation, described in Figure 5. This algorithm computes an interval of values of matching between each rule and the input, then discards all rules that can not be the winner rule, and approximates the output of the FRBS by the set of the consequents of the non-discarded rules. This set includes the theoretical output, but sometimes it also includes extra class labels. In Figure 6 we have also included a more accurate approximation which is based on a sample of values of the support of the input. This second approximation will be used in the next section to better determine the quality of a classifier, but our learning will be guided by the function in Figure 5, because of its lower cost.

2.4 Genetic selection and replacement

There are two other parts in the original algorithm that must be altered in order to use an imprecise fitness function: (a) the selection of the individuals in [2] is based on a tournament, that depends on a total order on the set of fitness values. And (b) the same happens with the removal of the worst individuals. In both cases, we have used the uniform dominance defined in [4] to impose such a total order. We leave for future works the application of a multicriteria genetic algorithm similar to those used in our previous works in regression modeling [8, 10].

3 Numerical results

This section contains a brief numerical analysis of the generalized algorithm. We have performed the experiments that follow:

1. Synthetic datasets: Gaussian distribution, known Bayesian error, and different amounts of observation error.

2. Crisp datasets: three standard benchmarks, for testing that the extended algorithm has the same performance as the original version in crisp problems.

3. Imprecise, real world datasets: we propose two real world datasets, of small and medium size. One of them has been specifically designed for the purpose of this research, and the other is part of a practical problem of medical diagnosis.

```
function assignImpreciseFitnessExhaustive(population,dataset)
1     for example in {1,...,N}
2       S = sample(example)
3       maxScore = 0
4       for s in S
5         winnerRule = 0
6         bestMatch = 0
7         for rule in {1,...,M}
8           m = membership(Antecedent[rule],s)
9           if (m > bestMatch) then
10            winnerRule = rule
11            bestMatch = m
12          end if
13        end for rule
14        if (consequent(winnerRule) == class(example)) then
15          score[winnerRule] = score[winnerRule] ⊕ 1
16        elif (consequent(winnerRule) ⊂ class(example)) then
17          score[winnerRule] = score[winnerRule] ⊕ {0,1}
18        end if
19        if (max(score[winnerRule]) > max(score[maxScore]))
20          then maxScore = winnerRule
21        end if
22      end for s
23      if (score[maxScore] > 0) then
24        if (score[maxScore] == size(S)) then
25          fitness[maxScore] = fitness[maxScore] ⊕ 1
26        else
27          fitness[maxScore] = fitness[maxScore] ⊕ {0,1}
28        end if
29      end if
30    end for example
return fitness
```

Figure 6: Other generalization of the function "assignFitness" to interval-valued data. This function is computationally too expensive for being used as a fitness function; it will be used instead for obtaining better estimations of the train and test errors of the final rule bases. Lines 14–18 deal with the case where an object has imprecise output, i.e. "the class is A or C"; otherwise, the value of the variable "score" is crisp.

All the datasets used in this paper are available in the website of the KEEL project: `http://www.keel.es`. All the experiments have been run with a population size of 100, probabilities of crossover and mutation of 0.9 and 0.1, respectively, and limited to 200 generations. The fuzzy partitions of the labels are uniform and their size is 3, except when mentioned otherwise.

3.1 Synthetic datasets

The set "Gaussian" comprises 699 points of two classes. The distribution of both classes is bidimensional Gaussian, with unity covariance matrix, and centered in $(0,0)$ and $(3,0)$ respectively. To this data we have added interval-valued imprecision of sizes $\beta = 0.03, 0.05, 0.1, 0.2, 0.5$. A 10-cv experimental design was applied, and the mean values of the test errors are shown in Table 1. The training error has been also included, to show the differences between the approximation of the fitness function seen before and the exhaustive computation that has been used to compute the test error. Observe that the approximate error computed by the fitness function is less specific than the actual error, and the difference is relevant

ISBN: 978-989-95079-6-8

	Crisp			Low Quality		
β	Theoretical	Train	Test	Exh. Train	Exh. Test	Approx. Train
0	0.084	0.083	0.086	[0.086,0.086]	[0.082,0.082]	[0.086,0.086]
0.03				[0.047,0.086]	[0.083,0.094]	[0.076,0.091]
0.05				[0.075,0.089]	[0.081,0.098]	[0.071,0.094]
0.1				[0.070,0.103]	[0.068,0.104]	[0.076,0,093]
0.2				[0.052,0.116]	[0.055,0.128]	[0.075,0.089]
0.5				[0.022,0.183]	[0.022,0.179]	[0.014,0.225]

Table 1: Results of the extended GFS in the synthetic dataset "Gauss" for crisp data (first column) and different degrees of observation error (second column). The approximate error computed by the fitness function is less specific than the actual error, and the difference is relevant when the observation error is high ($\beta = 0.2$ and $\beta = 0.5$, nevertheless it still guides the evolution correctly.

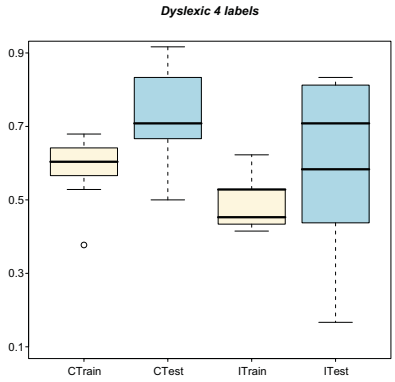


Figure 7: Boxplots illustrating the dispersion of the 10 repetitions of crisp and extended GFS in the problem "dyslexia-12", with 4 labels/partition. The boxplot of the imprecise experiments is not standard: we show respectively the 75% of the maximum and 25% percentile of the minimum fitness, thus the box displays at least the 50% of data; there are two marks inside the box, because the median of the data is an interval.

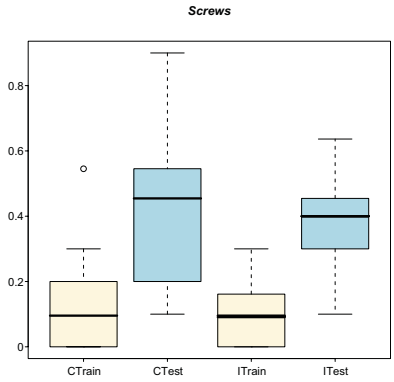


Figure 8: Boxplots illustrating the dispersion of the 10 repetitions of crisp and extended GFS in the problem "screws-50".

when the observation error is high ($\beta = 0.2$ and $\beta = 0.5$), nevertheless it still guides the evolution correctly.

3.2 Crisp datasets

Crisp datasets are included for assessing the performance of the generalized algorithm in standard problems. Since, in this case, the particularization of the algorithm to crisp data recovers the original algorithm in [2], the results are expected to be adequate. Nonetheless, we have included some estimations in Table 2. As expected, the results of both algorithms are similar; the differences are originated in the different random seeds.

3.3 Real world datasets

We propose two datasets for testing this and future learning algorithms with low quality data:

1. Dataset "Screws-50": 21 objects, 3 classes, 2 features (weight and length). We have weighed and measured 21 screws of three different types, taking into account the accuracy of the physical measurement: each feature is an interval. The class labels are precise. There are not outliers, i.e. 0% of error is attainable in absence of imprecision.

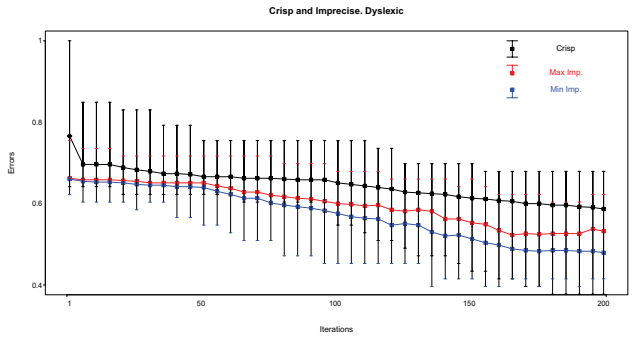


Figure 9: Compared evolution of crisp and imprecise GFS in the dataset "dyslexia-12". The ranges and means of 10 repetitions of the learning are shown, for both the crisp and the imprecise versions of the algorithm. The upper bound of the mean imprecise fitness is consistently lower than the mean of the crisp fitness.

2. Dataset "Dyslexia-12": 65 objects, 4 classes, 12 features. This is a selection of the original dataset described in [9], where the 12 most relevant variables have been hand-picked by a psychologist. There are imprecision in both the input and the output. The theoretical error is

ISBN: 978-989-95079-6-8

	Crisp		Low Quality		
Dataset	Train	Test	Exh. Train	Exh. Test	Approx. Train
Pima	0.254	0.287	[0.258,0.258]	[0.288,0.288]	[0.258,0.258]
Glass	0.323	0.365	[0.321,0.321]	[0.352,0.352]	[0.321,0.321]
Haberman	0.239	0.255	[0.238,0.238]	[0.248,0.248]	[0.238,0.238]

Table 2: Results in some crisp benchmarks, where the imprecise fitness function reduces to the crisp fitness function. The results of "crisp" and "low quality" columns are similar.

Crisp			Low Quality		
Dataset	Train	Test	Exh. Train	Exh. Test	Approx. Train
Screws-50	0.133	0.427	[0.096,0.096]	[0.377,0.377]	[0.068,0.106]
Dyslexia-12 (4 labels)	0.584	0.724	[0.481, 0.616]	[0.541, 0.675]	[0.477,0.516]
Dyslexia-12 (5 labels)	0.745	0.658	[0.581,0.667]	[0.608,0.641]	[0.537,0.549]

Table 3: Means of 10 repetitions of the generalized GFS for the imprecise datasets "Screws-50" and "Dyslexia-12" with 4 and 5 labels/variable

unknown.

We have compared the performance of the generalized algorithm to that of the original crisp algorithm. To that end, we have built a crisp dataset by removing the uncertainty in the imprecise dataset: each imprecise measurement was replaced by the mid-point of the corresponding interval, and those examples with imprecision in the independent variable were replicated for the different options. For instance, a point $(X = [1,3], C = \{A,B\})$ is converted into two points $(x = 2, c = A)$, $(x = 2, c = B)$.

We have used a 5x2cv design for the first problem, because of its small size, and 10cv for the second. The boxplots of the compared results, in both train and test sets, are depicted in Figures 7 and 8. Observe that the boxplots of the imprecise experiments are not standard. We propose using a box showing the 75% of the maximum and 25% percentile of the minimum fitness (thus the box displays at least the 50% of data) and also drawing two marks inside the box, because the median of the data is an interval. In Figure 9 the ranges and means of 10 repetitions of the learning are shown, for both the crisp and the imprecise versions of the algorithm. The upper bound of the mean imprecise fitness is consistently lower than the mean of the crisp fitness.

4 Concluding remarks

Extending a GFS to imprecise data in classification problems is based on the use of an interval or fuzzy valued fitness function. Most GFSs can be extended to low quality data if some changes are made in their reasoning method, and the genetic algorithm can deal with an imprecisely known fitness function. We have shown in detail how to apply this changes to a simple GCCL-type algorithm, and evaluated it with some synthetic and real-world benchmarks. The numerical results are as expected for an elementary algorithm like this; there is room for improvement and future works will address more complex GFSs that are based on a multicriteria fitness function.

Acknowledgements

This work was supported by the Spanish Ministry of Education and Science, under grants TIN2008-06681-C06-04, TIN2007-67418-C03-03, and by Principado de Asturias, PCTI 2006-2009.

References

[1] Couso, I., Sánchez, L. Higher order models for fuzzy random variables. Fuzzy Sets and Systems 159: 237–258 (2008)

[2] Ishibuchi, H., Nakashima, T., Murata, T, A fuzzy classifier system that generates fuzzy if-then rules for pattern classification problems. In Proc. of 2nd IEEE International Conference on Evolutionary Computation, 759-764 (1995)

[3] Koeppen, M., Franke, K., and Nickolay, B., Fuzzy-Pareto-Dominance driven multi-objective genetic algorithm. in Proc. 10th International Fuzzy Systems Assotiation World Congress (IFSA), Istanbul, Turkey, 2003: 450–453. (2003)

[4] Limbourg, P., Multi-objective optimization of problems with epistemic uncertainty. in EMO 2005: 413–427. (2005)

[5] Sánchez, L., Otero, J., Villar, J. R., Boosting of fuzzy models for high-dimensional imprecise datasets. Proc. IPMU 2006, Paris, France: 1965–1973. (2006)

[6] Sánchez L., Couso I. Advocating the use of imprecisely observed data in genetic fuzzy systems IEEE Transactions on Fuzzy Systems 15 (4): 551–562. (2007)

[7] Sánchez, L., Couso, I., Casillas, J. Modelling vague data with genetic fuzzy systems under a combination of crisp and imprecise criteria Proc. 2007 IEEE Symp. on Comp. Int. in Multicriteria Decision Making, Honolulu, USA: 30–37. (2007)

[8] Sánchez, L., Couso, I., Casillas, J. Genetic Learning of Fuzzy Rules based on Low Quality Data. Fuzzy Sets and Systems. Submitted.

[9] Sánchez, L., Palacios, A., Couso, I., A Minimum Risk Wrapper Algorithm for Genetically Selecting Imprecisely Observed Features, applied to the Early Diagnosis of Dyslexia. Lecture Notes in Computer Science 5271, 608–615 (2008)

[10] Sánchez, L., Otero, J., Couso, I., Obtaining linguistic fuzzy rule-based regression models from imprecise data with multiobjective genetic algorithms. Soft Computing. 13(5): 467–479 (2008)

[11] Teich, J., Pareto-front exploration with uncertain objectives. in EMO, 2001: 314–328. (2001)

ISBN: 978-989-95079-6-8

Pattern Recognition in Blur Motion Noisy Images using Fuzzy Methods for Response Integration in Ensemble Neural Networks

M. Lopez[1, 2] P. Melin[2] O. Castillo[2]

[1] PhD Student of Computer Science in the Universidad Autonoma de Baja California, Tijuana, B. C., México.
[2] Computer Science in the Graduate Division Tijuana, Institute of Technology Tijuana, B. C., Mexico,
Email: danym23@aol.com, pmelin@tectijuana.mx, ocastillo@tectijuana.mx

Abstract—*Linear Blur Motion is one of the most common degradation functions that corrupt images. Since 1976 many researchers have tried to estimate blur motion parameters and this problem can be solved for noise free images but in the case of noisy images this can be done when the image SNR is low. In this paper, we consider pattern recognition with ensemble neural networks for the case of fingerprints; we propose the use of fuzzy methods for Response Integration in Ensemble Neural Networks for blur motion noisy images. An ensemble neural network of three modules is used; each module is a local expert on person recognition based on a biometric measure (the fingerprints). The Response Integration method of the ensemble neural networks has the goal of combining the responses of the modules to improve the recognition rate of the individual modules when the SNR rate blur motion signal increases to a high level.*

Keywords— Pattern Recognition, Ensemble Neural Networks, Fuzzy Logic, Ratio SNR.

1 Introduction

Blur Motion occurs when there is relative motion between the camera and the object being captured [1]. When a changing scene is observed by a camera, all the classical algorithms assume that is possible to take pictures every δt instant, which means that every picture is taken with a dt ≈ 0 exposure time. If that is not the case, then the exposure time (dt = T) is large enough that different points in the scene are moving far enough and consequently their corresponding projections on the image plane travel several pixels. Therefore, during the capture process of an image, at any single image point, a certain number of scene points is projected during the exposure time, each one contributing to the final brightness of the image point; this effect is shown in figure 1. More formally, during the exposure time T in front of the pixel $P_{i,j}$ we could assume that they pass k scene points with brightness (C_1….C_k) respectively, then the resulting brightness value for pixel $P_{i,j}$ is given in equation (1), in the case it continues movement the summation is replaced by integration. This holds in general for every pixel that is moving points in the scene. It is clear that the blurring of the image exists only across the direction of the motion; this one dimensional blur is called Blur Motion.

$$P_{i,j} = \frac{1}{K}\sum_{l=1}^{k} C_l \qquad (1)$$

The Result of blur motion is shown in figure 1b where an image consistent of different value pixels is shown in figure 1a and then the blurred image is shown in figure 1b.

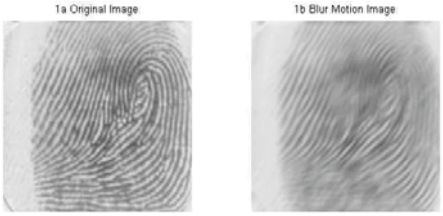


Figure 1: a) Original Image without noise, b) Blur Motion with 20 pixels distance d and 50 degrees angle α.

The blur motion can be described mathematically as the result of a linear filter b(x, y) = i(x, y)*h(x, y) where i is the theoreticalcal image taken with an exposure time T_e=0, b **is** the real blurred image and h the point spread function (PSF**).** Given an angle = α and the length d = V_o x T_e, which is the number of scene points that affect a specific pixel, the point spread function of blur motion is given in equation (2).

$$h(x,y) = \begin{cases} \frac{1}{d}, 0 \le |x| \le d * \cos(\alpha)\ \ y = \sin(\alpha) * d \\ 0, \text{otherwhise} \end{cases} \qquad (2)$$

In practical terms, this mean computing accurate estimates for the two parameters of the blur motion PSF, namely the length, d, and the angle, α. From these quantities, the relative velocity at this point can be easily recovered knowing the exposure time. In this paper for the experimental results we use fingerprint images and adding different levels of blur motion noise from 10 to 90 pixels in distance d and the angle α is zero. The measurement for noise in an image is the signal-to-noise ratio, SNR. The SNR measures the relative strength of the signal in a blurred and noisy image to the strength of the signal in a blurred image without noise. A SNR of 16 db is a low noise level, while a SNR of 4 db is a high noise level [1].

2 Noise and Registration Error

2.1 Signal-to-noise ratio

Noise is present in all digital imaging systems due to a number of sources such as photon shot noise, readout noise, dark current noise and quantization noise. While some sources can be effectively suppressed, such as the dark

ISBN: 978-989-95079-6-8

current noise by cooling and the quantization noise by using more bits, others not. Photon shot noise, for example, is unavoidable due to the particle nature of light. Readout noise increases with a higher readout rate, which is desirable in high-speed cameras. In equation 3 the calculus of the Variance and Standard deviation of the image are shown [2].

$$\sigma_I{}^2 = \frac{1}{2}\, var(I_1 - I_2), where\ I_1 = I + \eta_1, I_2 = I + \eta_2 \quad (3)$$

σ_I = the Standard deviation of the Fingerprint Image Original.
var = the variance between the Fingerprint Image Original and The Fingerprint Image with Blur Motion Noise.
$\eta_1 = 0$, is the noise to the Fingerprint Image Original.
η_2 = blur motion noise 10 to 90 distance pixels.
I_2 , is the Fingerprint Image with Blur Motion Noise.

The Signal-to-Noise Ratio (SNR) of an image I is then computed as:

$$SNR = 10\log\left(\frac{varI}{\sigma^2 I}\right) \quad (4)$$

The combined effect of these noise sources is often modeled from two images of the same scene captured under the same degradation is shown in figure 2.

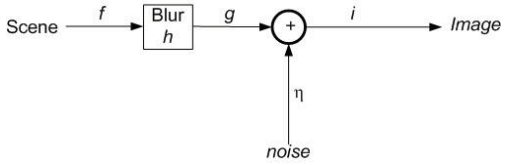


Figure 2: Image degradation by blur and additive noise.

2.2 Calculus of the SNR for Fingerprint Images

In this paper we use fingerprint images from the FCV2000 database of the Biometric System Laboratory at the University of Bologna [3, 4]. The image size is 300 pixels wide and 300 pixels high with a resolution of 500 ppi, and of a gray scale representation. The fingerprints were acquired by using a low-cost optical sensor; up to four fingers were collected from each volunteer (forefinger and middle finger of both the hands). The database is 10 fingers wide (w) and 8 impressions per finger deep (d) (80 fingerprints in all); the acquired fingerprints were manually analyzed to assure that the maximum rotation is approximately in the range [-15°, 15°] and that each pair of impressions of the same finger have a non-null overlapping area.
We added Blur Motion Level Noise from 10 to 90 distance pixels to all 80 fingerprint images, based on equations (3) and (4) the calculus of the Average SNR in decibels for the Fingerprint Image of each of the person obtained is shown in table 1.

Table I: The Average SNR in db of the Fingerprint Images.

Average SNR in db Person N° Fingerprint Image	Blur Motion Distance Pixels								
	10	20	30	40	50	60	70	80	90
P1	10.9	7.57	8.08	7.65	7.43	7.28	7.11	6.99	6.89
P2	8.36	8.02	7.52	7.38	7.16	7.01	6.89	6.78	6.68
P3	8.01	7.02	6,76	6.56	6.44	6.34	6.21	6.10	5.98
P4	7.67	5.75	5.57	5.46	5.26	5.17	5.07	5.01	4.93
P5	11.6	10.62	10.07	9.77	9.47	9.28	9.08	8.88	8.71
P6	7.29	7.28	6.88	6.75	6.58	6.46	6.33	6.20	6.09
P7	8.21	6.11	6.26	6.18	5.93	5.85	5.79	5.67	5.57
P8	10.6	10.07	9.4	9.08	8.8	8.56	8.47	8.15	7.97
P9	12.7	12	11.65	11.35	11.7	10.9	10.81	10.6	10.4
P10	8.94	6.43	6	5.79	5.52	5.33	5.24	5.14	5.03

In figure 3 we show the plot of the Average SNR of the Person Fingerprint with the blur motion level noise from 10 to 90 distance pixels.

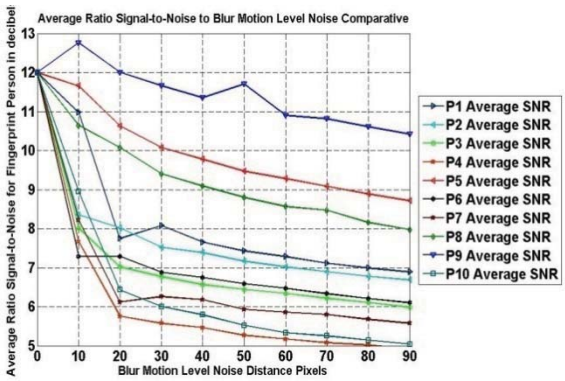


Figure 3: Average SNR with blur motion noise.

3 Ensemble Neural Networks Architecture

3.1 ANN´s Architecture

The Ensemble ANN´s architecture consists of three main modules [5, 6], in which each of them in turn consists of a set of neural networks trained with the same data (fingerprints image), If we now using different parameters for each module of the Ensemble Neural Networks, for module 1 the used parameters are learning rate =.001 , Goal Error=.001, for the module 2 learning rate =.0001 , Goal Error=.0001, for the module 3 the learning rate =.0001, and Goal Error=.00001, using 2 hidden layers, 36 neurons in the first layer and 17 neurons in the second layer for each module of the Ensemble Neural Network and is shown in figure 4.

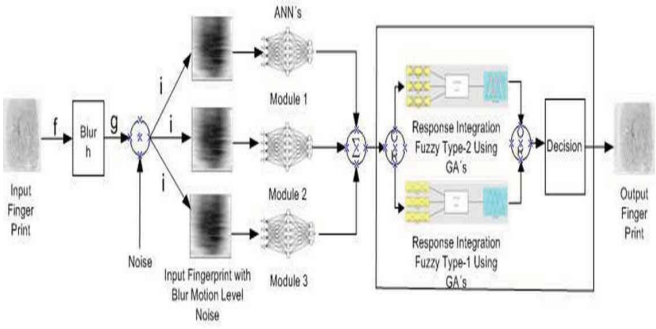


Figure 4: Ensemble Neural Networks with blur Motion Level Noise Added.

ISBN: 978-989-95079-6-8

The procedure to perform the tests was first to train the modules of the ensemble neural networks with the databases of the fingerprints without noise, in total using 8 fingerprints of 10 people, that is 80 fingerprints in total, until being able to find which architecture of the ensemble neural network responded better to arrive to the desired error.

We Added Blur Motion Level Noise from 10 to 90 distance pixels to the Fingerprint Input, to obtain a blur motion image that went into each of three modules ensemble neural networks. The output of each of the ensemble neural networks was obtained response integration with type-1 or type-2 fuzzy logic, and finally with the decision module obtained the output fingerprint.

3.2 Response Integration with Type-1 Fuzzy Logic

We also used a Type-1 Fuzzy Inference System as method of response integration of the ensemble neural network output, considering as input three linguistic variables, Activation Low, Activation Medium, and Activation High, and one output linguistic variable, Winning Activation of the three modules.

We show below one of the rules for the fuzzy inference system.

If (Module1 is ActMod1Low) and (Module2 is ActMod2Medium) and (Module3 is ActMod3Medium) then (Winner Module is Module3)

Three membership functions were used for each linguistic variable (input and output) of the triangular type, to be managed in a range from 0 to 1.

We show in figure 5 the membership functions designed using the editor of the fuzzy logic toolbox of MATLAB [14].

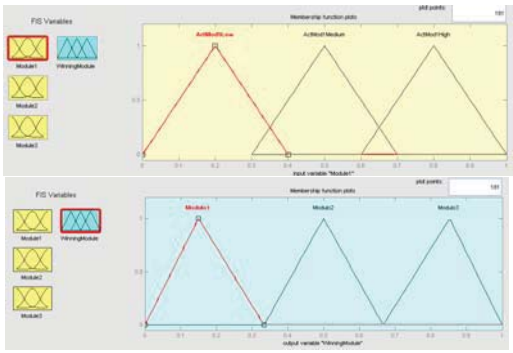


Figure 5: The Input Membership function and The Output Membership function of Fuzzy Logic Type-1.

Once the Ensemble Neural Network is trained, the fuzzy inference system integrates the outputs of the modules. We used the same 80 people's images to which we had applied noise with blur motion, and the type-1 fuzzy inference system gives an answer for the stage of the final decision.

3.3 Response Integration with Type-2 Fuzzy Logic

We used a Type-2 Fuzzy Inference System as method of response integration of the ensemble neural network output, considering as input three linguistic variables, Activation Low, Activation Medium, and Activation High, and one output linguistic variable, Winning Activation of the three modules, used the same rules of the type-1 fuzzy logic.

Three membership functions were used for each linguistic variable of the input and output of the Gaussian type managed in a range from 0 to 1.

We show in figure 6 the membership functions designed using the editor of IT2FUZZY fuzzy logic toolbox [12, 13].

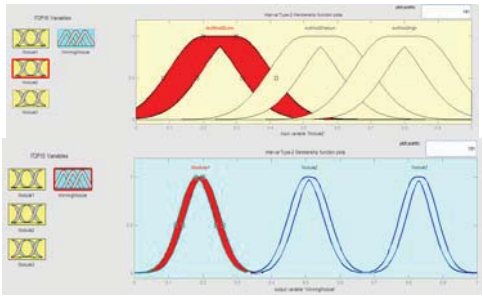


Figure 6: The Input Membership function and The Output Membership function of Type-2 Fuzzy Logic.

4 Simulation Results

4.1 Response Integration with Type-1 Fuzzy Logic

Once the Ensemble Neural Network is trained, the fuzzy inference system integrates the outputs of the modules. We used the same 80 persons images to which we had applied different levels of noise with blur motion, the type-1 inference systems give an output for the stage of the final decision, and we show the result if the fingerprint input was recognized. We show in figure 7 the simulation results.

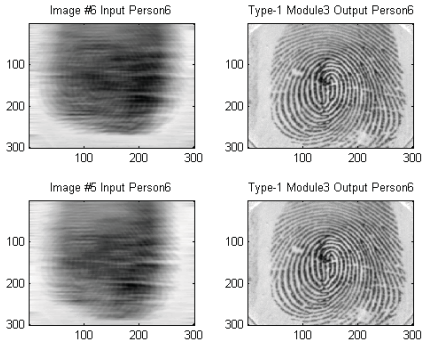


Figure 7: Simulation results for the fingerprints using the type-1 Fuzzy Inference System (blur motion 50 distance pixels).

4.2 Response Integration with Type-2 Fuzzy Logic

In the same way when the Ensemble Neural Network is trained, the fuzzy inference system integrates the outputs of the modules. We used the same 80 persons images to which we had applied different levels of noise with blur motion, the type-2 inference systems give an output for the stage of the final decision, and show the result if the fingerprint input was recognized. We show in figure 8 the simulation results.

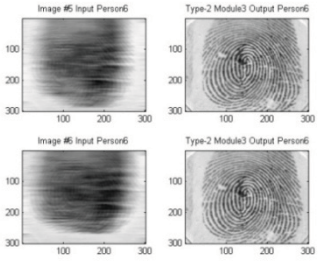


Figure 8: Simulation results for the fingerprints using the type-2 Fuzzy Inference System (blur motion 50 distance pixels).

ISBN: 978-989-95079-6-8

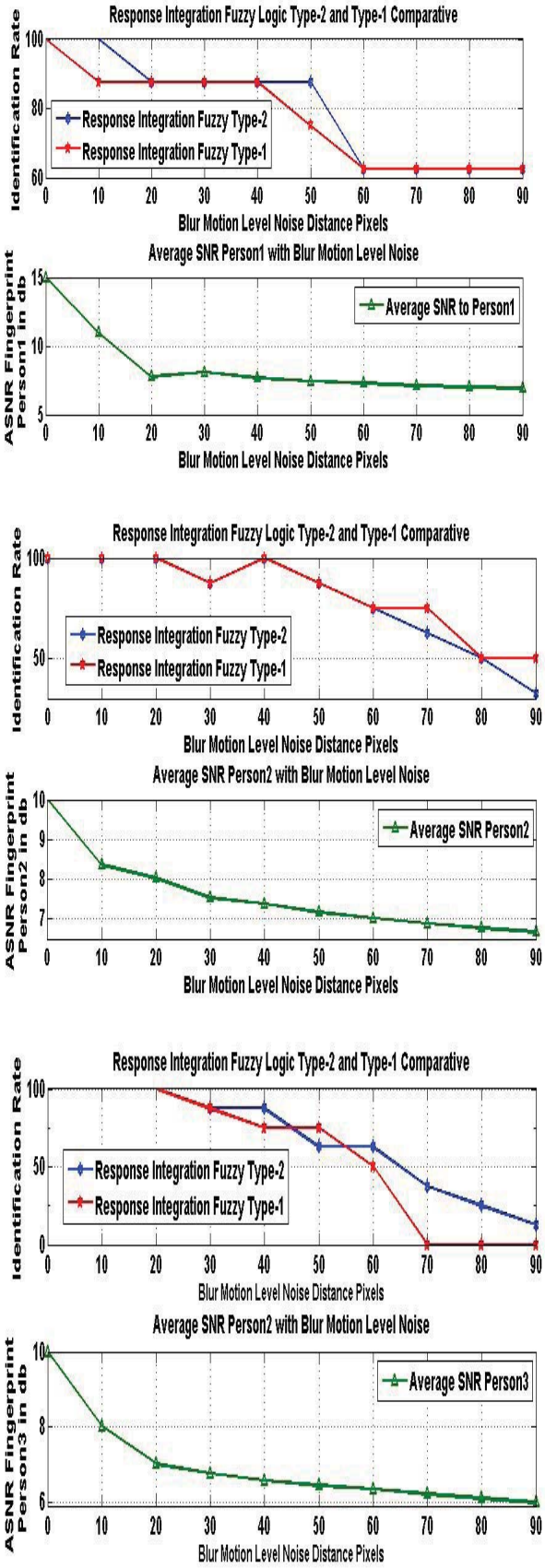
4.3 Comparison Pattern Recognition Blur Motion Noisy Images using Fuzzy Methods for Response Integration in Ensemble Neural Networks

The Response Integration method of the ensemble neural networks has the goal of combining the responses of the modules to improve the recognition rate of the individual modules when the blur motion signal increases to high level 10 to 90 distance pixels. In Table 2 we show a Comparison between Response Integration with Type-2 and Type-1 Fuzzy logic when the blur motion level noise increases and the Average SNR in db decrease.

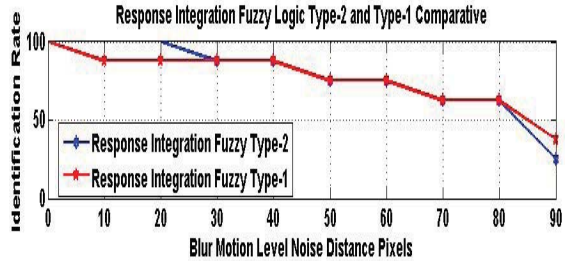
Table II: Comparison Response Integration Type-2 and Type-1 Fuzzy Logic with Blur Motion Level Noise and the Average SNR in db of the Fingerprint Images.

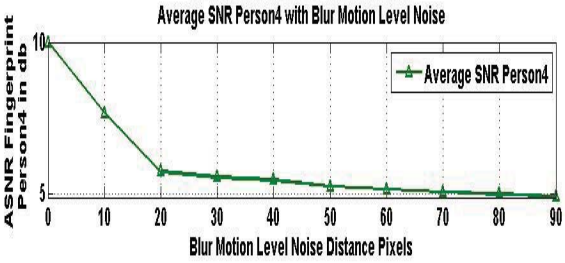
# Person Fingerprint	Blur Motion Level Noise								
	10	20	30	40	50	60	70	80	90
Average SNR P1 db	10.97	7.75	8.08	7.65	7.43	7.28	7.11	6.99	6.89
Fuzzy Type-2 %	100	87.5	87.5	87.5	87.5	62,5	62.5	62.5	62.5
Fuzzy Type-1 %	87.5	87.5	87.5	87.5	75	62.5	62.5	62.5	62.5
Average SNR P2 db	8.33	8.02	7.52	7.38	7.16	7.01	6.89	6.78	6.68
Fuzzy Type-2 %	100	100	87.5	100	87.5	75	62.5	50	37.5
Fuzzy Type-1 %	100	100	87.5	100	87.5	75	75	50	50
Average SNR P3 db	8.01	7.02	6.76	6.56	6.44	6.34	6.21	6.10	5,98
Fuzzy Type-2 %	100	100	87.5	87.5	62.5	62.5	37.5	25	12.5
Fuzzy Type-1 %	100	100	87.5	75	75	50	0	0	0
Average SNR P4 db	7.67	5.75	5.57	5.46	5.26	5.17	5.075	5.01	4.94
Fuzzy Type-2 %	100	100	87.5	87.5	75	75	62.5	62.5	25
Fuzzy Type-1 %	87.5	87.5	87.5	87.5	75	75	62.5	62.5	37.5
Average SNR P5 db	11.65	10.62	10.08	9.77	9.47	9.28	9.08	8.89	8.71
Fuzzy Type-2 %	100	100	100	100	100	87.5	87.5	75	50
Fuzzy Type-1 %	100	100	100	100	100	87.5	75	75	50
Average SNR P6 db	7.29	7.28	6.88	6.75	6.58	6.46	6.33	6.21	6.1
Fuzzy Type-2 %	100	100	100	100	100	100	100	87.5	75
Fuzzy Type-1 %	100	100	100	100	100	100	100	87.5	75
Average SNR P7 db	8.21	6.11	6.26	6.18	5.93	5.85	5.79	5.67	5.57
Fuzzy Type-2 %	100	100	100	87.5	87.5	87.5	62.5	50	37.5
Fuzzy Type-1 %	100	100	100	75	75	87.5	62.5	50	37.5
Average SNR P8 db	10.64	10.07	9.40	9.09	8.8	8.56	8.47	8.15	7.97
Fuzzy Type-2 %	100	100	100	87.5	87.5	87.5	62.5	50	37.5
Fuzzy Type-1 %	100	100	100	75	75	87.5	62.5	50	37.5
Average SNR P9 db	12.76	12	11.65	11.35	11.7	10.9	10.81	10.6	10.42
Fuzzy Type-2 %	100	100	100	100	100	100	100	100	100
Fuzzy Type-1 %	100	100	100	100	100	100	100	100	100
Average SNR P10 db	8.21	6.11	6.26	6.18	5.93	5.85	5.79	5.67	5.57
Fuzzy Type-2 %	100	100	100	100	100	100	100	100	100
Fuzzy Type-1 %	100	100	100	100	100	100	100	100	100

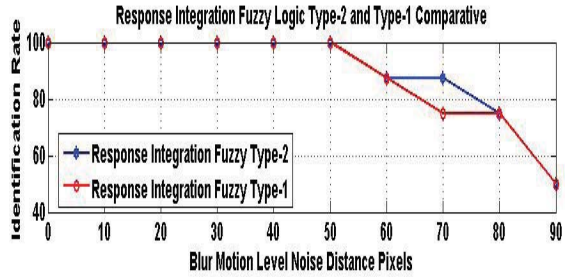
In Figure 9 we show in the up side the response integration of the type-2 and type-1 fuzzy system when the blur motion noise increases and the down side the Average SNR in db decrease for each of the Person Fingerprint Images.

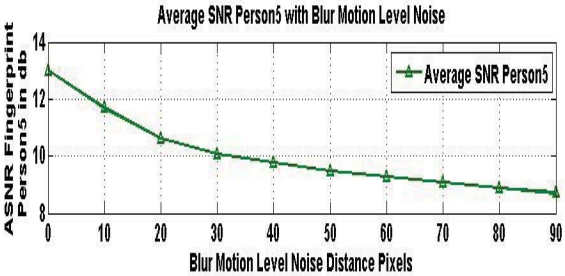


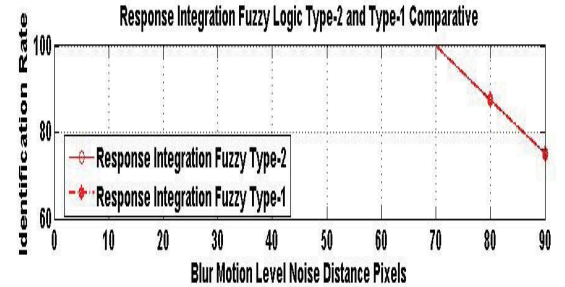
ISBN: 978-989-95079-6-8

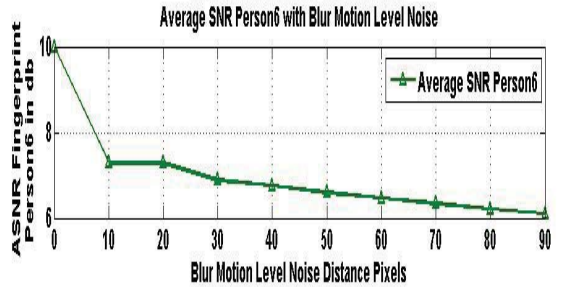


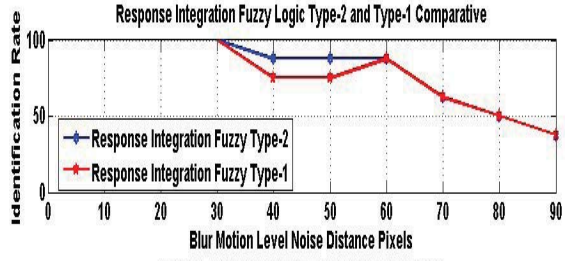


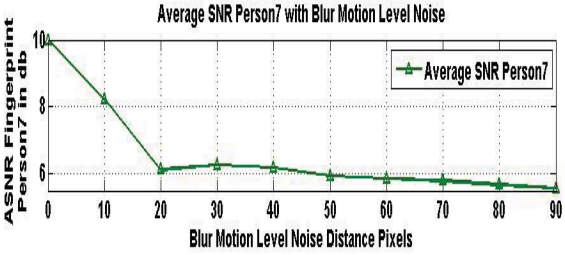


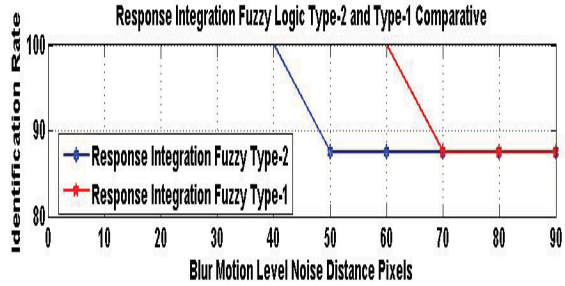


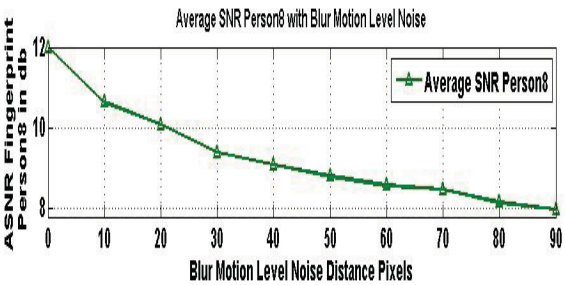


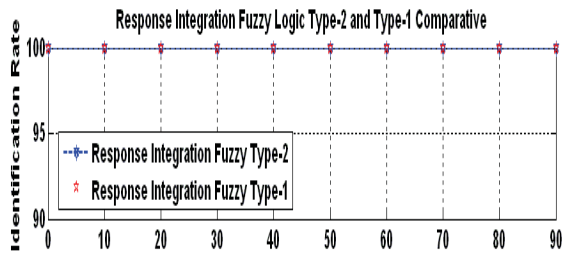


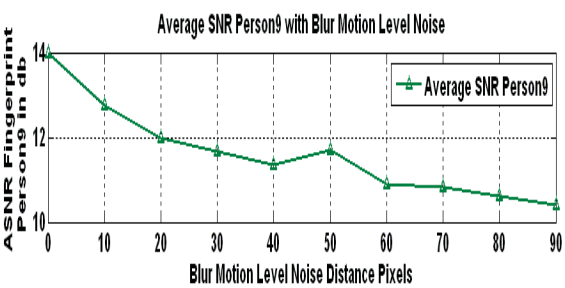












ISBN: 978-989-95079-6-8

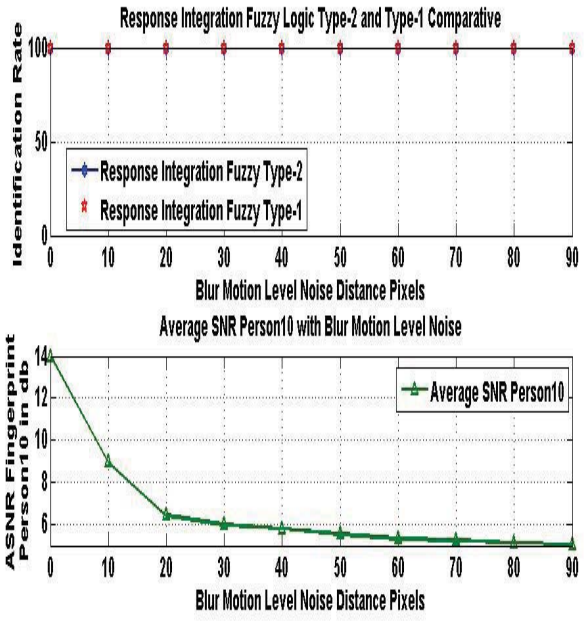


Figure 9: Response Integration Type-2 and Type-1 Fuzzy logic with the blur Motion Level Noise 10 to 90 Distance Pixels and Average SNR in db of each 10 Person Fingerprint Image.

5 Conclusions

Based on the experimental results, we can conclude that the behavior of the type-2 and type-1 fuzzy systems, as methods of response integration of ensemble neural networks for the fingerprints when added blur motion level noise increase and the Average SNR of the Fingerprint decrease, we obtain excellent results.

The identification rates obtained are between 37.5% to 100%for response integration type-2 fuzzy logic and 0% to 100% for response integration with type-1 fuzzy logic. The difference between the type-1 fuzzy and type-2 fuzzy inference systems for response integration of ensemble neural networks is appreciated when we uses the blur motion level noise. We think that there is an advantage in using a type-2 fuzzy inference system to manage the uncertainty of the knowledge base in pattern recognition problems. Future work will include, testing with more kinds of noise, other methods of response integration, using feature extraction, and image compression, with the goal of improving the identification rate.

Acknowledgment

We would like to express our gratitude to CONACYT, Universidad Autonoma de Baja California and Tijuana Institute of Technology for the facilities and resources granted for the development of this research project.

References

[1] Felix Krahmer_, Youzuo Lin, Bonnie McAdoo, Katharine Ott, Jiakou Wang, David Widemannk, Brendt Wohlberg, *Blind Image Deconvolution: Motion Blur Estimation, August 2006.*

[2] Tuan Q. Pham, Lucas J. van Vliet, Klamer Schutte, *Influence of signal-to-noise ratio and point spread function on limits of super resolution*, *Image Processing, Algorithms and Systems IV, Vol. 5672, pp. 169-180, 2005.*

[3] D. Maltoni, D. Maio, A.K. Jain, S. Prabhakar, *The full FVC2000 and FVC2002 databases are available in the DVD included in: Handbook of Fingerprint Recognition, Springer, New York, 2003.*

[4] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman and A. K. Jain, *"FVC2004: Third Fingerprint Verification Competition", Proc. International Conference on Biometric Authentication (ICBA), pp. 1-7, Hong Kong, July 2004.*

[5] M. Lopez, P. Melin: *Response integration in Ensemble Neural Networks using interval type-2 Fuzzy logic. IJCNN 2008: 1503-1508*

[6] O. Mendoza, P. Melin, G. Licea, *Modular Neural Networks and Type-2 Fuzzy Logic for Face Recognition, In Marek Reformat, editor, Proceedings of NAFIPS 2007, Number 1, pages CD Rom, San Diego, June 2007.*

[7] J. Urias, D. Solano, .M. Soto, M. Lopez, P. Melin, *¨Type-2 Fuzzy Logic as a Method of Response Integration in Modular Neural Networks¨. IC-AI 2006: 584-590.*

[8] P. Melin, F. González, G. Martínez: *Pattern Recognition Using Modular Neural Networks and Genetic Algorithms. IC-AI 2004: 77-83.*

[9] P. Melin, J. Urias, D. Solano, M. Soto, M. Lopez, O. Castillo: *Voice Recognition with Neural Networks, Type-2 Fuzzy Logic and Genetic Algorithms. Engineering Letters 13(2): 108-116 (2006).*

[10] P. Melin, A. Mancilla, M. Lopez, D. Solano, M. Soto, O. Castillo, *¨Pattern Recognition for Industrial Security using the Fuzzy Sugeno Integral and Modular Neural Networks¨, WSC11 11th Online World Conference on Soft Computing in Industrial Applications September 18-October 6, 2006.*

[11] P. Melin and O. Castillo, 2005*, ¨Hybrid Intelligent Systems for Pattern Recognition using Soft Computing¨, ISBN 3-540-24121-3 Springer Berlin Heidelberg New York.*

[12] J. R. Castro, O. Castillo, P. Melin, L.G. Martinez, S. Escobar, I. Camacho*, Building Fuzzy Inference Systems with Interval Type-2 Fuzzy Logic Toolbox Number 1 in Studies in Fuzziness and Soft Computing, 6, pages 53-62, Springer-Verlag, Germany, 1 edition, June 2007.*

[13] J. R. Castro, O. Castillo, P. Melin, *An Interval Type-2 Fuzzy Logic Toolbox for Control Applications; In Proc. FUZZ-IEEE 2007.*

[14] MATLAB Trade Marks, ©1994-2007 by The Math Works, Inc.

[15] S. Farsiu, D. Robinson, M. Elad and P. Milanfar. *Fast and robust multi-frame super-resolution. IEEE Trans. on Image Processing, 13(10):1327-1344, 2004.*

[16] R.C. Hardie, K.J. Barnard and E.E. Armstrong. *Joint map registration and high-resolution image estimation using a sequence of undersampled images. IEEE Trans.on Image Processing, 6(12):1621–1633, 1997.*

[17] R. Neelamani, H. Choi and R. Baraniuk. Forward: *Fourier-wavelet regularized deconvolution for Ill-Conditioned Systems. In IEEE Trans. Signal Processing, 52(2):418–433, 2004.*

[18] J. Biemond. *Iterative methods for image deblurring. Proceedings of the IEEE, 78(5):856–883*, 1990.

[19] M.Ebrahimi Moghadam, M.Jamzad, *Linear motion blur parameter estimation in noisy images using fuzzy sets and power spectrum, EUROSIP Journal on applied signal processing, Article number 68985, Volume 2007, pages 862–866..*

[20] Y. Yitzhaky, I. Mor, A. Lantzman, and N. S. Kopeika. *Direct method for restoration of motion blurred images. Journal of the Optical Society of America. A, Optics and image science, 15(6):1512–1519, 1998.*

Fuzzy Logic Adaptation of a Hybrid Evolutionary Method for Pattern Recognition

Fevrier Valdez[1] Patricia Melin[2] Oscar Castillo[2]

[1]PhD Student of Computer Science in the Universidad Autónoma de Baja California
Tijuana, B.C.
[2]Computer Science in the Graduate Division, Tijuana Institute of Technology
Tijuana, B.C.
Email: fvaldez0125@yahoo.com.mx, pmelin@tectijuana.mx, ocastillo@tectijuana.mx

Abstract— *We describe in this paper a new hybrid approach for optimization combining Particle Swarm Optimization (PSO) and Genetic Algorithms (GAs) using Fuzzy Logic to integrate the results. The new evolutionary method combines the advantages of PSO and GA to give us an improved FPSO+FGA hybrid method. Fuzzy Logic is used to combine the results of the PSO and GA in the best way possible. The new hybrid FPSO+FGA approach is compared with the PSO and GA methods with a set of benchmark mathematical functions. The proposed hybrid method is also tested with the problem of modular neural network optimization. The new hybrid FPSO+FGA method is shown to be superior with respect to both the individual evolutionary methods.*

Keywords— Fuzzy Logic, Evolutionary Computing, Genetic Algorithms**.**

1 Introduction

We describe in this paper a new evolutionary method combining PSO and GA, to give us an improved FPSO+FGA hybrid method. We apply the hybrid method to mathematical function optimization to validate the new approach. Also in this paper, the application of a Genetic Algorithm (GA) [1] and Particle Swarm Optimization (PSO) [2] for the optimization of mathematical functions is considered. In this case, we are using the Rastrigin's function, Rosenbrock's function, Ackley's function, Sphere's function Griewank's function, Michalewics's function and Zakharov's function [4][13] to compare the optimization results between a GA, PSO and FPSO+FGA. We also consider the problem of neural network architecture optimization, which is very important in the applications.

The paper is organized as follows: in section 2 a description about the Genetic Algorithms for optimization problems is given, in section 3 the Particle Swarm Optimization is presented, the neural networks is presented in section 4, in section 5 we can appreciate the proposed method FPSO+FGA and the fuzzy system, in section 6 we can appreciate the full model of FPSO+FGA that was used for this research, in section 7 the simulations results are described, finally we can see in section 8 the conclusions reached after the study of the proposed evolutionary computing methods.

2 Genetic Algorithms for Optimization

John Holland, from the University of Michigan initiated his work on genetic algorithms at the beginning of the 1960s. His first achievement was the publication of *Adaptation in Natural and Artificial System* [7] in 1975.

He had two goals in mind: to improve the understanding of natural adaptation process, and to design artificial systems having properties similar to natural systems [8].

The basic idea is as follows: the genetic pool of a given population potentially contains the solution, or a better solution, to a given adaptive problem. This solution is not "active" because the genetic combination on which it relies is split between several subjects. Only the association of different genomes can lead to the solution.

Holland's method is especially effective because it not only considers the role of mutation, but it also uses genetic recombination, (crossover) [9]. The crossover of partial solutions greatly improves the capability of the algorithm to approach, and eventually find, the optimal solution.

The essence of the GA in both theoretical and practical domains has been well demonstrated [1]. The concept of applying a GA to solve engineering problems is feasible and sound. However, despite the distinct advantages of a GA for solving complicated, constrained and multi-objective functions where other techniques may have failed, the full power of the GA in application is yet to be exploited [12] [14].

In Fig. 1 we show the reproduction cycle of the Genetic Algorithm.

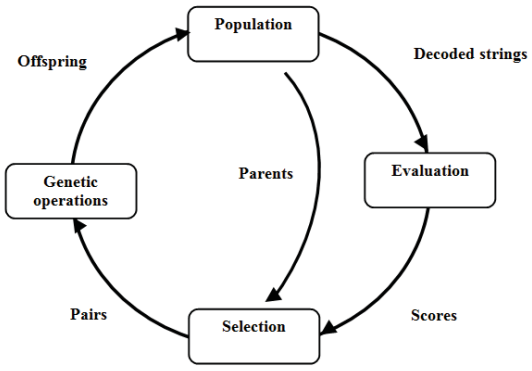


Figure 1: The Reproduction cycle

ISBN: 978-989-95079-6-8

3 Particle Swarm Optimization

Particle swarm optimization (PSO) is a population based stochastic optimization technique developed by Eberhart and Kennedy in 1995, inspired by social behavior of bird flocking or fish schooling [3].

PSO shares many similarities with evolutionary computation techniques such as Genetic Algorithms (GA) [6]. The system is initialized with a population of random solutions and searches for optima by updating generations.

However, unlike GA, the PSO has no evolution operators such as crossover and mutation. In PSO, the potential solutions, called particles, fly through the problem space by following the current optimum particles [10].

Each particle keeps track of its coordinates in the problem space, which are associated with the best solution (fitness) it has achieved so far (The fitness value is also stored). This value is called *pbest*. Another "best" value that is tracked by the particle swarm optimizer is the best value, obtained so far by any particle in the neighbors of the particle. This location is called *lbest*. When a particle takes all the population as its topological neighbors, the best value is a global best and is called *gbest* [19].

The particle swarm optimization concept consists of, at each time step, changing the velocity of (accelerating) each particle toward its p*best* and *lbest* locations (local version of PSO). Acceleration is weighted by a random term, with separate random numbers being generated for acceleration toward *pbest* and *lbest* locations [17].

In the past several years, PSO has been successfully applied in many research and application areas. It is demonstrated that PSO gets better results in a faster, cheaper way compared with other methods [11] [18].

Another reason that PSO is attractive is that there are few parameters to adjust. One version, with slight variations, works well in a wide variety of applications. Particle swarm optimization has been used for approaches that can be used across a wide range of applications, as well as for specific applications focused on a specific requirement [20].

4 Neural Networks

The discipline of neural networks originates from an understanding of the human brain. The average human brain consists of 3 × 1010 neurons of various types, with each neuron connecting to up to 104 synapses [21]. Many neural-network models, also called connectionist models, have been proposed [6].

Neural networks are attractive since they consist of many neurons, with each one processing information separately and simultaneously. All the neurons are connected by synapses with variable weights. Thus, neural networks are actually parallel distributed processing systems.

Research on neural networks dates back to the 1940s when McCulloch and Pitts found that the neuron can be modeled as a simple threshold device to perform logic function [22].

In this paper, we used a supervised learning for training the neural network. Supervised learning is based on a direct comparison between the actual network output and the desired output. Some recent advances in supervised learning have been reviewed in [5].

5 FPSO+FGA Method

This method combines the characteristics of PSO and GA using several fuzzy systems for integration of results and parameter adaptation. In this section, the proposed FPSO+FGA method is presented.

The general idea of the proposed FPSO+FGA method can be seen in Fig. 2. The method can be described as follows:

1. A problem optimization to be considered is received in this case; we are testing with a modular neural network and mathematical functions.
2. The method is randomly initialized to start with FPSO or FGA.
3. The system will solve the problem by FPSO or FGA depending of Error (E) and Derivate of Error (DE) generated after applying FPSO or FGA.
4. The function evaluated is called FEVAL, then the system records the Errors (E and DE) obtained to compare with the value desired.
5. After that the function is evaluated, E and DE are taken as inputs to one fuzzy system called 'FSDM', which the main function is to decide if is suitable to continue solving the problem with FPSO or FGA depending of the information generated by the fuzzy rules to calculate when can be necessary do a change
6. Another two fuzzy systems receive the values of Error and DError as inputs to evaluate if it is necessary to change the values of the parameters in FPSO or FGA depending of the fuzzy rules. The fuzzy systems responsible to do change the value of parameters in FPSO and FGA are called 'fuzzyga' and 'fuzzypso'. On Fig. 2 the two fuzzy systems are represented by el block FSPA (Fuzzy System Parameter Adaptation) but while the method is running in the FPSO the fuzzy system changes the values of the cognitive acceleration 'c_1', and social acceleration 'c_2'; and while the method is running in the FGA the fuzzy system change the value of crossover and mutation.
7. Repeat the above steps until the termination criterion of the algorithm is met. Finally BVALUE receive the best value obtained and when the method is finished we are obtaining the best result.

The fundamental characteristic of this method is to adapt the parameters to reach the best result, in this case; we are doing changes on the parameters of crossover (cr), mutation (mu), cognitive acceleration (c_1) and social acceleration (c_2). We are taking this parameters because are very important to the convergence with good results of the method, for example 'c_1' and 'c_2' are working as a memory able of record the knowledge of the particles more important in a population in the FPSO; also 'cr' and 'mu' are important to reach the best solution in the FGA. The values of these four parameters are fuzzy because depending of the output of a fuzzy system able of adaptation parameters.

ISBN: 978-989-95079-6-8

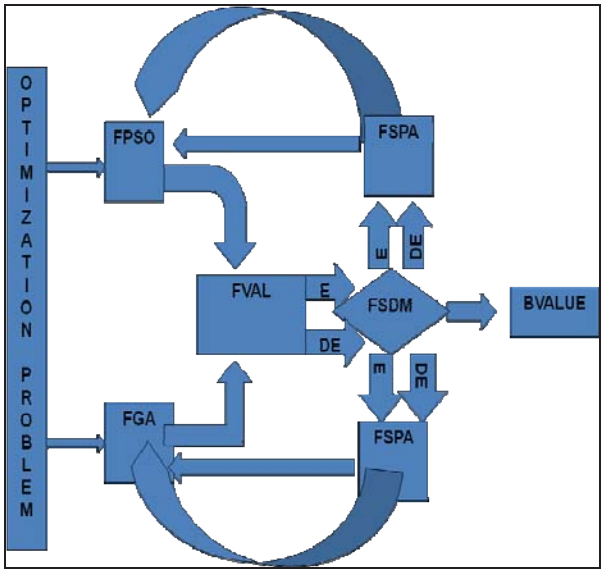


Figure 2: The FPSO+FGA scheme

6 Full Model of FPSO+FGA

The basic idea of the FPSO+FGA scheme is to combine the advantages of the individual methods using a fuzzy system for decision making and the others two fuzzy systems to improve the parameters of the FGA and FPSO when is necessary.

As can be seen in the proposed hybrid FPSO+FGA method, it is the internal fuzzy system structure, which has the primary function of receiving as inputs (Error and DError) the results of the outputs FGA and FPSO. The fuzzy system is responsible for integrating and deciding which are the best results being generated at run time of the FPSO+FGA. It is also responsible for selecting and sending the problem to the "fuzzypso" fuzzy system when the FPSO is activated or to the "fuzzyga" fuzzy system when FGA is activated. Also activating or temporarily stopping depending on the results being generated. The fuzzy system is of Mamdani type because it is more common in this type of fuzzy control and the defuzzification method is the centroid. In this case, we are using this type of defuzzification because in other papers we have achieved good results [4]. Also, the membership functions in the fuzzy main system were chosen of triangular form based on past experiences in this type of fuzzy control. However, in the simulation results we can see different architectures to the 'fuzzyga' and the 'fuzzypso'; where, we are changing the membership functions to observe which is more reliable to this application. In this research, we are testing with triangular and gaussian membership functions.

6.1 FPSO (Fuzzy Particle Swarm Optimization)

This section presents a detailed description of the FPSO model. The classical representation scheme for GAs is binary vectors of fixed length. In the case of an n_x – dimensional search space, each individual consists of n_x variables with each variable encoded as a binary string. The swarm is typically modeled by particles in multidimensional space that have a position and a velocity. These particles fly through hyperspace (i.e., R^n) and have two essential reasoning capabilities: their memory of their own best position and knowledge of the global or their neighborhood's best. In Fig. 3 we can see a simulation of Ackley's function. In a minimization optimization problem, "best" simply means the position with the smallest objective value. Members of a swarm communicate good positions to each other and adjust their own position and velocity based on these good positions. So a particle has the following information to make a suitable change in its position and velocity:

- A global best that is known to all and immediately updated when a new best position is found by any particle in the swarm.
- The neighborhood best that the particle obtains by communicating with a subset of the swarm.
- The local best, which is the best solution that the particle has seen.

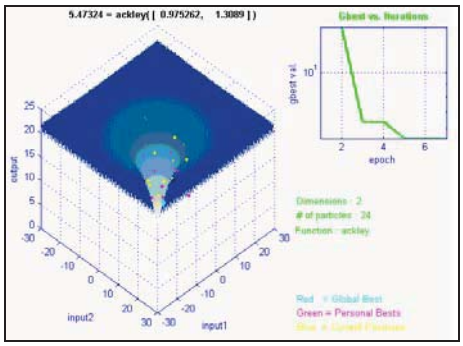


Figure 3: Simulation of ackley's function with FPSO

In this case, the social information is the best position found by the swarm, referred as $\hat{y}(t)$. For gbest FPSO, the velocity of particle i is calculated as

$$v_{ij}(t+1) = vi_j(t) + c_1 r_{1j}(t)[y_{ij}(t) - x_{ij}(t)] + c_2 r_{2j}(t)[\hat{y}_j(t) - x_{ij}(t)] \quad (1)$$

Where $vi_j(t)$ is the velocity of particle i in dimension j = 1,…,n_x at time step t, $xi_j(t)$ is the position of particle i in dimension j at time step t, 'c_1' and 'c_2' represents the cognitive and social acceleration. In this case, these values are fuzzy because they are changing dynamically when the FPSO is running, and $r_{1j}(t)$, r_{2j}~U(0,1) are random values in the range [0,1]. We are adding a fuzzy system called 'fuzzypso' that is able of change the 'c_1' and 'c_2' rate.

6.2 FGA (Fuzzy Genetic Algorithm)

This section presents a brief description of the FGA model. Several crossover operators have been developed for GAs, depending on the format in which individuals are represented. For binary representations, uniform crossover, one point crossover and two point crossover are the most popular. In this case we are using two points crossover with fuzzy crossover rate because we are adding a fuzzy system called 'fuzzyga' that is able of change the crossover and mutation rate.

6.3 Definition of the Fuzzy Systems used in FPSO+FGA

'fuzzypso': In this case we are using a fuzzy system called 'fuzzypso', and the structure of this fuzzy system is as follows:

Number of Inputs: 2
Number of Outputs: 2
Number of membership functions: 3

ISBN: 978-989-95079-6-8

Type of the membership functions: Triangular and Gaussian
Number of rules: 9
Defuzzification: Centroid
The main function of the fuzzy system called 'fuzzypso' is to adjust the parameters of the PSO. In this case, we are adjusting the following parameters: 'c_1' and 'c_2'; where:
'c_1' = Cognitive Acceleration
'c_2' = Social Acceleration
We are changing these parameters to test the proposed method. In this case, with 'fuzzypso' is possible to adjust in real time the 2 parameters that belong to the PSO.
'fuzzyga': In this case we are using a fuzzy system called **'fuzzyga'**, the structure of this fuzzy system is as follows:
Number of Inputs: 2
Number of Outputs: 2
Number of membership functions: 3
Type of membership functions: Triangular and Gaussian
Number of rules: 9
Defuzzification: Centroid
The main function of the fuzzy system called 'fuzzypso' is to adjust the parameters of the GA. In this case, we are adjusting the following parameters: 'mu', 'cr'; where:
'mu' = mutation
'cr' = crossover
'fuzzymain': In this case, we are using a fuzzy system called **'fuzzymain'**. The structure of this fuzzy system is as follows:
Number of Inputs: 2
Number of Outputs: 1
Number of membership functions: 3
Type of membership functions: Triangular
Number of rules: 9
Defuzzification: Centroid
The main function of the fuzzy system, called 'fuzzymain' is to decide on the best way for solving the problem, in other words if it is more reliable to use the FPSO or FGA. This fuzzy system is able to receive two inputs, called error and derror, it is to evaluate the results that are generated by FPSO and FGA in the last step of the algorithm. Fig. 4 shows the membership functions of the main fuzzy system that is implemented in this method. All fuzzy system consists of 9 rules. For example, one rule is if error is P and DError is P then best value is P (view Fig. 5). Fig. 6 shows the fuzzy system rule viewer. Fig. 7 shows the surface corresponding to the fuzzy main system. The other two fuzzy systems are similar to the main fuzzy system.

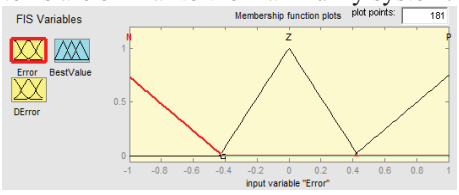
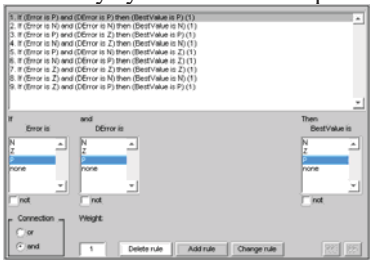


Figure 4: Fuzzy system membership functions


Figure 5: Fuzzy system rules

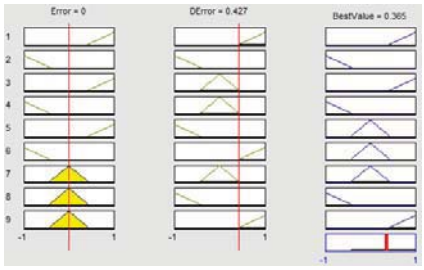


Figure 6: Fuzzy system rules viewer

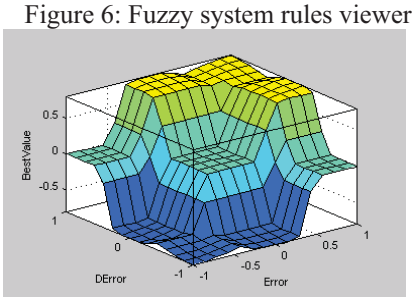


Figure 7: Surface of fuzzy system

7. Simulation Results for Modular Neural Network for Optimization

Several tests of the FPSO+FGA method for MNN optimization were made in the Matlab programming language.

All the implementations were developed using a computer with processor Intel core 2 quad of 64 bits that works to a frequency of clock of 2.5 GHz, 4 GB of RAM Memory and Windows Vista operating system.

We describe below simulation results of our approach for face recognition with modular neural networks (MNNs). We used two-layer feed-forward MNNs with the Conjugated-Gradient training algorithm [4]. The challenge is to find the optimal architecture of this type of neural network, which means finding out the optimal number of layers and nodes of the neural network [6]. We are using the Yale face database [20] that contains 165 grayscale images in GIF format of 15 individuals, for this paper only 10 subjects were used for training the MNN. There are 5 images per subject, one per different facial expression: center-light, happy, left-light, normal and right-light. In total 50 images were used (view Fig. 8). Three images per subject were used for training the MNN and the other two for the recognition. Regarding the genetic algorithm for NN evolution, we used a hierarchical chromosome for representing the relevant information of the network. First, we have the bits for representing the number of layers of the MNN; in this case, the initial topology was of 3 modules with 2 layers per module with 500 neurons in the first layer, 300 neurons in the second layer in each module. Therefore we used a representation the 2415 bits in total (view Fig. 9). The PSO is organized in a similar fashion, but there is less number of parameters. In Fig. 10 we can see the architecture of a MNN that we are using with the evolutionary proposed method FPSO+FGA.

The fitness function used in this case for the MNN combines the information of the error objective and also the information about the number of nodes as a second objective. This is shown in the following equation.

$$f(z) = \left(\frac{1}{\alpha * Ranking(ObjV1) + \beta * ObjV2} \right) * 10 \qquad (1)$$

ISBN: 978-989-95079-6-8

The first objective is basically the average sum of squared of errors as calculated by the predicted outputs of the MNN compared with real values of the function. This is given by the following equation.

$$f_1 = \frac{1}{N}\sum_{i=1}^{N}(Y_i - y_i)^2 \qquad (2)$$

The second objective is the complexity of the neural network, which is measured by the total number of nodes in the architecture.

The final topology of the neural network for the problem of face recognition is obtained by the hybrid evolutionary method FPSO+FGA. The comparison of the final objective values (errors) will be shown in the following section. In the final architecture, the result of the MNN evolution is a particular architecture with different number of nodes by layers. Several tests were made; we obtained optimized different architectures for this Modular Neural Network; the best architecture obtained was the following:

Layers = 2 x module

NNL1M1 = 90, NNL2M1 = 50. NNL1M2 = 100, NNL2M2 = 150. NNL1M3 = 70, NNL2M3 = 90. Total bits = 565

Where: NNL1M1= Number of neurons of the layer 1 in module 1. NNL1M1= Number of neurons of the layer 1 in module 1. NNL2M1= Number of neurons of the layer 2 in module 1. NNL1M2= Number of neurons of the layer 1 in module 2. NNL2M2= Number of neurons of the layer 2 in module 2. NNL1M3= Number of neurons of the layer 1 in module 3. NNL2M3= Number of neurons of the layer 2 in module 3. We can see in the Fig. 11 the binary representation for this optimized architecture. With this final topology the Neural Network was trained and the ten images were recognized. It can be seen in Table 1 the different architectures obtained with this method using a 'fuzzyga' and a 'fuzzypso' with triangular membership functions and a 'fuzzymain' with triangular membership functions, the other parameters to the 3 fuzzy systems are above described. Table 2, shows the simulation results with all fuzzy systems with gaussian membership functions. The proposed method optimizes the initial architecture proposed for the problem of face recognition.



Figure 8: Images of the Yale face database

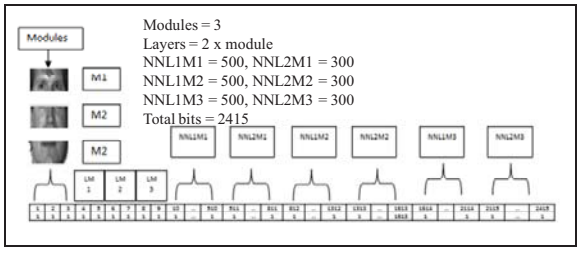


Figure 9: Binary representation for FPSO+FGA (No optimized)

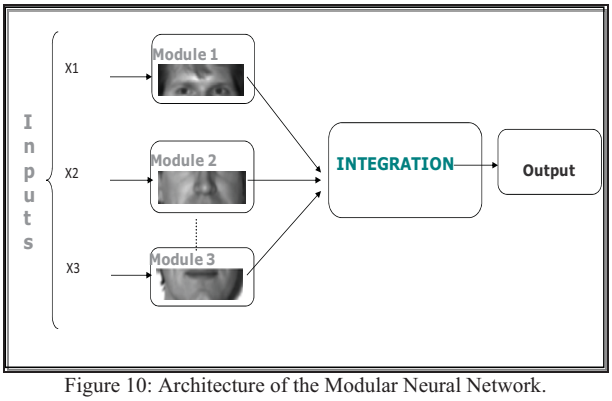


Figure 10: Architecture of the Modular Neural Network.

The Parameters of Table 1 and 2, are as follows:

LMod = Layers per module

NNL = Number of neurons Layer 1, 2 or 3.

GE = Goal Error.

RE = Reached Error.

IDENT = Number of images recognized.

VAR = Number of variables for the mathematical function.

It can be seen in Table 1 and Table 2 that this method is good alternative to solve this type of problems, in our case to optimize Modular Neural Networks. Also this FPSO+FGA have been applied, for optimization of complex mathematical functions to validate our approach. The parameters in FPSO+FGA; as crossover, mutation, cognitive acceleration and social are fuzzy, the population size was of 100 individuals, Table 3 shows the simulation results with 7 mathematical functions to testing this hybrid evolutionary method. The mathematical functions are evaluated with 2, 4, 8 and 16 variables.

Table 1. Simulation Results for the MNN with triangular membership functions

LMod	NN L1 M1	NN L2 M1	NN L1 M2	NN L2 M2	NN L1 M3	NN L2 M3	GE	RE	IDENT
2	20	60	80	50	60	120	0.01	0.03	8
2	90	50	100	150	70	90	0.01	0.005	10
2	70	40	80	40	90	30	0.01	0.02	8
2	150	135	200	90	84	40	0.01	0.003	9
2	100	120	100	145	100	70	0.01	0.001	10

Table 2. Simulation Results for the MNN with gaussian membership functions

LMod	NN L1 M1	NN L2 M1	NN L1 M2	NN L2 M2	NN L1 M3	NN L2 M3	GE	RE	IDENT
2	30	50	90	70	50	115	0.01	0.05	8
2	85	60	103	140	80	130	0.01	0.02	9
2	90	50	95	50	85	40	0.01	0.08	8
2	130	120	180	90	70	50	0.01	0.02	9
2	50	90	75	70	40	40	0.01	0.01	9

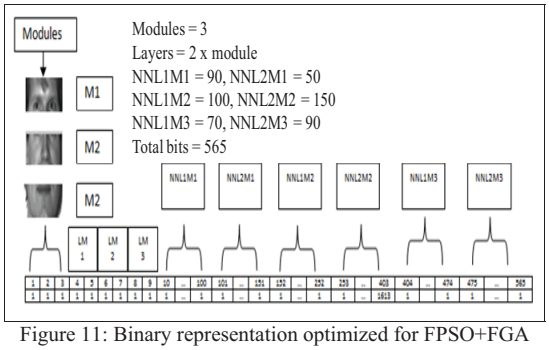


Figure 11: Binary representation optimized for FPSO+FGA

ISBN: 978-989-95079-6-8

Table 3. Simulation results for Mathematical Functions

Math F	NVar = 2		NVar = 4		NVar = 8		NVAR = 16	
Math	BEST	MEAN	BEST	MEAN	BEST	MEAN	BEST	MEAN
Ras	1.45E-06	3.05E-04	0.0003	0.0755	0.01318	0.9311	0.5483	5.0946
Ros	1.17E-02	1.17E-02	0.0285	0.5991	0.15800	3.8925	0.2555	4.33334
Ack	8.42E-04	4.98E-03	8.42e-01	4.98E-02	0.7	1.56	2.35	2.63
Sph	5.75E-11	1.05E-10	1.946e-05	4.5109e-004	0.00059	0.0057	0.0024 8	0.0211
Gri	7.88E-11	1.07E-07	7.18e-06	1.1182e-004	0.00016	9.299e-004	0.00040	0.0043
Mich	-1.8010	-1.8201	-1.80129	-1.8002	-1.80130	-1.8005	-1.801301	-1.8004
Zak	6.00E-07	0.00168	3.237e-07	8.4129e-005	1.3308e-07	6.4901e-005	8.63410e-07	7.0065e-005

In Table 4, It can be seen a comparison results between the 3 methods analyzed in this research, we can observe FPSO+FGA as a good alternative to resolve optimization problems. Also, the Neural Network (NN) was included to test this approach. Table 4 shows the simulation results for two variables.

Table 4. Comparison results between GA, PSO and FPSO+FGA

Math Funct	GA	PSO	FPSO+FGA	Objective Value
Ras	2.15E-03	5.47E-05	3.05E-04	0
Ros	1.02E-05	1.97E-03	1.17E-02	0
Ack	2.98	2.98	4.98E-03	0
Sph	1.62E-04	8.26E-11	1.05E-10	0
Grie	2.552E-05	2.56E-02	1.07E-07	0
Mich	-1.7829	-7.44E-01	-1.8201	-1.8013
Zakh	0.00146674	8.10	0.00168	0
NN	3.33E-01	2.3E-01	1.01E-03	0

8 Conclusions

The analysis of the simulation results of the evolutionary method considered in this paper, FPSO+FGA lead us to the conclusion that for the optimization of Modular Neural Networks with this method is a good alternative because it is easier to optimize the architecture of Modular Neural Network than to try it with PSO or GA separately. This is, because the combination PSO and GA with fuzzy rules gives a new hybrid method FPSO+FGA. It can be seen in Table 1 that the second and five architectures obtained after applying FPSO+FGA recognizes the ten images; we are demonstrating that it is reliable uses for this type of applications. Recently we are working with more images for test the effectiveness of this approach. In Table 4 it can be seen a comparison with PSO and GA used separately, we can observed as FPSO+FGA was better.

Acknowledgment

We would like to express our gratitude to CONACYT, Universidad Autónoma de Baja California and Tijuana Institute of Technology for the facilities and resources granted for the development of this research.

References

[1] Man K.F., Tang K.S., and Kwong S., "Genetic Algorithms: Concepts and Designs", Springer Verlag. 1999.

[2] Eberhart, R. C. and Kennedy, J. A new optimizer using particle swarm theory. Proceedings of the Sixth International Symposium on Micromachine and Human Science, Nagoya, Japan. pp. 39-43, 1995.J.-G. Lu, "Title of paper with only the first word capitalized," *J. Name Stand. Abbrev.*, in press.

[3] Kennedy, J. and Eberhart, R. C. Particle swarm optimization. Proceedings of IEEE International Conference on Neural Networks, Piscataway, NJ. pp. 1942-1948, 1995.

[4] Holland J.H., Adaptation in natural and artificial system, Ann Arbor, The University of Michigan Press, 1975.

[5] Valdez, F. and Melin P. "Parallel Evolutionary Computing using a cluster for Mathematical Function Optimization", Nafips. San Diego CA USA, 598-602. June 2007.

[6] Castillo, O.; Melin, P. "Hybrid intelligent systems for time series prediction using neural networks, fuzzy logic, and fractal theory" Neural Networks, IEEE Transactions on Volume 13, Issue 6, Page(s): 1395 – 1408. Nov. 2002.

[7] D.B. Fogel, "An introduction to simulated evolutionary optimization", IEEE transactions on neural networks, vol 5, n 1, Page(s):3 – 14, jan 1994.

[8] Goldberg D., Genetic Algorithms, Addison Wesley, 1988.

[9] Emmeche C., Garden in the Machine. The Emerging Science of Artificial Life, Princeton University Press, pp. 114. 1994.

[10] Angeline P. J., Using Selection to Improve Particle Swarm Optimization , In Proceedings 1998 IEEE World Congress on Computational Intelligence, Anchorage, Alaska, IEEE, 84-89. 1998.

[11] Angeline P.J., Evolutionary Optimization Versus Particle Swarm Optimization: Philosophy and Performance Differences, Evolutionary Programming VII, Lecture Notes in Computer Science 1447, Springer, :601-610. 1998.

[12] Back T., Fogel D.B., and Michalewicz Z., (Eds), Handbook of Evolutionary Computation. Oxford University Press, 1997.

[13] Montiel O, Castillo O, Melin P, Rodriguez A and Sepulveda R: "Human evolutionary model: A new approach to optimization." Inf. Sci. 177 (10): 2075-2098, 2007.

[14] Castillo O., Valdez F. and Melin P., "Hierarchical Genetic Algorithms for topology optimization in fuzzy control systems". International Journal of General Systems, Volume 36, Issue 5., pag 575-591, October 2007.

[15] Valdez F., Melin P. and Castillo O. "Evolutionary Computing for the Optimization of Mathematical functions". Analysis and Design of Intelligent Systems Using Soft Computing Techniques. Advances in Soft Computing 41. June 2007.

[16] Castillo O, Huesca G, and Valdez F. "Proceedings of the International Conference on Artificial" Intelligence, IC-AI '04, Las Vegas, Nevada,USA, Volume 1, pag. 98-104,June 21-24, 2004.

[17] Veeramachaneni K, Osadciw L and Yan W "Improving Classifier Fusion Using Particle Swarm Optimization", IEEE Fusion Conference, Italy, July, 2006.

[18] M and Kennedy J. "The particle swarm-explosion, stability, and convergence in a multidimensional complex space". IEEE Transactions on Evolutionary Computation, 6(1):58-73. 2002.

[19] Kennedy J and Mendes R. "Population structure and particle swarm performance".Proceeding of IEEE conference on Evolutionary Computation, pag. 1671-1676. 2002.

[20] Wei W, Jiatao S, Zhongzxiu Y and Zheru . "Wavelet-based Illumination Compensation for Face Recognition using Eifenface Method Intelligent Control and Automation". WCICA 2006. Volume 2, Issue , 21-23 June 2006 Page(s): 10356 – 10360, 2006.

[21] Muller B, Reinhardt J, StricklandM. Neural networks: An introduction. 2nd edn, Springer-Verlag, Berlin. 1995.

[22] McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bull of Math Biophysics 5:115–133. 1943.

Fuzzy Subgroups and the Teichmüller Space

Paavo Kukkurainen

Department of Mathematics and Physics, Lappeenranta University of Technology
Lappeenranta, Finland
Email: paavo.kukkurainen@lut.fi

Abstract— *There exists a generalization of the Teichmüller space of a covering group. In this paper we combine this generalized Teichmüller space $T(G)$ and any fuzzy subgroup $\mathcal{A}: G \longrightarrow \mathcal{F}$ where G is a subgroup of the group consisting of such orientation preserving and orientation reversing Möbius transformations which act in the upper half-plane of the extended complex plane. A partially ordered set $\mathcal{F} = (\mathcal{F}, \leq)$ consists of stabilizers of G and all of their intersections. After preliminaries we present two new results concerning this special case of fuzzy subgroups. These conclusions are then applied to the known theory of the parametrization of the generalized Teichmüller space . As consequence, the equivalence classes of fuzzy subgroups (with an equivalence relation) become the elements of $T(G)$ where G is generated by a finite set of hyperbolic Möbius transformations. Let the number of the generators be n. Then there is an embedding $\psi: T(G) \longrightarrow \mathbb{R}^{3n-3}$ and therefore a homeomorphism $T(G) \longrightarrow \psi(T(G))$. Through parametrization of $T(G)$ $\psi(\mathcal{A}(G))$ has $3n-3$ real coordinates which are also the coordinates of $\mathcal{A}(G)$ up to identification.*

Keywords— Fuzzy subgroups, Möbius groups, Möbius transformations, compact Riemann surfaces, Teichmüller spaces.

1 Introduction

Orientation preserving and orientation reversing Möbius transformations form a group denoted by $\widehat{M}$. Following [2], we say that every subgroup of $\widehat{M}$ is a Möbius group. In Section 2 we first take some basic knowledge of Möbius transformations and then list all the types of Möbius transformations acting in the upper half-plane H which means that they are self-mappings of H. Let G and G' be Möbius groups fixing H. Then we define geometric isomorphisms $j: G \longrightarrow G'$ which play an important role in our representation.

In Section 3 we present two new Propositons 1 and 2 which the author of this paper has proved in [1]. It is not relevant to give the proofs in this connection because they are too long to show as well as many preliminaries are required. The propositions constitute the relationship between subgroups of the group of Möbius transformations and certain kind of fuzzy subgroups.

Let G, G', G_1 and G_2 be Möbius groups with a fixed G. Defining a delation $\delta(j)$ of an isomorphism $j: G \longrightarrow G'$ by means of multipliers of Möbius transformations $g \in G$ we say that any two isomorphisms $j_1: G \longrightarrow G_1$ and $j_2: G \longrightarrow G_2$ with some properties are equivalent precisely when $\delta(j_2 j_1^{-1}) = 1$. The set $E(G)$ of equivalence classes $[j]$ forms a metric space $(E(G), d)$ with a metric d. In the theory of compact Riemann surfaces this space can be regarded as a generalization of the Teichmüller space of a covering group G. So, $(E(G), d)$ is denoted by $T(G)$. All the presented premiliminaries and the parametrization of the generalized Teichmüller space can be found from [2], [3] and [4]. We will consider them in Section 4 so much as needed for Section 5.

Suppose that there exists a set F in G such that the parametrization condition in Proposition 3 is satisfied. Then we obtain an embedding $\psi: T(G) \longrightarrow \mathbb{R}^{3n-3}$ where G is finitely generated by n hyperbolic elements. More precisely, $T(G) \longrightarrow \psi(T(G))$ is a homeomorphism. Then, up to homeomorphism, $[j]$ is an element of $\mathbb{R}^{3n-3}$ with coordinates $k(j(g))$ which are the multipliers of Möbius transformations $j(g)$ for every $g \in F$.

In Section 5 we apply fuzzy subgroups to the theory of the parametrization of the generalized Teichmüller space. According to Corollary 1 the type-preserving isomorphisms $j_1: G \longrightarrow G_1$ and $j_2: G \longrightarrow G_2$ as well as the fuzzy subgroups $\mathcal{A}(j_1(g))$ and $\mathcal{A}(j_2(g))$ are simultaneously equivalent. The definition of the fuzzy subgroups is given by (3) in Proposition 1 and the equivalence between two fuzzy subgroups is defined by conjugation with a Möbius transformation. Roughly speaking, it is now possible to replace any element $[j]$ by the corresponding $[\mathcal{A}(j(g))]$ in Proposition 4. This leads to the representation of $[\mathcal{A}(j(g))]$ with real coordinates (up to identification) in the Euclidean space $\mathbb{R}^{3n-3}$.

In the following special case, we refer to [3]: If G is a covering group of the upper half-plane H over some compact Riemann surface S, then $[j]$ has coordinates in $\mathbb{R}^{6s-6+3n}$ where S is of genus s and n is the number of conformal disks removed from S. The coordinates of $[\mathcal{A}(j(g))]$ are given in subsection 5.2.

2 Preliminaries for Möbius transformations

In this section we refer to [2].

2.1 Möbius transformations

Directly conformal automorphisms of the extended complex plane $\widehat{\mathbb{C}}$ are orientation preserving Möbius transformations

$$g(z) = \frac{az+b}{cz+d} \quad a,\, b,\, c,\, d \in \mathbb{C}, \quad ad-bc = 1, \qquad (1)$$

and $g(-\frac{d}{c}) = \infty$, $g(\infty) = \frac{a}{c}$.

Indirectly conformal automorphisms of $\widehat{\mathbb{C}}$ are of the form

$$g(z) = \frac{a\bar{z}+b}{c\bar{z}+d} \quad a,\, b,\, c,\, d \in \mathbb{C}, \quad ad-bc = -1 \qquad (2)$$

which are orientation reversing Möbius transformations. The mappings (1) and (2) form the group $\widehat{M}$. Two transformations g_1 and g_2 in $\widehat{M}$ are conjugate if $g_1 = h g_2 h^{-1}$ for some Möbius transformation h.

ISBN: 978-989-95079-6-8

A point z is said to be a fixed point of g if $g(z) = z$. Every non-identity orientation preserving Möbius transformation g has one fixed point or two fixed points. Then

- g with one fixed point is parabolic and it is conjugate to $z \mapsto z+1$.
- If g has two fixed points and it is conjugate to $z \mapsto kz$ for some $k \in \mathbb{C} \setminus \{0\}$, then g is loxodromic if $|k| \neq 1$ and elliptic if $|k| = 1, k \neq 1$.
- Loxodromic transformations fixing the upper half-plane H are called hyperbolic. Otherwise they are strictly loxodromic.

In the loxodromic case we set

$$P(g) = \lim_{n\to\infty} g^n \text{ and } N(g) = \lim_{n\to\infty} g^{-n},$$

which are the attracting and repelling fixed points of g. Because the fixed points of any loxodromic g are real iff $g(H) = H$, hyperbolic transformations have fixed points on the real axis $\mathbb{R}$. Therefore $P(g)$ and $N(g)$ are real if g is hyperbolic. A circle or line in H perpendicular to to the real axis is a non-Euclidean line. Now the non-Euclidean line through $P(g)$ and $N(g)$ is called the axis $ax(g)$ of g. In the elliptic case $g(H) = H$ iff the fixed points of g are complex conjugates.

The multiplier of a Möbius transformation g is defined by means of the cross-ratio

$$k(g) = (g(z), z, x, y) = \frac{g(z)-x}{g(z)-y}\,\frac{z-y}{z-x},$$

where x and y are two different fixed points of g. In the parabolic case we set $k(g) = 1$. For the above complex number k we have $k = k(g)$. The multiplier is invariant in conjugation.

We are only interested in such orientation reversing Möbius transformations which act in H (fix H). This leads to the next consideration: the axis $ax(\sigma) = \{z \mid \sigma(z) = z\}$ of a reflection σ fixing H is also a circle or line orthogonal to $\mathbb{R}$. Denote by x and y the real fixed points of σ. Then

- the reflection $\sigma(z) = \eta(\bar{z})$ where η is the elliptic transformation defined by $k(\eta) = -1$, $\eta(x) = x$ and $\eta(y) = y$,
- a glide-reflection s fixing H is of the form $s = \tau\sigma$ where τ is a hyperbolic transformation fixing H, σ is a reflection fixing H and $ax(\tau) = ax(\sigma)$.

The reflection σ has an infinite number of fixed points but two real fixed points (if $ax(\sigma)$ is not a line). For the multiplier we set $k(\sigma) = -1$. The glide-reflection s and the hyperbolic τ have the same two fixed points. Moreover, the multiplier $k(s) = -k(\tau)$.

It is known that transformations (1) and (2) fix the upper-half plane H iff the coefficients a, b, c, d are real. In fact, there are the following types of Möbius transformations fixing H: hyperbolic, parabolic and elliptic transformations, the identity transformation, reflections and glide-reflections.

Suppose that G and G' are groups of Möbius transformations (or Möbius groups) acting in H. An isomorphism $j : G \longrightarrow G'$ is induced by a Möbius transformation h if $j(g) = hgh^{-1}$ for all $g \in G$. We say that $j : G \longrightarrow G'$ is type-preserving if g and $j(g)$ are of the same type for all $g \in G$. A type-preserving isomorphism $j : G \longrightarrow G'$ is geometric on $\mathbb{R}$ if there exists a homeomorphism $\varphi : \widehat{\mathbb{R}} \longrightarrow \widehat{\mathbb{R}}$ inducing j on $\widehat{\mathbb{R}}$, which is the boundary of the upper half-plane. Especially $\varphi(P(g)) = P(j(g))$ and $\varphi(N(g)) = N(j(g))$ for any hyperbolic g. Since every Möbius transformation is a homeomorphism $\widehat{\mathbb{C}} \longrightarrow \widehat{\mathbb{C}}$, then for every φ which induces a geometric isomorphism j we have $h \mid \widehat{\mathbb{R}} = \varphi$ ($h : \widehat{\mathbb{C}} \longrightarrow \widehat{\mathbb{C}}$ is a Möbius transformation).

3 Connection between fuzzy subgroups and Möbius transformations

Let $(P, \leq)$ be a partially ordered set and A a nonempty set.

(i) A mapping $\mathcal{A} : A \longrightarrow P$ is a P-(fuzzy)set on A.

(ii) For every $p \in P$, $A_p = \{x \in A \mid \mathcal{A}(x) \geq p\}$ is a p-level set subset (in short a level subset) of $\mathcal{A}$.

(iii) Let $\mathcal{G} = (G, \circ)$ be a group such that $(A_p, \circ)$ is a subgroup of $\mathcal{G}$ for every $p \in P$. Then a P-set on G, $\mathcal{A} : G \to P$, is said to be a P-(fuzzy) subgroup of $\mathcal{G}$.

Proposition 1. *[1] Let G be a subgroup of the group of Möbius transformations and let $G_z = \{g \in G \mid g(z) = z\}$ be a stabilizer of G at $z \in \widehat{\mathbb{C}}$. If $\mathcal{F}$ is a set of stabilizers of G and all of their intersections, and $\mathcal{F} = (\mathcal{F}, \leq)$ is a partially ordered set under $p \leq q$ iff $p \supseteq q$ ($p, q \in \mathcal{F}$), then $\mathcal{A} : G \to \mathcal{F}$*

$$\mathcal{A}(g) = \bigcap (p \in \mathcal{F} \mid g \in p) = \bigcap (G_z \mid g \in G_z) \qquad (3)$$

is a $\mathcal{F}$-subgroup of G.

We set an equivalence relation between fuzzy subgroups (3) by conjugation: $\mathcal{A}(g_1) \sim \mathcal{A}(g_2)$ iff there exists a Möbius transformation h such that $\mathcal{A}(g_1) = h\mathcal{A}(g_2)h^{-1}$. Observe that stabilizers form a group and the conjugation of the intersection of the groups G_z occurs by elements in formula (3). Recall that a Möbius group is a subgroup of the group $\widehat{M}$.

Proposition 2. *[1] Let G and G' be Möbius groups acting in the upper half-plane, $\mathcal{F}$ a set of stabilizers of G and all of their intersections. Let $\mathcal{A} : G \longrightarrow \mathcal{F}$ be defined by (3) and suppose that there is a type-preserving isomorphism $j : G \longrightarrow G'$. Then the following conditions are equivalent:*

(i) $\mathcal{A}(g) \sim \mathcal{A}(j(g))$ for every $g \in G$,

(ii) $k(g) = k(j(g))$ for every $g \in G$,

(iii) j is induced by a Möbius transformation,

(iv) $j : G \longrightarrow G'$ is a geometric isomorphism.

Further, let $j_1 : G \longrightarrow G_1$ and $j_2 : G \longrightarrow G_2$ be type-preserving isomorphisms. Then

$$\mathcal{A}(j_1(g)) \sim \mathcal{A}(j_2(g)) \quad \text{for every } g \in G$$

iff

$$j_2 j_1^{-1} : G_1 \longrightarrow G_2 \quad \text{is induced by a Möbius transformation.}$$

4 Coordinates of the generalized Teichmüller space

In this section we refer to [3].

ISBN: 978-989-95079-6-8

4.1 The set E

Let G and G' be Möbius groups acting in the upper half-plane H and $j : G \longrightarrow G'$ an isomorphism. We suppose that G has a set $E = \{g_1, g_2, \cdots\}$ of hyperbolic generators satisfying the following conditions:

(i) $ax(g_1) \cap ax(g_2)$ is a set of two points,

(ii) $(ax(g_1) \cup ax(g_2)) \cap ax(g_i) = \emptyset$, $i = 3,4,\cdots$,

(iii) $(N(g_i), P(g_i), N(g_1), P(g_1)) < 1$, $i = 3,4,\cdots$.

4.2 The dilation

Let $j : G \longrightarrow G'$ be an isomorphism between Möbius groups. The dilation $\delta(j)$ of j is the smallest number of the numbers $a \geq 1$ satisfying

$$|k(g)|^{\frac{1}{a}} \leq |k(j(g))| \leq |k(g)|^a \tag{4}$$

for all $g \in G$. If there exists no number $a \geq 1$ we define $\delta(j) = \infty$.

For a fixed G, let $J(G)$ be a set of all isomorphisms $j : G \longrightarrow G'$ with the following properties:

(i) G and G' are Möbius groups acting in H,

(ii) the dilation $\delta(j)$ is finite,

(iii) there exists a homeomorphism $\varphi : \widehat{\mathbb{R}} \longrightarrow \widehat{\mathbb{R}}$ such that $\varphi(P(g_i)) = P(j(g_i))$ and $\varphi(N(g_i)) = N(j(g_i))$ for all $g_i \in E$.

Define in $J(G)$ an equivalence relation $\sim$ by setting $j_1 \sim j_2$ iff $\delta(j_2 j_1^{-1}) = 1$. Then the set $E(G)$ of equivalence classes $[j]$ becomes a metric space $(E(G), d)$ with the dilation metric d defined by $d([j_1],[j_2]) = log\ \delta(j_2 j_1^{-1})$ [4].

Definition 1. *[3] The space $T(G) = (E(G), d)$ is called the generalized Teichmüller space of a Möbius group G.*

4.3 Parametrization of the Teichmüller space $T(G)$

Let us construct a set F containing the next elements:

(1) g_i, $i = 1,2,\cdots$,

(2) $g_i g_1$, $i = 2,3,\cdots$,

(3) $g_2 g_i g_2^{-1} g_i^{-1}$, $i = 3,4,\cdots$,

where every $g_i \in E$. The following proposition shows the significance of the set F if we want to find the condition for the parametrization of $T(G)$. We say that F parametrizes the space $T(G)$ or the set $J(G)$.

Proposition 3. *[3] Suppose that there exists a homeomorphism $\varphi : \widehat{\mathbb{R}} \longrightarrow \widehat{\mathbb{R}}$ for which $\varphi(P(g_i)) = P(j(g_i))$ and $\varphi(N(g_i)) = N(j(g_i))$ for all $g_i \in E$. If*

$$k(j(g)) = k(g)$$

for all $g \in F$, then j is induced by a Möbius transformation $h : \widehat{\mathbb{C}} \longrightarrow \widehat{\mathbb{C}}$.

Since multipliers are invariant in conjugation the converse holds trivially.

Proposition 4. *[3] Suppose that $E = \{g_1, \cdots, g_n\}$ is finite. Then there is an embedding*

$$\psi : T(G) \longrightarrow \mathbb{R}^{3n-3},\ \psi([j]) = (x_1, \cdots, x_{3n-3}),$$

where

$$\begin{aligned} x_i &= k(j(g_i)),\ i = 1,\cdots,n, \\ x_{n-1+i} &= k(j(g_i)j(g_1)),\ i = 2,\cdots n, \\ x_{2n-3+i} &= k(j(g_2)j(g_i)j(g_2)^{-1}j(g_i)^{-1}),\ i = 3,\cdots,n. \end{aligned}$$

Proof. The mapping $\psi : T(G) \longrightarrow \mathbb{R}^{3n-3}$ is continuous with the dilation metric d [3], [4]. By Proposition 3, ψ is injective [3]. Furthermore, the inverse $\psi^{-1} : \psi(T(G)) \longrightarrow T(G)$ is continuous. Hence $\psi : T(G) \longrightarrow \psi(T(G))$ is a homeomorphism. □

5 Fuzzy subgroups in the Teichmüller space

5.1 The main results

Let G, G_1 and G_2 be Möbius groups. According to Proposition 2 and inequalities (4) we conclude

Corollary 1. *Let $j_1 : G \longrightarrow G_1$ and $j_2 : G \longrightarrow G_2$ be type-preserving isomorphisms. Then the following statements are equivalent:*

(i) $j_1 \sim j_2$,

(ii) $\delta(j_2 j_1^{-1}) = 1$,

(iii) $k(j_2 j_1^{-1}(g)) = k(g)$ *for every* $g \in G$,

(iv) $j_2 j_1^{-1} : G_1 \longrightarrow G_2$ *is induced by a Möbius transformation,*

(v) $\mathcal{A}(j_1(g)) \sim \mathcal{A}(j_2(g))$ *for every* $g \in G$.

Let $j \in J(G)$ and suppose that $j : G \longrightarrow G'$ is geometric. Then j is type-preserving and satisfies the condition *(iii)* in subsection 4.2. Following the definition of $E(G)$, let us denote the set of equivalence classes $[\mathcal{A}(j(g))]$ by $E'(G)$. By Corollary 1,*(i)*,*(v)*, $(E'(G), d)$ becomes also a metric space having the same metric d. This leads to the new form of the generalized Teichmüller space, $T'(G) = (E'(G), d)$. Reformulation of Proposition 4 for the fuzzy subgroups $\mathcal{A}(j(g))$ yields

Proposition 5. *Suppose that $E = \{g_1, \cdots, g_n\}$ is finite. Then there is an embedding*

$$\psi' : T'(G) \longrightarrow \mathbb{R}^{3n-3},\ \psi'([\mathcal{A}(j(g)]) = (x_1, \cdots, x_{3n-3}),$$

for every $g \in G$, and where

$$\begin{aligned} x_i &= k(j(x_i)),\ i = 1,\cdots,n, \\ x_{n-1+i} &= k(j(g_i)j(g_1)),\ i = 2,\cdots n, \\ x_{2n-3+i} &= k(j(g_2)j(g_i)j(g_2)^{-1}j(g_i)^{-1}),\ i = 3,\cdots,n. \end{aligned}$$

ISBN: 978-989-95079-6-8

5.2 The covering group of a compact Riemann surface

Let G be a Möbius group acting in the upper half-plane H and suppose that G is a covering group of H over the compact Riemann surface S. Since any non-identity cover transformation of G has no fixed points in H, we know that all elements of G are hyperbolic or parabolic Möbius transformations, or it is the identity mapping. We say that G is of signature (s,n) if $S = H/G$ is of genus s from which n conformal disks are removed.

Assume that G is a fixed covering group of signature (s,n) with $s > 0$ and $n > 0$. Then G has $2s+n$ hyperbolic generators satisfying the defining relation:

$$c_n \cdots c_1\, b_s^{-1}\, a_s^{-1}\, b_s\, a_s \cdots b_1^{-1}\, a_1^{-1}\, b_1\, a_1 \;=\; id$$

from which c_n can be solved. As conclusion the set

$$E \;=\; \{a_1, b_1, a_2^{-1}, b_2, \cdots, a_s^{-1}, b_s, c_1^{-1}, \cdots, c_{n-1}^{-1}\}$$

generates G freely. Moreover, E satisfies the conditions $(i)-(iii)$ in subsection 4.1. Then the mapping $\psi' : T'(G) \longrightarrow \mathbb{R}^{3n-3}$ in Proposition 5 takes the form

$$\psi' : T'(G) \longrightarrow \mathbb{R}^{6s-6+3n},\;\; \psi'([\mathcal{A}(j(g)]) = (x_1, \cdots, x_{6s-6+3n})$$

with coordinates

$$\begin{aligned}
x_t \;=\;& k(j(a_i)),\;\; i = 1, \cdots, s,\\
& k(j(b_i)),\;\; i = 1, \cdots, s\\
& k(j(c_i)),\;\; i = 1, \cdots, n-1,\\
& \text{for } t = 1, \cdots, 2s+n-1,\\
x_{2s+n-2+t} \;=\;& k(j(b_i)j(a_1)),\;\; i = 1, \cdots, s,\\
& k(j(a_i)^{-1})j(a_1)),\;\; i = 2, \cdots, s,\\
& k(j(c_i)^{-1})j(a_1))\;\; i = 1, \cdots, n-1,\\
& \text{for } t = 2, \cdots 2s+n-1,\\
x_{4s+2n-5+t} \;=\;& k(j(b_1)j((a_i)^{-1})j((b_1)^{-1})j(a_i)),\\
& i = 2, \cdots, s\\
& k(j(b_1)(j(b_i)j((b_1)^{-1})j((b_i)^{-1})),\\
& i = 2, \cdots, s\\
& k(j(b_1)j((c_i)^{-1})j((b_1)^{-1})j(c_i)),\\
& i = 1, \cdots, n-1,\\
& \text{for } t = 3, \cdots, 2s+n-1.
\end{aligned}$$

Example 1. Let us construct a compact Riemann surface $S = H/G$ with a covering group G of signature $(1,1)$. Then S consists of one handle from which one conformal disk is removed. Therefore G has a set $E = \{a, b, c^{-1}\}$ of hyperbolic generators with the defining relation

$$c^{-1}\, a\, b \;=\; id,$$

where c is a boundary mapping of the disk satisfying $c = ab$. Then a and b generates G freely, $G =< a, b >$. In fact, the hyperbolic a and b have no comman fixed points and G is a purely hyperbolic group. It does not contain parabolic or elliptic elements. The first one is concluded from [3] and the second one holds because G is a covering group.

An embedding

$$\begin{aligned}
&\psi' : E'(G) \longrightarrow \mathbb{R}^3,\\
&\psi'([\mathcal{A}(j(g)]) = (k(j(a_1)), k(j(b_1)), k(j(b_1)j(a_1)))
\end{aligned}$$

gives the coordinates $(k(j(a_1)), k(j(b_1)), k(j(b_1)j(a_1)))$ of $[\mathcal{A}(j(g)]$ up to homeomorphism for any $g \in G$ and some geometric isomorphism j of G and $j \in J(G)$.

References

[1] P. Kukkurainen, Fuzzy subgroups, algebraic and topological points of view and complex analysis. Thesis for the degree of Doctor of Philosophy, Acta Universitatis Lappeenrantaensis, 2009. In printed.

[2] M. Seppälä and T. Sorvali, Geometry of Riemann Surfaces and Teichmüller Spaces, Elsevier Science Publishers B.V., North-Holland Amsterdam, 1992.

[3] T. Sorvali, Parametrization of free Möbius groups, Annales Academiae Scientiarum Fennicae. Series A. I.Mathematica 579, 1974, 1-12.

[4] T. Sorvali, The boundary mapping induced by an isomorphism of covering groups, Annales Academiae Scientiarum Fennicae. Series A. I.Mathematica 526, 1972, 4-31.

ISBN: 978-989-95079-6-8

Reasoning about Actions in Fuzzy Environment

Emad Saad[1] Shaimaa A. Elmorsy[2] Mahmoud M. H. Gabr[2] Yasser F. M. Hassan[2]

1. Department of Computer Science, Gulf University for Science and Technology
West Mishref, Kuwait
2. Department of Mathematics and Computer Science, Faculty of Science, Alexandria University
Alexandria, Egypt
Email: saad.e@gust.edu.kw, shaimaa_elmorsy@hotmail.com, mahgabr@yahoo.com, y.fouad@sci.alex.edu.eg

Abstract— *Reasoning in the presence of imprecision and vagueness is inevitable in many real-world applications including those in robotics and intelligent agents. Although, reasoning about actions is a major component in these real-world applications, current actions languages for reasoning about actions lack the ability to represent and reason about actions in the presence of imprecision and vagueness that stem from effects of actions in these real-world applications. In this paper we present a new action language called fuzzy action language, $\mathcal{A}_F$, that allows the representation and reasoning about actions with vague (fuzzy) effects. In addition we define the notions of fuzzy planning and fuzzy plan in the fuzzy action language $\mathcal{A}_F$. Furthermore, we describe a fuzzy planner based on the fuzzy action language A_F that is developed by translating a fuzzy action theory,* $\mathbf{FT}$*, in A_F into a normal logic program with answer set semantics,* Π*, where trajectories in* $\mathbf{FT}$ *are equivalent to the answer sets of* Π*. In addition, we formally prove the correctness of our translation. Furthermore, we show that fuzzy planning problems can be encoded as a SAT problem.*

Keywords— Answer set programming, fuzzy planning, reasoning about actions with fuzzy effects, reasoning under uncertainty.

1 Introduction

Reasoning about the properties of actions is fundamental in many real-world applications. Therefore, many action languages that allow representing and reasoning about actions have been developed which include [2, 5, 9, 10, 11, 12, 18, 21, 22, 24]. Furthermore, uncertainty is a main issue in representing and reasoning about actions in these real-world applications. Uncertainty in these real-world applications stems from different sources including incompleteness, erroneous, or imprecision (vagueness). For reasoning under uncertainty, probability theory is used to reason in the presence of incompleteness or erroneous, however, fuzzy set theory is used to reason in the presence of imprecision or vagueness. To deal with probabilistic uncertainty in reasoning about actions, different proposals have been presented. These proposals include [2, 5, 12, 21, 22]. Although, these proposals deal with uncertainty probabilistically, they are inappropriate in situations where actions have imprecise or vague effects, which need a different formalism from the one required in reasoning about actions with probabilistic effects as in [2, 5, 12, 21, 22]. This is because the underlying formalism in reasoning about actions with probabilistic effects is the probability theory, however, the formalism required for reasoning about actions with imprecise effects is the fuzzy set theory. Actions with imprecise effects appear in many domains including robotics and intelligent agents. Consider the following example adapted from [23]. Consider a planner that controls an autonomous car that is moving in a highway with the goal to keep the car close to the middle of a lane. The autonomous car planner's uses the action navigate whose effect is to make the car close to the middle of the lane. The effect of the action navigate is neither probabilistically nor precisely defined. Therefore, action languages, for example [9, 10, 11, 18, 24], that reason about actions whose effects are precisely defined (either entirely true or entirely false), and action languages, for example [2, 5, 12, 21, 22], that reason about actions whose effects are probabilistically defined (either true or false to a probabilistic degree), can not be used to correctly represent and reason about these kind of actions whose effects are imprecisely or vaguely defined.

In this paper we present a new action language called fuzzy action language, $\mathcal{A}_F$, that allows the representation and reasoning about actions with vague (fuzzy) effects. In addition, we define the notions of fuzzy planning and fuzzy plan in the fuzzy action language $\mathcal{A}_F$. Furthermore, we describe a fuzzy planner based on the fuzzy action language $\mathcal{A}_F$ that is developed by translating a fuzzy action theory, $\mathbf{FT}$, in $\mathcal{A}_F$ into a normal logic program with answer set semantics, Π, where trajectories in $\mathbf{FT}$ are equivalent to the answer sets of Π. In addition, we formally prove the correctness of our translation. Furthermore, we show that fuzzy planning problems can be encoded as SAT problems.

This paper is organized as follows. Section 2 reviews the answer set semantics of normal logic programs and the fundamental notions of the fuzzy set theory. The syntax and semantics of the fuzzy action language, $\mathcal{A}_F$, is introduced in section 3. In section 4, a translation from a fuzzy planning problem in the fuzzy action language $\mathcal{A}_F$ into a normal logic program with answer set semantics is presented. The correctness of the translation is introduced in section 5. Finally, conclusion and related work is presented in section 6.

2 Preliminaries

In this section we review the answer set semantics of normal logic programs [7] and the basic notions of fuzzy set theory [27].

2.1 Normal Logic Programs with Answer Set Semantics

Let $\mathcal{L}$ be a first-order language with finitely many predicate symbols, function symbols, constants, and infinitely many variables. The Herbrand base of $\mathcal{L}$ is denoted by $\mathcal{B}$. A Her-

ISBN: 978-989-95079-6-8

brand interpretation is a subset of the Herbrand base $\mathcal{B}$. A normal logic program is a finite set of rules of the form

$$a \leftarrow a_1, \ldots, a_n, not\ b_1, \ldots, not\ b_m$$

where a, $a_1, \ldots, a_n, b_1, \ldots, b_m$ are atoms and not is the negation-as-failure. Intuitively, the meaning of the above rule is that if it is believable (provable) that a_i is true and it is not believable that b_i is true then it is believable that a is true. A normal logic program is ground if no variables appear in any of its rules.

Let Π be a ground normal logic program, S be a Herbrand interpretation, and r be a rule as above. Then, we say that

- S satisfies a_i (denoted by $S \models a_i$) iff $a_i \in S$.
- S satisfies $not\ b_j$ (denoted by $S \models not\ b_j$) iff $b_j \notin S$.
- S satisfies $(a_1, \ldots, a_n, not\ \ b_1, \ldots, not\ \ b_m)$ (denoted by $S \models (a_1, \ldots, a_n, not\ \ b_1, \ldots, not\ \ b_m)$) iff $\{a_1, \ldots, a_n\} \subseteq S$ and $\{b_1, \ldots, b_m\} \nsubseteq S$.
- S satisfies $a \leftarrow a_1, \ldots, a_n, not\ b_1, \ldots, not\ b_m$ iff $S \models a$ or $S \not\models (a_1, \ldots, a_n, not\ b_1, \ldots, not\ b_m)$.

A Herbrand interpretation is said to satisfy a normal logic program Π iff it satisfies every rule in Π. A Herbrand model of a normal logic program Π is a Herbrand interpretation of Π that satisfies Π. A Herbrand interpretation S of a normal logic program Π is said to be an answer set of Π if S is the minimal Herbrand model (with respect to the set inclusion) of the reduct, denoted by Π^S, of Π w.r.t. S, where Π^S is a set of rules of the form

$$a \leftarrow a_1, \ldots, a_n$$

such that

$$a \leftarrow a_1, \ldots, a_n, not\ b_1, \ldots, not\ b_m \in \Pi$$

and $\{b_1, \ldots, b_m\} \nsubseteq S$

Intuitively, for any $not\ \ b_j$ in the body of a rule in Π with $b_j \notin S$ is simply satisfied by S, and $not\ b_j$ is safely removed from the body of the rule. If $b_j \in S$ then the body of the rule is not satisfied and the rule is trivially ignored.

2.2 Fuzzy Sets

In this section we review the basic notions of fuzzy sets as presented in [27]. Let U be a set of objects. A fuzzy set, F, in U is defined by the grade membership function $\mu_F : U \rightarrow [0, 1]$, where for each element $x \in U$, μ_F assigns to x a value $\mu_F(x)$ in $[0, 1]$. The support for F denotes the set of all objects x in U for which the grade membership of x in F is a non-zero value. Formally, $support(F) = \{x \in U \mid \mu_F(x) > 0\}$. The intersection (conjunction) of two fuzzy sets F and F' in U, denoted by $F \wedge_f F'$ is a fuzzy set G in U where the grade membership function of G is $\mu_G(x) = \min(\mu_F(x), \mu_{F'}(x))$ for all $x \in U$. However, the union (disjunction) of two fuzzy sets F and F' in U, denoted by $F \vee_f F'$ is a fuzzy set G in U where the grade membership function of G is $\mu_G(x) = \max(\mu_F(x), \mu_{F'}(x))$ for all $x \in U$. The complement (negation) of a fuzzy set F in U is a fuzzy set in U denoted by $\overline{F}$ where the grade membership function of $\overline{F}$ is $\mu_{\overline{F}}(x) = 1 - \mu_F(x)$ for all $x \in U$. A fuzzy set F in U is said to be contained in another fuzzy set G in U if and only if $\mu_F(x) \leq \mu_G(x)$ for all $x \in U$. Notice that we use the notations $\wedge_f$ and $\vee_f$ to denote fuzzy conjunction and fuzzy disjunction respectively to distinguish them from $\wedge$ and $\vee$ for propositional conjunction and disjunction respectively. Furthermore, other function characterizations for the fuzzy conjunction and fuzzy disjunction operators can be used. However, we will stick with the min and max function characterizations for the fuzzy conjunction and fuzzy disjunction as originally proposed in [27].

3 Fuzzy Action Language $\mathcal{A}_F$

In this section we develop a novel action language, called *fuzzy action language*, $\mathcal{A}_F$, that allows the representation of actions with vague (fuzzy) effects. An action theory in $\mathcal{A}_F$ is capable of representing the grade membership distribution of the possible initial states, the executability conditions of actions, and the grade membership distribution of the fuzzy effects of actions. The semantics of $\mathcal{A}_F$ is based on a transition function that maps an action and a set of states to a set of states. The fuzzy action language $\mathcal{A}_F$ is inspired by the action languages in [21, 22].

3.1 Language syntax

A predicate that describes a property of the environment, which may contain variables, is called a fluent. Let $\mathcal{A}$ be a set of action names that can contain variables and $\mathcal{F}$ be a set of fluents. A fluent $f \in \mathcal{F}$ or $\neg f$, the negation of f is called a fluent literal. A conjunction of fluent literals of the form $l_1 \wedge \ldots \wedge l_n$ is called a conjunctive fluent formula, where $l_1, \ldots, l_n$ are fluent literals. Sometimes we abuse the notation and refer to a conjunctive fluent formula as a set of fluent literals ($\emptyset$ denotes $true$).

A fuzzy action theory, $\mathbf{FT}$, in $\mathcal{A}_F$ is a tuple of the form $\mathbf{FT} = \langle S_0, \mathcal{AD} \rangle$, where S_0 is a proposition of the form (1), $\mathcal{AD}$ is a set of propositions from (2-3) as follows:

$$\textbf{initially} \begin{cases} \psi_1 \ : \ v_1 \\ \psi_2 \ : \ v_2 \\ \quad \ldots \\ \psi_n \ : \ v_n \end{cases} \tag{1}$$

$$\textbf{executable}\ a\ \textbf{if}\ \psi \tag{2}$$

$$a\ \textbf{causes} \begin{cases} \phi_1 \ : \ v_1\ \textbf{if}\ \psi_1 \\ \phi_2 \ : \ v_2\ \textbf{if}\ \psi_2 \\ \quad \ldots \\ \phi_n \ : \ v_n\ \textbf{if}\ \psi_n \end{cases} \tag{3}$$

where $\psi, \psi_1, \ldots, \psi_n, \phi_1, \ldots, \phi_n$ are conjunctive fluent formulae, $a \in \mathcal{A}$ is an action, and for all $1 \leq i \leq n$, we have $v_i \in [0, 1]$. The set of all ψ_i must be mutually exclusive. Let $\mathcal{S}$ be the set of all states formed from the fluents in $\mathcal{F}$.

Proposition (1) is a fuzzy set in the set of all states $\mathcal{S}$. It represents the grade membership of the possible initial states. We consider that only ψ_i with grade membership that is non-zero is listed in (1). Proposition (1) says that the grade membership of the possible initial state ψ_i is v_i for all $1 \leq i \leq n$. Proposition (2) represents the *executability condition* of actions, where each variable that appears in a also appears in ψ. It states that an action a is executable in any state in which ψ holds. A proposition of the form (3) represents the fuzzy (vague) effects resulting from executing an action a in the

ISBN: 978-989-95079-6-8

states in which a is executable. For each $1 \leq i \leq n$, all variables that appear in ϕ_i also appear in a and ψ_i. It describes that a causes ϕ_i to hold with grade membership v_i in a successor state to a state in which a is executed and ψ_i holds for all $1 \leq i \leq n$. For each $1 \leq i \leq n$, ψ_i is called a precondition of an action a that corresponds to an effect ϕ_i, ϕ_i is called an effect of a, v_i is the grade membership that ϕ_i holds given that ψ_i holds, where v_i is a non-zero value in $[0, 1]$. For any proposition of the form (3), the set of ground preconditions ψ_i are mutually exclusive and exhaustive.

It will be more convenient for the subsequent results to represent an action a as a set of the form $a = \{a_1, \ldots, a_n\}$, where each a_i corresponds to ϕ_i, v_i, and ψ_i. Therefore, alternatively, for each $1 \leq i \leq n$, proposition (3) can be represented as

$$a_i \textbf{ causes } \phi_i \ : \ v_i \textbf{ if } \psi_i$$

A fuzzy action theory is ground if it does not contain any variables.

3.2 Semantics

A consistent set of ground literals ϕ is a set of literals that does not contain a pair of complementary literals, i.e., l and $\neg l \notin \phi$. If a literal $l \in \phi$, then we say l is true (holds) in ϕ (denoted by $\phi \models l$), and l is false (does not hold) in ϕ if $\neg l$ is in ϕ (denoted by $\phi \models \neg\, l$). If a set of literals σ is contained in ϕ then we say σ is true (holds) in ϕ (denoted by $\phi \models \sigma$), otherwise, σ is false (does not hold) in ϕ (denoted by $\phi \not\models \sigma$). A state s is a complete and consistent set of literals that describes the world at a certain time point.

Definition 1 *Let $\mathbf{FT} = \langle S_0, \mathcal{AD} \rangle$ be a ground fuzzy action theory, $\mathcal{S}$ is the set of all states in $\mathbf{FT}$, s be a state whose grade membership $\mu_S(s)$ is v, a_i **causes** ϕ_i : v_i : **if** ψ_i ($1 \leq i \leq n$) be a proposition in $\mathcal{AD}$, and $a = \{a_1, \ldots, a_n\}$ be an action, where each a_i corresponds to ϕ_i, v_i, and ψ_i for $1 \leq i \leq n$. Then, $s' = \Phi(a_i, s)$ is the state resulting from executing a in s, given that a is executable in s, where:*

- *$l \in \Phi(a_i, s)$ and $\neg\ l \notin \Phi(a_i, s)$ if $l \in \phi_i$ and the precondition ψ_i holds in s.*
- *$\neg\ l \in \Phi(a_i, s)$ and $l \notin \Phi(a_i, s)$ if $\neg\ l \in \phi_i$ and the precondition ψ_i holds in s.*
- *Otherwise, $l \in \Phi(a_i, s)$ iff $l \in s$ and $\neg\ l \in \Phi(a_i, s)$ iff $\neg\ l \in s$.*

where the grade membership of the resulting state $s' = \Phi(a_i, s)$ is $\mu_{\mathcal{S}}(s') = \min(\mu_{\mathcal{S}}(s), v_i) = \min(v, v_i)$. We call Φ a fuzzy transition function.

Example 1 *Consider the following fuzzy planning task adapted from [14]. An arm of a robot is grasping a block from a table, where the* pickup *action the robot performs has effects that are imprecisely defined. The arm is able to tightly hold a block (hb) with a grade membership 0.9 after executing the* pickup *action in the state in which the gripper is dry (gd), and the arm tightly cannot hold the block ($\neg hb$), after executing the* pickup *action in the same state, with grade membership value 0.4. On the other hand, when executing the* pickup *action in the state while the gripper is not dry ($\neg\ gd$) causes the block to be tightly held (hb) with grade of membership equal to 0.3 and tightly not held ($\neg\ hb$) with grade membership 0.6. We assume the initial grade membership distribution of the initial states of the world is given by the grade membership function $\mu_{\mathcal{S}}$ such that $\mu_{\mathcal{S}}(s_1) = 0.9$ and $\mu_{\mathcal{S}}(s_2) = 0.5$, where $s_1 = \{gd, \neg hb\}$ and $s_2 = \{\neg gd, \neg hb\}$, with the understanding that the grade membership of the other states of the world is 0. This fuzzy planning domain can be represented in the fuzzy action language $\mathcal{A}_F$ as the fuzzy action theory $\mathbf{FT} = \langle S_0, \mathcal{AD} \rangle$ where*

$$S_0 = \textbf{initially} \left\{ \begin{array}{lcl} \{gd, \neg hb\} & : & 0.9 \\ \{\neg gd, \neg hb\} & : & 0.5 \end{array} \right.$$

and $\mathcal{AD}$ consists of:

$$\textbf{executable } pickup \textbf{ if } \emptyset$$

$$pickup \textbf{ causes} \left\{ \begin{array}{lcllll} \{hb\} & : & 0.9 & \textbf{if} & \{gd\} \\ \{\neg hb\} & : & 0.4 & \textbf{if} & \{gd\} \\ \{hb\} & : & 0.3 & \textbf{if} & \{\neg\, gd\} \\ \{\neg\, hb\} & : & 0.6 & \textbf{if} & \{\neg\, gd\} \end{array} \right.$$

The $pickup$ action can be represented as the set $pickup = \{pickup_1, pickup_2, pickup_3, pickup_4\}$, where

$$\begin{array}{llllll} pickup_1 & \textbf{causes} & \{hb\} & : & 0.9 & \textbf{if } \{gd\} \\ pickup_2 & \textbf{causes} & \{\neg hb\} & : & 0.4 & \textbf{if } \{gd\} \\ pickup_3 & \textbf{causes} & \{hb\} & : & 0.3 & \textbf{if } \{\neg\, gd\} \\ pickup_4 & \textbf{causes} & \{\neg\, hb\} & : & 0.6 & \textbf{if } \{\neg\, gd\} \end{array}$$

The grade membership distribution resulting from executing the $pickup$ action in the initial states s_1 and s_2 is given by:

- $s_1^{'} = \{gd, hb\} = \Phi(pickup_1, \{gd, \neg hb\})$, where $\mu_S(s_1^{'}) = \min(\mu_S(s_1), 0.9) = \min(0.9, 0.9) = 0.9$.
- $s_2^{'} = \{gd, \neg hb\} = \Phi(pickup_2, \{gd, \neg hb\})$, where $\mu_S(s_2^{'}) = \min(\mu_S(s_1), 0.4) = \min(0.9, 0.4) = 0.4$.
- $s_3^{'} = \{\neg gd, hb\} = \Phi(pickup_3, \{\neg gd, \neg hb\})$, where $\mu_S(s_3^{'}) = \min(\mu_S(s_2), 0.3) = \min(0.5, 0.3) = 0.3$.
- $s_4^{'} = \{\neg gd, \neg hb\} = \Phi(pickup_4, \{\neg gd, \neg hb\})$, where $\mu_S(s_4^{'}) = \min(\mu_S(s_2), 0.6) = \min(0.5, 0.6) = 0.5$.

Definition 2 (Fuzzy Plan) *A sequence of actions $\langle a_0, a_1, \ldots, a_{n-1} \rangle$ is called a fuzzy plan, where each $(0 \leq i \leq n-1)$ a_i is an action with fuzzy effects.*

The grade membership that a fuzzy plan P satisfies a conjunctive fluent formula $\mathcal{G}$ after executing P in a given state s is given by the following definition.

Definition 3 *Let $\mathbf{FT}$ be a ground fuzzy action theory, s, s' be states, s_I be a variable ranging over the possible initial states, $\mathcal{G}$ be a conjunctive fluent formula, and $\langle a_0, a_1, \ldots, a_{n-1} \rangle$ be a fuzzy plan. Then the grade membership that $\mathcal{G}$ is true in a state s' that results after executing $\langle a_0, a_1, \ldots, a_{n-1} \rangle$ in the possible initial states s_I is given by*

$$\mu'_{\mathcal{P}}(s' | s_I, \langle a_0, a_1, \ldots, a_{n-1} \rangle) =$$

$$\max_{s''}(\min_{s_I = s}(\mu'_{\mathcal{P}}(s'' | s, a_0),\ \mu'_{\mathcal{P}}(s' | s'', \langle a_1, \ldots, a_{n-1} \rangle)))$$

where $\mathcal{P}$ is the set of all plans in $\mathbf{FT}$, and $\mu'_{\mathcal{P}}(s'' | s, a_0) = \mu_{\mathcal{S}}(s'') = \min(\mu_S(s), v)$ such that $s'' = \Phi(a_0, s)$. In general, for any action $a \in \mathcal{A}$, $\mu'_{\mathcal{P}}(s'' | s, a) = \mu_{\mathcal{S}}(s'') = \min(\mu_S(s), v)$ such that $s'' = \Phi(a, s)$.

ISBN: 978-989-95079-6-8

Definition 4 *A fuzzy planning problem is a 4-tuple* $\mathbf{FP} = \langle S_0, \mathcal{AD}, \mathcal{G}, \mathcal{T} \rangle$*, where* S_0 *is a fuzzy set in the set of all states* S *that represents the initial agent knowledge about the world states at the time of execution (the initial grade membership distribution over states),* $\mathcal{AD}$ *is a fuzzy action description,* $\mathcal{G}$ *is conjunctive fluent formula represents the goal to be satisfied, and* $0 \leq \mathcal{T} \leq 1$ *is the fuzzy threshold for the goal* $\mathcal{G}$ *to be achieved. We say* $\langle a_0, \ldots, a_{n-1} \rangle$ *is a fuzzy plan for* $\mathbf{FP}$ *iff each* a_i *appears in* $\mathcal{AD}$ *and* $\mu'_{\mathcal{P}}(s'|s_I, \langle a_0, a_1, \ldots, a_{n-1} \rangle) \geq \mathcal{T}$*, where* s_I *is a variable ranging over the possible initial states, and* $\mathcal{G}$ *is true in* s'*.*

4 Fuzzy Planning Using Answer Set Programming

This section presents a translation from a fuzzy planning problem $\mathbf{FP} = \langle S_0, \mathcal{AD}, \mathcal{G}, \mathcal{T} \rangle$ into a normal logic program with answer set semantics, $\Pi_{\mathbf{FP}}$, where the rules in $\Pi_{\mathbf{FP}}$ encode (1) the initial grade membership distribution S_0, (2) the fuzzy transition function Φ, (3) the fuzzy action description $\mathcal{AD}$, (4) and the goal $\mathcal{G}$. The answer sets of $\Pi_{\mathbf{FP}}$ correspond to valid trajectories in $\mathbf{FP}$. The normal logic program translation of a fuzzy planning problem $\mathbf{FP}$ is mainly adapted from [25]. We assume that the length of the fuzzy plan that we are looking for is known. We use the predicates $holds(L, T)$ to represent the fact that a literal L holds at time moment T and $occ(AC, T)$ to describe that an action AC executes at time moment T. We use lower case letters to represent constants and upper case letters to represent variables.

Let $\Pi_{\mathbf{FP}}$ be the normal logic program translation of a fuzzy planning problem $\mathbf{FP} = \langle S_0, \mathcal{AD}, \mathcal{G}, \mathcal{T} \rangle$, where $\Pi_{\mathbf{FP}}$ is the set of rules described as follows. In addition, given p is a predicate and $\psi = \{l_1, \ldots, l_n\}$, we use $p(\psi)$ to denote $p(l_1), \ldots, p(l_n)$.

- For each action $a = \{a_1, \ldots, a_n\} \in \mathcal{A}$, we add to $\Pi_{\mathbf{FP}}$ the set of facts

$$action(a_i) \leftarrow \tag{4}$$

for each $1 \leq i \leq n$. States of the world are described by literals that are encoded in $\Pi_{\mathbf{FP}}$ by the rules

$$literal(A) \leftarrow atom(A) \tag{5}$$
$$literal(\neg A) \leftarrow atom(A) \tag{6}$$

where $atom(A)$ is a set of facts that describe the properties of the world. To present that A and $\neg A$ are contrary literals, the following rules are added to $\Pi_{\mathbf{FP}}$.

$$contrary(A, \neg A) \leftarrow atom(A) \tag{7}$$
$$contrary(\neg A, A) \leftarrow atom(A) \tag{8}$$

- The initial grade membership distribution **initially** ψ_i : v_i for $1 \leq i \leq n$ is represented in $\Pi_{\mathbf{FP}}$ as follows. Consider that $s_1, s_2, \ldots, s_n$ form the set of possible initial states, where for each $1 \leq i \leq n$, $s_i = \{l_1^i, \ldots, l_m^i\}$, and the grade membership of s_i is $\mu_{\mathcal{S}}(s_i) = v_i$. Moreover, let $s = s_1 \cup s_2 \cup \ldots \cup s_n$, $s' = s_1 \cap s_2 \cap \ldots \cap s_n$, $\widehat{s} = s - s'$, and $s'' = \{\, l \,|\, l \in \widehat{s} \vee \neg l \in \widehat{s}\}$. To generate the set of all possible initial states, the following set of rules are added to $\Pi_{\mathbf{FP}}$. For each literal $l \in s'$, the fact

$$holds(l, 0) \leftarrow \tag{9}$$

is in $\Pi_{\mathbf{FP}}$. This fact presents that the literal l holds at time moment 0. This set of facts specifies the set of literals that hold in every possible initial state. Moreover, for each literal $l \in s''$, we add to $\Pi_{\mathbf{FP}}$ the rules

$$holds(l, 0) \leftarrow not\ holds(\neg l, 0) \tag{10}$$
$$holds(\neg l, 0) \leftarrow not\ holds(l, 0) \tag{11}$$

The above rules say that the literal l (similarly $\neg l$) holds at time moment 0, if $\neg l$ (similarly l) does not hold at the time moment 0.

- Each executability condition proposition of an action $a = \{a_1, \ldots, a_n\}$ of the form (2) is encoded in $\Pi_{\mathbf{FP}}$ for each $1 \leq i \leq n$ as

$$exec(a_i, T) \leftarrow holds(\psi, T) \tag{12}$$

- For each proposition of the form a_i **causes** ϕ_i : v_i **if** ψ_i ($1 \leq i \leq n$), in $\mathcal{AD}$, we proceed as follows. Let $\phi_i = \{l_i^1, \ldots, l_i^m\}$. Then, $\forall$ $(1 \leq i \leq n)$, we have for each $1 \leq j \leq m$,

$$holds(l_i^j, T+1) \leftarrow occ(a_i, T), exec(a_i, T), holds(\psi_i, T) \tag{13}$$

belongs to $\Pi_{\mathbf{FP}}$. This rule states that if the action a occurs at time moment T and the precondition ψ_i holds at the same time moment, then the literal l_i^j holds at the time moment $T+1$.

- The frame axioms are presented in $\Pi_{\mathbf{FP}}$ as below. For any literal L we have the rule

$$holds(L, T+1) \leftarrow holds(L, T), not\ holds(L', T+1), contrary(L, L') \tag{14}$$

in $\Pi_{\mathbf{FP}}$. The above rule states that L holds at the time moment $T+1$ if it holds at the time moment T and its contrary does not hold at the time moment $T+1$.

- To encode the fact that a literal A and its negation $\neg A$ cannot hold at the same time, we add the following rule to $\Pi_{\mathbf{FP}}$

$$\leftarrow holds(A, T), holds(\neg A, T) \tag{15}$$

- Action generation rules are described by

$$occ(AC^i, T) \leftarrow action(AC^i), not\ abocc(AC^i, T) \tag{16}$$
$$abocc(AC^i, T) \leftarrow action(AC^i), action(AC^j), occ(AC^j, T), AC^i \neq AC^j \tag{17}$$

The above rules generate action occurrences once at a time, where AC^i and AC^j are variables representing actions.

- Let $\mathcal{G} = g_1 \wedge \ldots \wedge g_m$ be a goal expression, then $\mathcal{G}$ is encoded in $\Pi_{\mathbf{FP}}$ as

$$goal \leftarrow holds(g_1, T), \ldots, holds(g_m, T) \tag{18}$$

ISBN: 978-989-95079-6-8

Example 2 *The normal logic program translation, $\Pi_{\mathbf{FP}}$, of the fuzzy planning problem $\mathbf{FP} = \langle S_0, \mathcal{AD}, \mathcal{G}, \mathcal{T} \rangle$, described in Example 1 proceeds as follows, where S_0 and $\mathcal{AD}$ are as presented in Example 1, $\mathcal{G} = \{hb\}$, and $\mathcal{T} \in [0,1]$. In addition to the rules (5), (6), (7), (8), (14), (15), (16), and (17), $\Pi_{\mathbf{FP}}$ contains the rules*

$$\begin{aligned} action(pickup_1) &\leftarrow \\ action(pickup_2) &\leftarrow \\ action(pickup_3) &\leftarrow \\ action(pickup_4) &\leftarrow \end{aligned}$$

where pickup is in $\mathcal{A}$. Properties of the world are described by the atoms gd (gripper dry) and hb (holding block), which are encoded in $\Pi_{\mathbf{FP}}$ by the rules

$$\begin{aligned} atom(gd) &\leftarrow \\ atom(hb) &\leftarrow \end{aligned}$$

Executability conditions of action pickup are encoded in $\Pi_{\mathbf{FP}}$ by the rules

$$\begin{aligned} exec(pickup_1, t) &\leftarrow \\ exec(pickup_2, t) &\leftarrow \\ exec(pickup_3, t) &\leftarrow \\ exec(pickup_4, t) &\leftarrow \end{aligned}$$

where $0 \leq t \leq n$. The set of possible initial states are encoded in $\Pi_{\mathbf{FP}}$ by the rules:

$$\begin{aligned} holds(\neg hb, 0) &\leftarrow \\ holds(gd, 0) &\leftarrow not\ holds(\neg gd, 0) \\ holds(\neg gd, 0) &\leftarrow not\ holds(gd, 0) \end{aligned}$$

Effects of the pickup action are encoded in $\Pi_{\mathbf{FP}}$ by the rules

$$\begin{aligned} holds(hb, T+1) &\leftarrow occ(pickup_1, T), \\ &\quad exec(pickup_1, T), holds(gd, T) \\ holds(\neg hb, T+1) &\leftarrow occ(pickup_2, T), \\ &\quad exec(pickup_2, T), holds(gd, T) \\ holds(hb, T+1) &\leftarrow occ(pickup_3, T), \\ &\quad exec(pickup_3, T), holds(\neg gd, T) \\ holds(\neg hb, T+1) &\leftarrow occ(pickup_4, T), \\ &\quad exec(pickup_4, T), holds(\neg gd, T) \end{aligned}$$

The goal is encoded in $\Pi_{\mathbf{FP}}$ by the rule

$$goal \leftarrow holds(hb, T)$$

5 Correctness

In this section we prove the correctness of our translation. We show that the answer sets of the normal logic program translation, $\Pi_{\mathbf{FP}}$, of a fuzzy planning problem, $\mathbf{FP}$, correspond to trajectories in $\mathbf{FP}$. Consider that the domain of T is $\{0, \ldots, n\}$. Assume that Φ is the fuzzy transition function associated with $\mathbf{FP}$, s_0 is a possible initial state, and $a_0, \ldots, a_{n-1}$ be a set of actions in $\mathcal{A}$. Any action a_i can be represented as a set where $a_i = \{a_{1_i}, \ldots, a_{m_i}\}$. Therefore, a trajectory in $\mathbf{FP}$ is $s_0\ a_{j_0}\ s_1 \ldots a_{j_{n-1}}\ s_n$ for $(1 \leq j \leq m)$ and $(0 \leq i \leq n)$, such that $\forall (0 \leq i \leq n)$, s_i is a state, a_i is an action, $a_{j_i} \in a_i = \{a_{1_i}, \ldots, a_{m_i}\}$, and $s_i = \Phi(a_{j_{i-1}}, s_{i-1})$.

Theorem 1 *Let $\mathbf{FP} = \langle S_0, \mathcal{AD}, \mathcal{G}, \mathcal{T} \rangle$ be a fuzzy planning problem, P be a fuzzy plan in $\mathbf{FP}$, and T_P be the set of all trajectories in P. Then, $s_0\ a_{j_0}\ s_1 \ldots a_{j_{n-1}}\ s_n$ is a trajectory in T_P iff $occ(a_{j_0}, 0), \ldots, occ(a_{j_{n-1}}, n-1)$ is true in an answer set of $\Pi_{\mathbf{FP}}$.*

Theorem 1 presents that any fuzzy planning problem, $\mathbf{FP}$, can be translated into a normal logic program with answer set semantics, $\Pi_{\mathbf{FP}}$, such that a trajectory in $\mathbf{FP}$ is equivalent to an answer set of $\Pi_{\mathbf{FP}}$. Theorem 1 shows that normal logic programs with answer set semantics can be used to find fuzzy plans for fuzzy planning problems in two steps. The first step is to translate a fuzzy planning problem, $\mathbf{FP}$, into a normal logic program whose answer sets correspond to valid trajectories in $\mathbf{FP}$. From the answer sets of the normal logic program translation of $\mathbf{FP}$, the set of trajectories T_P that correspond to a fuzzy plan P in $\mathbf{FP}$ is determined. The second step is to calculate the grade membership of the fuzzy plan P using the formula

$$\max_{s_0\ a_{j_0}\ s_1 \ldots a_{j_{n-1}}\ s_n \in T_P} \left(\min_{0 \leq i \leq n-1} (\mu_{\mathcal{S}}(s_i), \mu_{\mathcal{S}}(s_{i+1})) \right)$$

Furthermore, we show that any fuzzy planning problem can be encoded as a SAT formula. Hence, state-of-the-art SAT solvers can be used to find fuzzy plans for fuzzy planning problems. Any normal logic program, Π, can be translated into a SAT formula, $\mathcal{S}$, where the models of $\mathcal{S}$ are equivalent to the answer sets of Π [19]. Therefore, the normal logic program translation of a fuzzy planning problem $\mathbf{FP}$ can be encoded into an equivalent SAT formula, where the models of $\mathcal{S}$ correspond to valid trajectories in $\mathbf{FP}$.

Theorem 2 *Let $\mathbf{FP}$ be a fuzzy planning problem and $\Pi_{\mathbf{FP}}$ be the normal logic program encoding of $\mathbf{FP}$. Then, the models of the SAT encoding of $\Pi_{\mathbf{FP}}$ are equivalent to valid trajectories in $\mathbf{FP}$.*

6 Conclusions and Related Work

We described a novel action language called fuzzy action language, $\mathcal{A}_F$, that allows the representation and reasoning about actions with fuzzy effects. In addition we introduced the notions of fuzzy planning and fuzzy plan in the fuzzy action language $\mathcal{A}_F$. Furthermore, we described a fuzzy planner based on the fuzzy action language A_F that is developed by translating a fuzzy planning problem, $\mathbf{FP}$, in $\mathcal{A}_F$ into a normal logic program with answer set semantics, $\Pi_{\mathbf{FP}}$, where trajectories in $\mathbf{FP}$ are equivalent to the answer sets of $\Pi_{\mathbf{FP}}$. In addition, we formally proved the correctness of our planner. Furthermore, we showed that a fuzzy planning problem can be encoded as a SAT problem.

The literature is rich with action languages that are capable of representing and reasoning about actions in the presence of probabilistic uncertainty, which include [2, 3, 5, 12, 14, 21, 22]. In [22], a probabilistic action language $\mathcal{P}$ is described that allows the representation of imperfect sensing actions (with probabilistic outcomes), non-sensing actions with probabilistic effects, the initial probability distribution over the possible initial states, and the indirect effects of actions. The action language $\mathcal{E}^+$ [12] allows sensing actions under the assumption that the agent's sensors are perfect, actions with probabilistic effects, and actions with non-deterministic effects. In

ISBN: 978-989-95079-6-8

addition, the semantics of $\mathcal{E}^+$ is based on description logic. Other high level probabilistic action description languages are described in [2, 5]. These languages are similar to $\mathcal{E}^+$ in the sense that they represent and reason about actions with probabilistic effects, except that they do not allow actions with non-deterministic effects or sensing actions. In [21], a high level action language called $\mathcal{A}_{MD}$ is presented that allows the factored representation and reasoning about Markov Decision Processes for reinforcement learning. In addition to the fact that $\mathcal{A}_F$ is a high level language, the major difference between $\mathcal{A}_F$ and these languages is that $\mathcal{A}_F$ represent and reason about actions under the presence of fuzzy uncertainty or imprecision. This is achieved in $\mathcal{A}_F$ by allowing description of actions with fuzzy effects, the executability conditions of actions, and the initial grade membership distribution of the initial states.

References

[1] C. Baral. *Knowledge representation, reasoning, and declarative problem solving*. Cambridge University Press, 2003.

[2] C. Baral, N. Tran, L. C. Tuan. *Reasoning about actions in a probabilistic setting*. In *AAAI*, 2002.

[3] C. Boutilier, T. Dean, and S. Hanks. Decision-theoretic planning: structural assumptions and computational leverage. *Journal of AI Research*, 11:1–94, 1999.

[4] C. da Costa Pereira and A. Tettamanzi. Reasoning about actions with imprecise and incomplete state descriptions. *Fuzzy Sets and Systems*, 160(10):13831401, 2009.

[5] T. Eiter and T. Lukasiewicz. Probabilistic reasoning about actions in nonmonotonic causal theories. In *19th Conference on Uncertainty in Artificial Intelligence*, 2003.

[6] T. Eiter et al. Declarative problem solving in dlv. In *Logic Based Artificial Intelligence*, 2000.

[7] M. Gelfond and V. Lifschitz. The stable model semantics for logic programming. *ICSLP*, 1988, MIT Press.

[8] M. Gelfond and V. Lifschitz. Classical negation in logic programs and disjunctive databases. *New Generation Computing*, 9(3-4):363-385, 1991.

[9] M. Gelfond and V. Lifschitz Representing action and change by logic programs. *Journal of Logic Programming*, 17:301–321, 1993.

[10] M. Gelfond and V. Lifschitz. Action languages. In *Electronic Transactions on AI*, 3(16), 1998.

[11] E. Giunchiglia and V. Lifschitz. An action language based on causal explanation: preliminary report. In *AAAI*, 1998.

[12] L. Iocchi, T. Lukasiewicz, D. Nardi, and R. Rosati. Reasoning about actions with sensing under qualitative and probabilistic uncertainty. In *16th European Conference on Artificial Intelligence*, 2004.

[13] H. Kautz and B. Selman. Pushing the envelope: planning, propositional logic, and stochastic search. In *13th National Conference on Artificial Intelligence*, 1996.

[14] N. Kushmerick, S. Hanks, and D. Weld. An algorithm for probabilistic planning. *Artificial Intelligence*, 76(1-2):239-286, 1995.

[15] J. Lee and V. Lifschitz. Describing additive fluents in action language C+. In *International Joint Conference in AI*, 2003.

[16] V. Lifschitz. Answer set planning. In *ICLP*, 1999.

[17] V. Lifschitz. Two components of an action language. *Annals of Mathematics and Artificial Intelligence*, 21:305–320, 1997.

[18] V. Lifschitz and W. Ren. A modular action description language. In *AAAI*, 2006.

[19] F. Lin and Y. Zhao. ASSAT: Computing answer sets of a logic program by SAT solvers. *Artificial Intelligence*, 157(1-2):115-137, 2004.

[20] I. Niemela and P. Simons. Efficient implementation of the well-founded and stable model semantics. *In Joint International Conference and Symposium on Logic Programming*, 289-303, 1996.

[21] E. Saad. A logical framework to reinforcement learning using hybrid probabilistic logic programs. In *Second International Conference on Scalable Uncertainty Management*, 2008.

[22] E. Saad. Probabilistic planning with imperfect sensing actions using hybrid probabilistic logic programs. In *Ninth International Workshop on Computational Logic in Multi-Agent Systems (CLIMA'08)*, 2008.

[23] A. Saffiotti, K. Konolige, and E. Ruspini. A Multivalued logic approach to integrating planning and control. *Artificial Intelligence*, 76(1-2): 481–526, 1995.

[24] T. Son and C. Baral. Formalizing sensing actions - a transition function based approach. *Artificial Intelligence*, 125:(1-2): 19 - 91, 2001.

[25] T. Son, C. Baral, T. Nam, and S. McIlraith. Domain-dependent knowledge in answer set planning. *ACM Transactions on Computational Logic*, 7(4):613–657, 2006.

[26] V. S. Subrahmanian and C. Zaniolo. Relating stable models and AI planning domains. *In International Conference of Logic Programming*, 233-247, 1995.

[27] L. Zadeh. Fuzzy Sets. *Information and Control*, 8(3): 338 –353, 1965.

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

ISBN: 978-989-95079-6-8

Interval-valued restricted equivalence functions applied on Clustering Techniques

Aranzazu Jurio[1] Miguel Pagola[1] Daniel Paternain[1] Carlos Lopez-Molina[1] Pedro Melo-Pinto[2]

1.Departamento de Automática y Computación, Universidad Pública de Navarra
Pamplona, Spain
2.Department of Engineering, CITAB, Universidade de Trás-os-Montes e Alto Douro
Vila Real, Portugal
Email: aranzazu.jurio@unavarra.es, miguel.pagola@unavarra.es, pmelo@utad.pt

Abstract— *In this work we use interval-valued fuzzy sets in the Fuzzy C-Means algorithm for image segmentation. We introduce interval-valued restricted equivalence functions as a way of measuring the equivalence between the intervals associated to different pixels. We propose two construction methods for those new functions. We also prove experimentally that with these new concepts, the algorithm provides better results in ultrasound image segmentation that than those not using intervals.*

Keywords— Decomposability, Fuzzy C-Means, Interval-Valued Fuzzy Sets, Restricted Equivalence Functions, Ultrasound image.

1 Introduction

Ultrasound images are used to evaluate breast anomalies during the diagnosis in breast exams. When any breast anomaly is studied to determine whether it is benign or not, the radiologist studies its size, area and other morphological characteristics [12]. In this sense, the development of algorithms that automatically separates the lesion and the rest of the image is very important for a quick and correct diagnosis.

Ultrasound images segmentation is very limited by its bad quality. Speckle noise, shades and attenuation make the segmentation problem a very difficult topic [10]. There exist some methods used for this purpose; a comprehensive survey can be found in [9].

A very common method for segmenting images is clustering. The purpose of this work is to adapt the well known algorithm Fuzzy C-Means (FCM) [1, 2] to ultrasound image segmentation. This algorithm tries to classify a set of objects, in this case image pixels, in a determined number of clusters. That number must be chosen a priori. The main idea is to find as many centers as clusters must be, trying to minimize the distance between each pixel and the cluster centroids.

Due to the complexity of ultrasound images, there exists a lot of uncertainty within them. This uncertainty is also present in the images obtained after processing the ultrasound images. We are going to use interval-valued fuzzy sets [13] in the FCM algorithm to deal with this uncertainty.

As already mentioned, our proposed algorithm is based on the FCM. The new contributions we propose are: the use of interval-valued fuzzy sets and the search of the greatest interval equivalence between the image pixels and the centroids of the clusters. To find the greatest equivalence between intervals we introduce the concept of interval-valued restricted equivalence function.

This paper is organized in the following way: first of all we present some preliminary concepts. In Section 3 we define the concept of interval-valued restricted equivalence functions. In Section 4, we explain the way to apply these functions to ultrasound images segmentation and we show some experimental results. Finally some conclusions and future researchs are shown.

2 Preliminaries

We know that in fuzzy set theory a function $n : [0,1] \to [0,1]$ with $n(0) = 1$, $n(1) = 0$ that is strictly decreasing and continuous is called strict negation. If, in addition, n is involutive, then it is said that it is a strong negation.

We denote by $L([0,1])$ the set of all closed subintervals of the closed interval $[0,1]$; that is,

$$L([0,1]) = \{\mathbf{x} = [\underline{x}, \overline{x}] \mid (\underline{x}, \overline{x}) \in [0,1]^2 \text{ and } \underline{x} \leq \overline{x}\}$$

$L([0,1])$ is a partially ordered set with respect to the relation $\leq_L$ defined in the following way: given $\mathbf{x}, \mathbf{y} \in L([0,1])$,

$$\mathbf{x} \leq_L \mathbf{y} \text{ if and only if } \underline{x} \leq \underline{y} \text{ and } \overline{x} \leq \overline{y};$$

$$\mathbf{x} =_L \mathbf{y} \text{ if and only if } \underline{x} = \underline{y} \text{ and } \overline{x} = \overline{y};$$

The relation above is transitive, antisymmetric and it expresses the fact that $\mathbf{x}$ strongly links to $\mathbf{y}$, so that $(L([0,1]), \leq_L)$ is a complete lattice, where the smallest element is $0_L = [0,0]$, and the largest is $1_L = [1,1]$.

Definition 1 *An interval-valued fuzzy set A on the universe $U \neq \emptyset$ is a mapping $A : U \to L([0,1])$.*

Definition 2 *An IV negation is a function $N : L([0,1]) \to L([0,1])$ that is decreasing (with respect to $\leq_L$) and such that $N(1_L) = 0_L$ and $N(0_L) = 1_L$. If for all $\boldsymbol{x} \in L([0,1])$, $N(N(\boldsymbol{x})) = \boldsymbol{x}$, it is said that N is involutive.*

Theorem 1 *[7, 3] A function $N : L([0,1]) \to L([0,1])$ is an involutive IV negation if and only if there exists an involutive negation n such that*

$$N(\boldsymbol{x}) = [n(\overline{x}), n(\underline{x})].$$

Definition 3 *[4] A function $REF : [0,1]^2 \to [0,1]$ is called a restricted equivalence function, if it satisfies the following conditions:*

(1) $REF(x,y) = REF(y,x)$ for all $x, y \in [0,1]$;

(2) $REF(x,y) = 1$ if and only if $x = y$;

ISBN: 978-989-95079-6-8

(3) $REF(x,y) = 0$ *if and only if* $x = 1$ *and* $y = 0$ *or* $x = 0$ *and* $y = 1$*;*

(4) $REF(x,y) = REF(n(x), n(y))$ *for all* $x, y \in [0,1]$*,* n *being a strong negation;*

(5) For all $x, y, z \in [0,1]$*, if* $x \leq y \leq z$*, then* $REF(x,y) \geq REF(x,z)$ *and* $REF(y,z) \geq REF(x,z)$*.*

Definition 4 *[8] A function* $F : L([0,1])^2 \to L([0,1])$ *is said decomposable if there exist two functions* $F_1, F_2 : [0,1]^2 \to [0,1]$ *such that for all* $[\underline{x}, \overline{x}], [\underline{y}, \overline{y}] \in L([0,1])$ *they satisfy that*

$$F([\underline{x}, \overline{x}], [\underline{y}, \overline{y}]) = [F_1(\underline{x}, \underline{y}), F_2(\overline{x}, \overline{y})]$$

Definition 5 *[6] A binary aggregation function is defined as a function* $Ag : [0,1]^2 \to [0,1]$ *such that:*

(1) $Ag(x_1, x_2) \leq Ag(y_1, y_2)$ *whenever* $x_1 \leq y_1$ *and* $x_2 \leq y_2$*;*

(2) $Ag(0,0) = 0$ *and* $Ag(1,1) = 1$*.*

This is an idempotent aggregation function if

(3) $Ag(x,x) = x$ *for all* $x \in [0,1]$

3 Interval Valued Restricted Equivalence Functions

Definition 6 *An Interval Valued Restricted Equivalence Function (*REF_{IV}*) is a function*

$$REF_{IV} : L([0,1]) \times L([0,1]) \to L([0,1])$$

such that:

(1) $REF_{IV}(\boldsymbol{x}, \boldsymbol{y}) = REF_{IV}(\boldsymbol{y}, \boldsymbol{x})$ *for all* $\boldsymbol{x}, \boldsymbol{y} \in L([0,1])$*;*

(2) $REF_{IV}(\boldsymbol{x}, \boldsymbol{y}) = 1_L$ *if and only if* $\boldsymbol{x} = \boldsymbol{y}$*;*

(3) $REF_{IV}(\boldsymbol{x}, \boldsymbol{y}) = 0_L$ *if and only if* $\boldsymbol{x} = 1_L$ *and* $\boldsymbol{y} = 0_L$ *or* $\boldsymbol{x} = 0_L$ *and* $\boldsymbol{y} = 1_L$*;*

(4) $REF_{IV}(\boldsymbol{x}, \boldsymbol{y}) = REF_{IV}(N(\boldsymbol{x}), N(\boldsymbol{y}))$ *being N an involutive IV negation;*

(5) For all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in L([0,1])$*, if* $\boldsymbol{x} \leq_L \boldsymbol{y} \leq_L \boldsymbol{z}$*, then* $REF_{IV}(\boldsymbol{x}, \boldsymbol{y}) \geq_L REF_{IV}(\boldsymbol{x}, \boldsymbol{z})$ *and* $REF_{IV}(\boldsymbol{y}, \boldsymbol{z}) \geq_L REF_{IV}(\boldsymbol{x}, \boldsymbol{z})$*.*

Theorem 2 *There is not any decomposable* REF_{IV}*.*

Proof. We suppose there exists a decomposable REF_{IV}. This means that there exist two functions $F_1, F_2 : [0,1] \times [0,1] \to [0,1]$ such that:

$$REF_{IV}([\underline{x}, \overline{x}], [\underline{y}, \overline{y}] = [F_1(\underline{x}, \underline{y}), F_2(\overline{x}, \overline{y})]$$

in such form that REF_{IV} fulfills the five properties demanded in Definition 6. It means:

(1) As $REF_{IV}(\mathbf{x}, \mathbf{y}) = REF_{IV}(\mathbf{y}, \mathbf{x})$, then F_1 and F_2 must satisfy that $F_1(a,b) = F_1(b,a)$ and $F_2(a,b) = F_2(b,a)$ for all $a, b \in [0,1]$.

(2) As $REF_{IV}([\underline{x}, \overline{x}], [\underline{y}, \overline{y}]) = 1_L$ if and only if $\underline{x} = \underline{y}$ and $\overline{x} = \overline{y}$, then $F_1(\underline{x}, \underline{x}) = 1$ and $F_2(\overline{x}, \overline{x}) = 1$, so for all $x \in [0,1]$ $F_1(x,x) = F_2(x,x) = 1$.

(3) As $REF_{IV}([\underline{x}, \overline{x}], [\underline{y}, \overline{y}]) = 0_L$ if and only if $\mathbf{x} = 0$ and $\mathbf{y} = 1$ or vice versa, then $F_1(0,1) = 0$ and $F_2(0,1) = 0$.

(4) As $REF_{IV}([\underline{x}, \overline{x}], [\underline{y}, \overline{y}]) = REF_{IV}([n(\overline{x}), n(\underline{x})], [n(\overline{y}), n(\underline{y})]$ then $[F_1(\underline{x}, \underline{y}), F_2(\overline{x}, \overline{y})] = [F_1(n(\overline{x}), n(\overline{y})), F_2(n(\underline{x}), n(\underline{y}))]$. Separating each term, $F_1(\underline{x}, \underline{y}) = F_1(n(\overline{x}), n(\overline{y}))$ and $F_2(\overline{x}, \overline{y}) = F_2(n(\underline{x}), n(\underline{y}))$. In particular, if $\underline{x} = 1$ and $\underline{y} = 0$, and $\overline{x} = \overline{y} = 1$ then $F_2(\overline{x}, \overline{y}) = F_2(1,1) = 1$ and $F_2(n(\underline{x}), n(\underline{y})) = F_2(0,1) = 0$. $1 \neq 0$, so there is not any decomposable REF_{IV}.

Theorem 3 *Let* REF *be a restricted equivalence function, and* Ag_1 *and* Ag_2 *be two symmetric, idempotent aggregation functions such that:*

(i) $Ag_1 \leq Ag_2$*;*

(ii) $Ag_1(a,b) = 1$ *if and only if* $a = b = 1$*;*

(iii) $Ag_2(a,b) = 0$ *if and only if* $a = b = 0$*.*

Under these conditions, the function

$$REF_{IV} : L([0,1]) \times L([0,1]) \to L([0,1])$$

given by

$$REF_{IV}(\boldsymbol{x}, \boldsymbol{y}) = \begin{array}{l}[Ag_1(REF(\underline{x}, \underline{y}), REF(\overline{x}, \overline{y})), \\ Ag_2(REF(\underline{x}, \underline{y}), REF(\overline{x}, \overline{y}))]\end{array}$$

is an interval-valued restricted equivalence function in the sense of Definition 6.

Proof. By hypothesis, $Ag_1 \leq Ag_2$, so the REF_{IV} is a well defined interval.

(1) REF, Ag_1 and Ag_2 are commutative functions, so REF_{IV} is a commutative function too.

(2) (Sufficiency) If $REF_{IV}(\mathbf{x}, \mathbf{y}) = 1_L$, then $Ag_1(REF(\underline{x}, \underline{y}), REF(\overline{x}, \overline{y})) = 1$. So, by (ii), $REF(\underline{x}, \underline{y}) = REF(\overline{x}, \overline{y}) = 1$. Any REF is 1 if and only if both arguments are equal, so $\underline{x} = \underline{y}$ and $\overline{x} = \overline{y}$. This means $\mathbf{x} = \mathbf{y}$.

(Necessity) If $\mathbf{x} = \mathbf{y}$, it means $\underline{x} = \underline{y}$ and $\overline{x} = \overline{y}$, then $REF_{IV}(\mathbf{x}, \mathbf{x}) = \begin{array}{l}[Ag_1(REF(\underline{x}, \underline{x}), REF(\overline{x}, \overline{x})), \\ Ag_2(REF(\underline{x}, \underline{x}), REF(\overline{x}, \overline{x}))]\end{array}$ $[Ag_1(1,1), Ag_2(1,1)]$ and by the property (2) in Definition 5 this is $[1,1]$.

(3) (Sufficiency) If $REF_{IV}(\mathbf{x}, \mathbf{y}) = 0_L$, then $Ag_2(REF(\underline{x}, \underline{y}), REF(\overline{x}, \overline{y})) = 0$, therefore, by (iii), $REF(\underline{x}, \underline{y}) = REF(\overline{x}, \overline{y}) = 0$. Any REF is 0 if and only if one of the arguments is 0 and the other is 1, so $\underline{x} = 1$ and $\underline{y} = 0$ or vice versa, and $\overline{x} = 1$ and $\overline{y} = 0$ or vice versa. This means $\underline{x} = \overline{x} = 1$ and $\underline{y} = \overline{y} = 0$ or vice versa.

(Necessity) If $\mathbf{x} = 1_L$ and $\mathbf{y} = 0_L$, then $REF_{IV}(\mathbf{x}, \mathbf{y}) = [Ag_1(REF(1,0), REF(1,0)), Ag_2(REF(1,0), REF(1,0))] = [Ag_1(0,0), Ag_2(0,0)]$, which is equal to $[0,0]$ by property (2) in Definition 5.

ISBN: 978-989-95079-6-8

(4) $REF_{IV}(N(\mathbf{x}), N(\mathbf{y})) = [Ag_1(REF(n(\overline{x}), n(\overline{y})), REF(n(\underline{x}), n(\underline{y}))), Ag_2(REF(n(\overline{x}), n(\overline{y})), REF(n(\underline{x}), n(\underline{y})))]$. As $REF(\underline{x}, \underline{y}) = REF(n(\underline{x}), n(\underline{y}))$ and $REF(\overline{x}, \overline{y}) = REF(n(\overline{x}), n(\overline{y}))$, then $REF_{IV}(N(\mathbf{x}), N(\mathbf{y})) = [Ag_1(REF(\overline{x}, \overline{y}), REF(\underline{x}, \underline{y})), Ag_2(REF(\overline{x}, \overline{y}), REF(\underline{x}, \underline{y}))]$. As Ag_1 and Ag_2 are commutative functions, this expression is equal to $[Ag_1(REF(\underline{x}, \underline{y}), REF(\overline{x}, \overline{y})), Ag_2(REF(\underline{x}, \underline{y}), REF(\overline{x}, \overline{y}))] = REF_{IV}(\mathbf{x}, \mathbf{y})$.

(5) If $\mathbf{x} \leq_L \mathbf{y} \leq_L \mathbf{z}$, then $\underline{x} \leq \underline{y} \leq \underline{z}$ and $\overline{x} \leq \overline{y} \leq \overline{z}$. By (5) in Definition 3, $REF(\underline{x}, \underline{y}) \geq REF(\underline{x}, \underline{z})$ and $REF(\overline{x}, \overline{y}) \geq REF(\overline{x}, \overline{z})$. By (1) in Definition 5 $Ag_1(REF(\underline{x}, \underline{z}), REF(\overline{x}, \overline{z})) \leq Ag_1(REF(\underline{x}, \underline{y}), REF(\overline{x}, \overline{y}))$ and $Ag_2(REF(\underline{x}, \underline{z}), REF(\overline{x}, \overline{z})) \leq Ag_2(REF(\underline{x}, \underline{y}), REF(\overline{x}, \overline{y}))$, so $REF_{IV}(\mathbf{x}, \mathbf{z}) \leq REF_{IV}(\mathbf{x}, \mathbf{y})$. The reasoning for $REF_{IV}(\mathbf{x}, \mathbf{z}) \leq REF_{IV}(\mathbf{y}, \mathbf{z})$ is analogous.

Corollary 1 *Given a restricted equivalence function REF, and given T and S any t-norm and any t-conorm,*

$$REF_{IV}(\boldsymbol{x},\boldsymbol{y}) = [T(REF(\underline{x}, \underline{y}), REF(\overline{x}, \overline{y})), S(REF(\underline{x}, \underline{y}), REF(\overline{x}, \overline{y}))]$$

is an interval-valued restricted equivalence function.

Proof. We need to see that any t-norm fulfills the properties demanded to Ag_1 and any t-conorm fulfills the properties demanded to Ag_2.

For the t-norm T it is necessary that $T(a, b) = 1$ if and only if $a = b = 1$. (Sufficiency) If the t-norm is the minimum, if $min(a, b) = 1$, then $a = b = 1$. As the minimum is the greatest of all t-norms, to be 1 the arguments of any t-norm must be greater than or equal to 1. As they are defined in the unit interval, both arguments must be 1.

(Necessity) We know that $T(1, x) = x$ for any $x \in [0, 1]$. If $x = y = 1$, then $T(1, 1) = 1$.

For the t-conorm S it is necessary that $S(a, b) = 0$ if and only if $a = b = 0$. (Sufficiency) If the t-conorm is the maximum, if $max(a, b) = 0$, then $a = b = 0$. As the maximum is the smallest of all t-conorms, to be 0 the arguments of any t-conorm must be smaller than or equal to 0. As they are defined in the unit interval, both arguments must be 0.

(Necessity) We know that $S(x, 0) = x$ for any $x \in [0, 1]$. If $x = y = 0$, then $S(0, 0) = 0$.

Example 1 *If we choose $T = min$, $S = max$ and $REF(x, y) = 1 - |x - y|$, we have that:*

$$REF_{IV}(\boldsymbol{x},\boldsymbol{y}) = [min(1 - |\underline{x} - \underline{y}|, 1 - |\overline{x} - \overline{y}|), max(1 - |\underline{x} - \underline{y}|, 1 - |\overline{x} - \overline{y}|)] =$$

$$[1 - max(|\underline{x} - \underline{y}|, |\overline{x} - \overline{y}|), 1 - min(|\underline{x} - \underline{y}|, |\overline{x} - \overline{y}|)] =$$

$$N([min(|\underline{x} - \underline{y}|, |\overline{x} - \overline{y}|), max(|\underline{x} - \underline{y}|, |\overline{x} - \overline{y}|)].$$

4 Application of $\mathbf{REF_{IV}}$ to clustering algorithm

The algorithm proposed in this paper is a modification of the Fuzzy C-Means (FCM) algorithm [1, 2], which aims to find the most characteristic point of each cluster, considered its centroid. Such a way, the membership degree of every object to each cluster is achieved via minimizing the target function:

$$J_m(U, V) = \sum_{j=1}^{n} \sum_{i=1}^{K} w_{ij}^m \parallel x_j - C_i \parallel^2$$

In the new algorithm we propose, the function to minimize is quite similar. The only difference is the way of calculating the distance between each pixel and the cluster centroids. Instead of using the euclidean distance, we use a negation of an interval restricted equivalence function. Therefore, our objective function to minimize is:

$$J_m(U, V) = \sum_{j=1}^{n} \sum_{i=1}^{K} w_{ij}^m (1 - Ag(REF_{IV}(\mathbf{x_j}, \mathbf{C_i})))$$

The main purpose of the new algorithm is to segment the image into K areas represented by K interval centroids, maximizing the equivalence in each area, using the interval valued restricted equivalence functions, that we have previously defined. As the algorithm is partially supervised, the number of areas must be selected by the user before the execution. Notice that the REF_{IV} is an interval. As we want to maximize the equivalence, we are going to use an aggregation of the bounds of the equivalence interval.

The algorithm starts with the initialization of all the required parameters. First of all it is necessary to build the interval-valued fuzzy set associated with the image. For that, each pixel is assigned an interval depending on the values of its neighborhood. The rest of the parameters to be initialized are the number of areas to be found, the fuzzification degree, the weight matrix, the maximum number of iterations and the finishing threshold.

After the initialization, the algorithm runs in a loop. In this loop the centroid of each cluster and the weight matrix are calculated until the improvement is not significant. To update that weight matrix, we use the interval-valued restricted equivalence functions.

When the loop is finished, the algorithm labels each pixel to show the results. For this purpose, each pixel is associated to the cluster with the biggest weight.

4.1 Algorithm structure

1. Given an image X with n pixels, build its interval-valued fuzzy set $A = \{[\underline{x_i}, \overline{x_i}] | x_i \in X\}$ with $i = 1 \ldots n$.
2. Initialize
   - 2.1 K = number of areas to be segmented.
   - 2.2 T = maximum number of iterations.
   - 2.3 m = fuzzification degree, $1 \leq m \leq \infty$.
   - 2.4 ε = finishing threshold, $\varepsilon > 0$.
   - 2.5 W = weight matrix, $0 \leq w_{ij} \leq 1$ with $i = 1 \ldots n$ and $j = 1 \ldots K$.

ISBN: 978-989-95079-6-8

2.6 t = current iteration = 0.

3. REPEAT

3.1 Increase current iteration

3.2 Calculate the centroids

$$C_j = \frac{\sum_{i=1}^{n} w_{ij}^m [\underline{x_i}, \overline{x_i}]}{\sum_{i=1}^{n} w_{ij}^m}$$

3.3 Update the weight matrix

$$w_{ij} = \left(\sum_{k=1}^{K} \left(\frac{1 - (Ag(REF_{IV}(\mathbf{x_i}, \mathbf{C_j})))}{1 - (Ag(REF_{IV}(\mathbf{x_i}, \mathbf{C_k})))} \right)^{\frac{1}{m-1}} \right)^{-1}$$

where $\mathbf{x_i} = [\underline{x_i}, \overline{x_i}]$, $\mathbf{C_j} = [\underline{C_j}, \overline{C_j}]$, $\mathbf{C_k} = [\underline{C_k}, \overline{C_k}]$ and $Ag(REF_{IV}(\mathbf{x}, \mathbf{y})) = \frac{REF_{IV}(\mathbf{x},\mathbf{y}) + \overline{REF_{IV}(\mathbf{x},\mathbf{y})}}{2}$.

3.4 Calculate the error.

UNTIL $t = T$ or $error \leq \varepsilon$.

4. Label the image.

4.2 Preprocessing

To prove the performance of the new algorithm, we are going to segment some ultrasound images. In this case we work with breast ultrasound images, where the goal is to classify every pixel of the image. It means, decide whether they belong to the lesion or not.

Due to the characteristics of these images, which may contain several similar regions, it is necessary to select the central pixel of the desired zone. Since the user has to contribute with some information, we say that the proposed system is partially supervised.

To improve the results the images are preprocessed. In order to do so, we create two new images for every ultrasound image. The first one is a false edges image while the second one is a enhanced image. An example of these new images can be seen in Fig. 1.



Figure 1: False edges and enhanced images.

4.2.1 Enhanced image

To create the enhanced image we use the algorithm proposed by Sahba et al. [11] for ultrasound images. The main idea is to remove the image noise and to enhance the gray level in the selected area. First of all the noise is removed using the median filter (7x7 or 9x9). After that, every pixel is fuzzified depending on its gray level using a membership function considering the average gray level in the pixel neighborhood and the position of the central pixel selected.

4.2.2 False Edges image

The false edges image represents to what degree each pixel is part of an edge. In [5] a method is presented to extract false edges from t-norms and t-conorms. For each pixel, the first step is to build a submatrix (3x3, 5x5, etc.). Applying any t-norm to the elements of that submatrix, the lower bound of an interval is obtained. Analogously, the upper bound is got applying any t-conorm to the same submatrix. The width of that interval, it means, the upper bound minus the lower bound, is called false edge. If the t-norm and the t-conorm chosen are the minimum and maximum respectively, the false edge is basically the substraction of the biggest element in the submatrix minus the lowest one.

4.3 Experiment

The experiment we show in this section is based on 8 ultrasound images, shown in Fig. 2. They are breast images in which there exist some zones that can be tumors. To create the interval-valued fuzzy set associated to the image, we use a net of 5x5 neighbors for every pixel. In this sense, each pixel is corresponded with an interval value that depends not only on its value but also on its neighbors' values. To construct the lower bound we take the minimum of the values of the mesh, and to calculate the upper bound we take the greatest one.

The REF_{IV} we use is the one shown in example 1:

$$\begin{aligned} REF_{IV}(\mathbf{x}, \mathbf{y}) = \quad & [min(1 - |\underline{x} - \underline{y}|, 1 - |\overline{x} - \overline{y}|), \\ & max(1 - |\underline{x} - \underline{y}|, 1 - |\overline{x} - \overline{y}|)] = \\ N([min(|\underline{x} - \underline{y}|, |\overline{x} - \overline{y}|), & max(|\underline{x} - \underline{y}|, |\overline{x} - \overline{y}|)]. \end{aligned}$$

To calculate the performance of the segmentation method we propose, we use the areas overlapping measure (S_A) between the image we have obtained and the ideal one. The latter is created by an expert radiologist.

$$S_A = \frac{|Ideal \ \& \ Obtained|}{|Ideal \ + \ Obtained|} \tag{1}$$

where $Ideal$ is the binary image segmented by the expert radiologist and $Obtained$ is the binary image got by the studied method.

We are going to compare the results from our algorithm with the ones obtained with the FCM algorithm by comparing their accuracy with respect to the ideal images. In Fig. 2 some results are illustrated. The first column matches the original image, while the second and the third columns match the segmentation using the FCM algorithm and the proposed one respectively. Finally, the fourth column shows the ideal segmentation.

As it can be seen, the results from our algorithm are more similar to the ideal ones than the ones got by the FCM. This improvement can be numerically observed by the areas overlapping percentage between each image and its corresponding ideal one (see table 1, where A, B, ...H refer to the images shown in Fig. 2).

With these results, our proposed algorithm improves the results obtained by the FCM in 8%. This percentage is the difference between the average overlap of the ideal image with the one obtained by the simple FCM and with the one obtained by our method in the images of Fig. 2. We can say this is a very good result, as it means a significant progress towards the correct segmentation of this kind of images.

ISBN: 978-989-95079-6-8

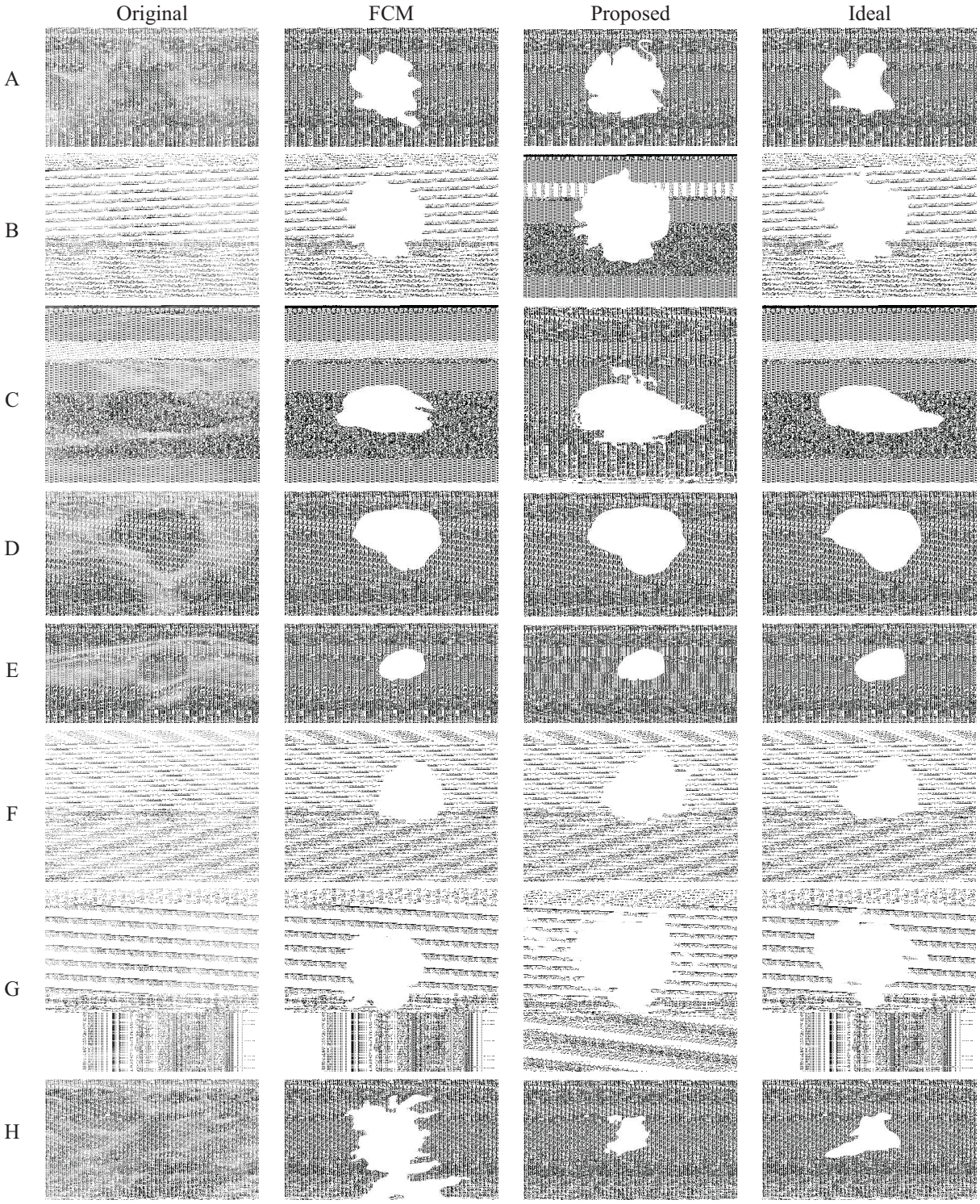


Figure 2: Obtained results.

Table 1: Analytic solutions.

	A	B	C	D
FCM	75.17	76.73	73.28	84.48
Proposed	70.81	89.43	76.37	91.47
	E	F	G	H
FCM	74.11	72.65	57.87	27.55
Proposed	76.77	79.63	71.68	51.33

The use of interval-valued fuzzy sets in this sort of images increases the accuracy of the segmentation. In some cases, it allows us to discover regions that are not easily identified at first sight, as it can be seen in the images we have segmented.

5 Conclusions and future research

In this work we show the advantage of using interval-valued fuzzy sets in the FCM algorithm. For this reason, we have introduced the concept of interval-valued restricted equivalence function and we have presented two different construction

ISBN: 978-989-95079-6-8

methods. The effectiveness of this method has been proved with the experimental results.

As future research we want to work with intervals within the FCM instead of aggregating the bounds of the equivalence interval. Also, we are going to study other processings of ultrasound images to improve the results.

Acknowledgment - This paper has been partially supported by the National Science Foundation of Spain, Reference TIN2007-65981.

References

[1] J.C. Bezdek. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum, NY (1981).

[2] J.C. Bezdek, J. Keller, R. Krisnapuram, N.R. Pal. Fuzzy Models and Algorithms for Pattern Recognition and Image Processing. (Kluwer Academic Publishers) (1999).

[3] H. Bustince, E.Barrenechea, M. Pagola. Generation of Interval-Valued Fuzzy and Atanassov's Intuitionistic Fuzzy Connectives from Fuzzy Connectives and from K_α Operators: Laws for Conjunctions and Disjunctions, Amplitude. International Journal of Intelligent Systems, vol. 23 (2008) 680-714.

[4] H. Bustince, E. Barrenechea, M. Pagola. Restricted equivalence functions. Fuzzy Sets and Systems 157 (2006) 2333-2346.

[5] H. Bustince, E. Barrenechea, M. Pagola, R. Orduna. Construction of interval type 2 fuzzy images to represent images in grayscale. False edges. FUZZY IEEE 2007. London, U.K.

[6] T. Calvo, A. Kolesarova, M.Komornikova, R. Masiar. Aggregation operators: properties, classes and construction methods. In T. Calvo, G. Mayor and R. Mesiar (eds): Aggregation Operators New Trends and Applications (Physica-Verlag, Heidelberg, 2002) 3-104.

[7] G. Deschrijver, C. Cornelis, EE.Kerre. On the representation of intuitionistic fuzzy T-norms and T-conorms. IEEE Trans Fuzzy Syst. 12(1) (2004) 45-61.

[8] P. Drygas. The problem of distributivity between binary operations in bi-fuzzy set theory. Proceedings of IPMU'08. L. Magdalena, M. Ojeda-Aciego, J.L. Verdegay (eds). (2008) 1648-1653.

[9] J.A. Noble, D. BouKerroui. Ultrasound image segmentation: A survey. IEEE Transactions on Medical Imaging 25 (2006) 987-1010.

[10] K. Ogawa, M. Fukushima, K. Kubota, N. Hisa. Computer-aided diagnosis system for diffuse liver-diseases with ultrasonography by neural networks. IEEE Transactions on Nuclear Science. 45(6) (1998) 3069-3074.

[11] F. Sahba, H.R. Tizhoosh, M.M. Salama. A coarseto-fine approach to prostate boundary segmentation in ultrasound images. BioMedical Engineering OnLine. (2005) 4-58.

[12] M.H. Yap, E.A. Edirisinghe, H.E. Bez. A novel algorithm for initial lesion detection in ultrasound breast images.Journal of applied clinical medical physics (Fall 2008). Vol. 9, number 4.

[13] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning-I. Information Science. 8 (1975) 199-249.

ISBN: 978-989-95079-6-8

Transitive Closure of Interval-valued Relations

Ramón González del Campo[1] Luis Garmendia[2] Jordi Recasens[3]

1.DSIC Universidad Complutense de Madrid
2.DISIA Universidad Complutense de Madrid
3.Universitat Politécnica de Catalunya
Email: rgonzale@estad.ucm.es, lgarmend@fdi.ucm.es, j.recasens@upc.edu

Abstract— *This paper introduces some concepts of interval-valued fuzzy relations and some of their properties: reflexivity, symmetry, T-transitivity, composition and locally reflexivity. It is also introduced the concept of T-transitive closure for an interval-valued fuzzy relation. An algorithm to compute the T-transitive closure of finite interval-valued relations, some properties and some examples are given.*

Keywords— Generalized t-norms, Interval-valued Fuzzy Relations, Interval-valued Fuzzy Sets, t-norms, t-representable t-norms, t-transitive closure.

1 Introduction

Fuzzy sets, $\mathcal{FS}$, were introduced by Zadeh in 1.965 [1]. Since then many generalizations of fuzzy sets have been proposed to model the uncertainty and the vagueness in linguistic variables replacing the unit interval by another structure such as posets or lattices [2, 3, 4]. One of these generalizations are type-2 fuzzy sets, $\mathcal{FS}2$, [5, 6, 7] were introduced by Zadeh. A Type-2 fuzzy set on a universe of discourse X, $\mathcal{FS}2$, is a fuzzy set whose membership function is another fuzzy set on [0,1]:

$$A = \{((x,u), \mu_A(x,u)) \mid \forall x \in X, \forall u \in [0,1]\}$$

Type-2 fuzzy sets have been widely studied and applied since in many cases the uncertainty can be better expressed by a fuzzy set than by a single numeric value. The problem with type2 fuzzy sets though is their computational complexity and the difficulty for an expert to select the adequate fuzzy subset as membership degree of an object to a linguistic label. This is why some simplifications have been proposed, such as the use of only some families of fuzzy sets such as triangular and trapezoidal ones. One of the most popular and useful simplification is the use of interval-valued fuzzy sets.

Interval-valued fuzzy sets ($\mathcal{IVFS}$) were introduced in the 60s by Grattan-Guinness [8], Jahn [9], Sambuc [10] and Zadeh [5]. They are extensions of classical fuzzy sets where the membership value between 0 and 1 is replaced an interval in [0,1]. They easily allow to model uncertainty and vagueness because sometimes it is easier for experts to give a "membership interval" than a membership degree to objects on a universe. $\mathcal{IVFS}$ are a special case of type-2 fuzzy sets that simplifies the calculations while preserving their richness as well. Intuitionistic fuzzy sets on X ($\mathcal{IFS}$) introduced by Atanassov [11]. In intuitionistic fuzzy sets each element has a membership degree, μ, and a non-membership degree, ν satisfying: $\mu + \nu \leq 1$.

$A = \{(x, \mu(x), \nu(x)) \mid x \in X, \mu(x), \nu(x) \in [0,1]\}$

The value $\pi = 1 - \mu - \nu$ is a measure of the uncertainty. This paper is organized as follows: In section three we introduce some important properties of interval-valued fuzzy relations such as reflexivity, symmetry, T-transitivity, composition and local reflexivity. Traditionally, the study of conjunctions between interval-valued fuzzy sets has been reduced to be modeled with t-representable t-norms. However, not all generalized t-norms are t-representable. Moreover, some of the non t-representable t-norms sometimes satisfy even more properties than t-representable t-norms [12]. Probably the most important property a fuzzy relation can fulfil is transitivity with respect a given t-norm. Since many times the data are given by a proximity relation P (i.e.: a reflexive and symmetric but not necessarily transitive fuzzy relation), there are some methods to obtain a transitive relation close to P to replace it when transitivity is required. The most popular way to do this is calculating its transitive closure. In section four we introduce the concept of T-transitive closure for an interval-valued fuzzy relation and its expression in a finite universe for any t-norm. A few methods to compute it and some examples are given.

2 Preliminaries

2.1 Interval-valued fuzzy sets

Definition 2.1 *[13] Let $\mathcal{L}^I = (L, \leq_L)$ be a lattice that satisfies:*

1. $L = \{[x_1, x_2] \in [0,1]^2$ *with* $x_1 \leq x_2\}$.
2. $[x_1, x_2] \leq_L [y_1, y_2]$ *if and only if* $x_1 \leq y_1$ *and* $x_2 \leq y_2$

Trivially:

$[x_1, x_2] <_L [y_1, y_2] \Leftrightarrow x_1 < y_1, x_2 \leq y_2$ or $x_1 \leq y_1, x_2 < y_2$
$[x_1, x_2] =_L [y_1, y_2] \Leftrightarrow x_1 = y_1, x_2 = y_2$.

$0_L =_L [0,0]$ and $1_L =_L [1,1]$ are the smallest and the greatest elements in L respectively.

$\mathcal{L}^I$ is a complete lattice an the supremum and infimum are defined as follows.

Definition 2.2 *[12] Let $\{[v_i, w_i]\}$ be a set of intervals on L. Then*

1. *Infimum:*
   $Inf_L\{[v_i, w_i]\} \equiv [infimun\{v_i\}, infimun\{w_i\}]$
2. *Supremum:*
   $Sup_L\{[v_i, w_i]\} \equiv [supremun\{v_i\}, supremun\{w_i\}]$

Definition 2.3 *[13] An interval-valued fuzzy set A on a universe X is a mapping:*

ISBN: 978-989-95079-6-8

$$A = \{(a, [x_1, x_2]) \mid a \in X, [x_1, x_2] \in L\}$$

Definition 2.4 *[13] Let X be a universe and A and B two interval-valued fuzzy sets. The equality between A and B is defined as: $A =_L B$ if and only if $A(a) =_L B(a)\ \forall a \in X$.*

Definition 2.5 *[13] Let X be a universe and A and B two interval-valued fuzzy sets. The inclusion of A in to B is defined as: $A \subseteq_L B$ if and only if $A(a) \leq_L B(a)\ \forall a \in X$.*

Definition 2.6 *[13] A negation function for interval-valued fuzzy sets $\mathcal{N}$ is a decreasing function, $\mathcal{N} : L \to L$, that satisfies:*

1. $\mathcal{N}(0_L) =_L 1_L$
2. $\mathcal{N}(1_L) =_L 0_L$

If $\mathcal{N}(\mathcal{N}([x_1, x_2])) =_L [x_1, x_2]$ for all $[x_1, x_2]$ in L then $\mathcal{N}$ is called an involutive negation.

Definition 2.7 *A strong negation function for interval-valued fuzzy sets, $\mathcal{N}$, is a involutive function, $\mathcal{N} : L \to L$, that satisfies:*

1. $\mathcal{N}(0_L) =_L 1_L$
2. $\mathcal{N}(1_L) =_L 0_L$

T-norms are generalized to the lattice $\mathcal{L}^I$ in a straightforward way.

Definition 2.8 *A generalized t-norm function [13], $\mathcal{T}$, is a monotone increasing, symmetric and associative operator, $\mathcal{T} : L^2 \to L$, that satisfies: $\mathcal{T}(1_L, [x_1, x_2]) =_L [x_1, x_2]$ for all $[x_1, x_2]$ in L.*

Due to monotony, it is easy to show:

$$\mathcal{T}(Sup_L\{[v_i, w_i]\}, [y_1, y_2]) \geq_L Sup_L\{\mathcal{T}([v_i, w_i], [y_1, y_2])\}$$
$$\mathcal{T}(Inf_L\{[v_i, w_i]\}, [y_1, y_2]) \leq_L Inf_L\{\mathcal{T}([v_i, w_i], [y_1, y_2])\}$$

Due to the associativity of $\mathcal{T}$ the conjunction of three or more intervals can be defined inductively as:

$$\mathcal{T}(a, \mathcal{T}(b, c)) =_L \mathcal{T}(\mathcal{T}(a, b), c) =_L a \triangle b \triangle c \text{ where } \triangle =_L \mathcal{T}.$$

where $a =_L [a_1, a_2]$, $b =_L [b_1, b_2]$ and $c =_L [c_1, c_2]$.

$\underline{T([x_1, x_2], [y_1, y_2])}$ and $\overline{T([x_1, x_2], [y_1, y_2])}$ will denote the lower and the higher values of $T([x_1, x_2], [y_1, y_2])$.

Definition 2.9 *[14] Let $\{x_i\}$ in $[0, 1]$. A t-norm T in $([0, 1], \leq)$ is left-continuous if it satisfies:*

$$T(Sup\ x_i, y) = Sup\ T(x_i, y)$$

Right-continuity can be defined in a similar way. This property is also called sup-preserving.

Definition 2.10 *[13] A generalized t-norm operator $\mathcal{T}$ is t-representable in $\mathcal{L}^I$ if there are two t-norms: T_1 and T_2 (T_1, T_2, in ([0,1],$\leq$)) that satisfy:*

$$\mathcal{T}([x_1, x_2], [y_1, y_2]) =_L [T_1(x_1, y_1), T_2(x_2, y_2)]$$

where $T_1(v, w) \leq T_2(v, w)\ \forall v, w \in [0, 1]$.

Let $x =_L [x_1, x_2]$ and $y =_L [y_1, y_2]$ be two intervals on L:

Example 2.1
$Inf_L(\{[x_1, x_2], [y_1, y_2]\}) =_L [min(x_1, y_1), min(x_2, y_2)]$ is t-representable by means of $T = min$ in ([0,1],$\leq$):

Example 2.2 *The following product generalized t-norm $*_L$ is t-representable:*

$$*_L([x_1, x_2], [y_1, y_2]) =_L [x_1 * y_1, x_2 * y_2]$$

Example 2.3 *There are two generalizations of the Lukasiewicz t-norm [12]:*

- $T_w([x_1, x_2], [y_1, y_2]) =_L$ $[max(0, x_1 + y_1 - 1), max(0, x_2 + y_2 - 1)]$
- $T_W([x_1, x_2], [y_1, y_2]) =_L$ $[max(0, x_1 + y_1 - 1), max(0, x_1 + y_2 - 1, x_2 + y_1 - 1)]$

Note that T_w is t-representable but T_W is not.

Definition 2.11 *[13] A generalized t-conorm $\mathcal{S}$ is an increasing, commutative, and associative operator $\mathcal{S} : L^2 \to L$, that satisfies: $\mathcal{S}(0_L, [x_1, x_2]) =_L [x_1, x_2]$ and $\mathcal{S}(1_L, [x_1, x_2]) =_L 1_L$.*

Due to the associativity of $\mathcal{S}$ we can write:

$\mathcal{S}(a, \mathcal{S}(b, c)) =_L \mathcal{S}(\mathcal{S}(a, b), c) =_L a \bigtriangledown b \bigtriangledown c$ where $\bigtriangledown =_L \mathcal{S}$.

For example, $\mathcal{S} = Sup_L$ is a generalized t-conorm.
In this paper the next definition is proposed.

Definition 2.12 *Let $\{[v_i, w_i]\}$ be in L. A generalized t-norm operator $\mathcal{T}$ is left-continuous if and only if:*

$$\mathcal{T}(Sup_L\{[v_i, w_i]\}, [y_1, y_2]) =_L Sup_L\{\mathcal{T}([v_i, w_i], [y_1, y_2])\}$$

Right-continuity can be defined in a similar way. A t-norm, a negation and the dual t-conorm of $\mathcal{T}$ with respect to $\mathcal{N}$ is called a De Morgan triplet.

3 Interval-valued fuzzy relations

This section contains some definitions on interval-valued fuzzy relations that generalize the ones of fuzzy relations valued on [0,1]. In the next section, $\mathcal{T}$-transitive relations will be generated using the $Sup - \mathcal{T}$ product.

Definition 3.1 *[15] Let X_1 and X_2 be two universes of discourse. An interval-valued fuzzy relation $R : X_1 \times X_2 \to L$ (where $\mathcal{L}^I = (L, \leq_L)$ is a lattice) is a mapping:*

$$R = \{((a, b), [x, y]) \mid a \in X_1, b \in X_2, [x, y] \in L\}$$

where $x = \underline{R(a, b)}$ and $y = \overline{R(a, b)}$.
In the rest of the paper, it is assumed that $X_1 = X_2 = X$.

Definition 3.2 *An interval-valued relation $R : X^2 \to L$ is reflexive if:*

$$R(a, a) =_L 1_L\ \forall a \in X$$

ISBN: 978-989-95079-6-8

Definition 3.3 *An interval-valued relation $R : X^2 \to L$ is $[\alpha_1 - \alpha_2]-$reflexive if:*

$$R(a,a) \geq_L [\alpha_1, \alpha_2] \; \forall a \in X$$

where $\alpha_1, \alpha_2 \in [0,1]$ and $\alpha_1 \leq \alpha_2$.

Definition 3.4 *An interval-valued relation $R : X^2 \to L$ is symmetric if:*

$$R(a,b) =_L R(b,a) \; \forall a, b \in X$$

Definition 3.5 *Let $\mathcal{T}$ be a generalized t-norm operator and let R interval-valued fuzzy relation on X. R is $\mathcal{T}$-transitive if:*

$$\mathcal{T}(R(a,b), R(b,c)) \leq_L R(a,c) \; \forall a, b, c \in X$$

Proposition 3.1 *If $\mathcal{T}$ is t-representable with T_1 and T_2 ($\mathcal{T} = [T_1, T_2]$) then an interval-valued relation $R : X^2 \to L$ is $\mathcal{T}$-transitive if and only if $\underline{R}$ is T_1-transitive and $\overline{R}$ is T_2-transitive:*

Definition 3.6 *Let $\mathcal{T}$ a generalized t-norm. Let $\mathcal{S}$ a generalized t-conorm from $\mathcal{S}$ and let be $\bigtriangledown$ the n-ary generalized t-conorm defined by associativity. Let R_1 and R_2 be two interval-valued fuzzy relations on a finite set $X = \{c_1, ...c_m\}$. The $\mathcal{S} - \mathcal{T}$-composition of R_1 and R_2 is defined as follows:*

$$(R_1 \diamond_{(\mathcal{ST})} R_2)(c_q, c_r) =_L \bigtriangledown_{c_k \in X} \mathcal{T}(R_1(c_q, c_k), R_2(c_k, c_r))$$

where $\bigtriangledown_{c_k \in X} \mathcal{T}(R_1(c_q, c_k), R_2(c_k, c_r)) =_L \mathcal{T}(R_1(c_q, c_1), R_2(c_1, c_r)) \bigtriangledown ... \bigtriangledown \mathcal{T}(R_1(c_q, c_m), R_2(c_m, c_r))$.

Definition 3.7 *The local equality relation of a fuzzy interval-valued fuzzy relation R on X is the fuzzy relation E_R is:*

$$E_R(a,b) = \begin{cases} Sup_{L\ \forall c \in X} max(R(a,c), R(c,a)) & a = b; \\ [0,0], & a \neq b. \end{cases}$$

Definition 3.8 *An interval-valued relation R is locally reflexive if $E_R \subseteq_L R$.*

It is easy to prove that if R is an reflexive interval-valued relation then R is a locally reflexive interval-valued relation.

4 $\mathcal{T}$-transitive closure for IVFSs

In this section it is proved the existence of the $\mathcal{T}$-transitive closure of an interval valued fuzzy relation and its calculation via $sup - \mathcal{T}$. The $sup - \mathcal{T}$ product is associative if $\mathcal{T}$ is left continuous of the universe is finite. Moreover, it is proved that the $sup - \mathcal{T}$ product is continuous and so is the calculation of the $\mathcal{T}$-transitive closure for finite universes. An algorithm to calculate the $\mathcal{T}$-transitive closure is provided.

Definition 4.1 *Let $\mathcal{T}$ be a generalized t-norm and let $R : X^2 \to L$ be an interval-valued relation on a finite universe X. The $\mathcal{T}$-transitive closure of R is the relation $R^{\mathcal{T}} : X \times X \to L$ that satisfies:*

1. $R \subseteq_L R^{\mathcal{T}}$.
2. *If $R \subseteq_L R'$ and R' is $\mathcal{T}$-transitive then $R^{\mathcal{T}} \subseteq_L R'$.*

Lemma 4.1 *R is $\mathcal{T}$-transitive if and only if $R \diamond_{Sup_L \mathcal{T}} R \subseteq_L R$.*

Proof.

- R is $\mathcal{T}$-transitive $\Rightarrow R \diamond_{Sup_L \mathcal{T}} R \subseteq_L R$:
  $(R \diamond_{Sup_L \mathcal{T}} R)(a,b) =_L$
  $Sup_{L\ c \in X}\{\mathcal{T}(R(a,c), R(c,b))\} \leq_L R(a,b)$
  due to $\mathcal{T}(R(a,c), R(c,b)) \leq_L R(a,b) \; \forall a, b, c \in X$
- $R \diamond_{Sup_L \mathcal{T}} R \subseteq_L R \Rightarrow R$ is $\mathcal{T}$-transitive:
  $Sup_{L\ d \in X}\{\mathcal{T}(R(a,d), R(d,b))\} \leq_L R(a,b) \; \forall a, b \in X$
  Then $\forall c \in X \; \mathcal{T}(R(a,c), R(c,b)) \leq_L R(a,b) \square$

Lemma 4.2 *Let R be an interval-valued relation. If $R^{\mathcal{T}}$ exists then it is unique.*

Proof. Let S_1 and S_2 be two relations. If S_1 and S_2 are transitive closures of R then according to definition 4.1: $S_1 \subseteq_L S_2$ and $S_2 \subseteq_L S_1$, consequently $S_1 =_L S_2 \square$.

Lemma 4.3 *Let R be an interval-valued relation in a universe X and an let $\mathcal{T}$ be an arbitrary generalized t-norm. Then the transitive closure of R always exists.*

Proof. Consider the set Ω_R of $\mathcal{T}$-transitive fuzzy relations containing R. Let us define the fuzzy relation $S_R^*(a,b) =_L Inf_{L;S \in \Omega_R}\{S(a,b)\}$. We will prove that S_R^* is $\mathcal{T}$-transitive. Due to monotony of generalized $\mathcal{T}$-norms it is obtained:
$\mathcal{T}(S_R^*(a,b), S_R^*(b,c)) =_L$
$\mathcal{T}(Inf_{L\ S_1 \in \Omega_R}\{S_1(a,b)\}, Inf_{L\ S_2 \in \Omega_R}\{S_2(b,c)\})$
$\leq_L Inf_{L\ S_1 \in \Omega_R} Inf_{L\ S_2 \in \Omega_R}\{\mathcal{T}(S_1(a,b), S_2(b,c))\}$
$\leq_L Inf_{L\ S \in \Omega_R}\{\mathcal{T}(S(a,b), S(b,c))\} =_L S_R^*(a,c) \square$

Proposition 4.1 *Let X be an arbitrary universe. Let $\{S_1, ..., S_n\}$ a set of interval-valued relations. If $\mathcal{T}$ is a left-continuous generalized $\mathcal{T}$-norm, then it satisfies:*

$$R \diamond_{Sup_L \mathcal{T}} (Sup_{L\ \forall i=1..n}\{S_i\}) =_L Sup_{L\ \forall i=1..n}\{R \diamond_{Sup_L \mathcal{T}} S_i\}$$

Proof.

$$\begin{aligned} &R \diamond_{Sup_L \mathcal{T}} (Sup_{L\ \forall i=1..n}\{S_i\})(a,c) =_L \\ &Sup_{L\ b \in X}\{\mathcal{T}(R(a,b), Sup_{L\ \forall i=1..n}\{S_i(b,c)\})\} \\ &=_L Sup_{L\ \forall i=1..n}\ Sup_{L\ b \in X}\{\mathcal{T}(R(a,b), S_i(b,c))\} \\ &=_L (Sup_{L\ \forall i=1..n}\{R \diamond_{Sup_L \mathcal{T}} S_i\})(a,c) \square \end{aligned}$$

Definition 4.2 *Given a generalized t-norm $\mathcal{T}$, the $\mathcal{T}$-power $R^{(n)_{\mathcal{T}}}$ of a fuzzy relation R on X is recursively defined as follows:*

1. $R^{(1)_{\mathcal{T}}} \equiv R$
2. $R^{(n)_{\mathcal{T}}} \equiv R^{(n-1)_{\mathcal{T}}} \diamond_{Sup_L \mathcal{T}} R$

Lemma 4.4 *If $A \subseteq_L B$ then $A^{(k)} \subseteq_L B^{(k)}$*

Proof.
Trivial due to monotony of $\mathcal{T} \square$.

Theorem 4.1 *Let X be an arbitrary universe and $\mathcal{T}$ a left-continuous t-norm. The transitive closure of R is:*

$$R^{\mathcal{T}} =_L Sup_{L\ \forall k \in \mathbb{N}}\{R^{(k)_{\mathcal{T}}}\}$$

Proof.
Let $S = Sup_{L\ \forall k \in \mathbb{N}}\{R^{(k)_{\mathcal{T}}}\}$ be. By proposition 4.1:

ISBN: 978-989-95079-6-8

$$S^{(2)_\mathcal{T}} =_L$$
$$(Sup_{L\ \forall k\in\mathbb{N}}\{R^{(k)_\mathcal{T}}\}) \diamond_{Sup_L\mathcal{T}} (Sup_{L\ \forall l\in\mathbb{N}}\{R^{(l)_\mathcal{T}}\}) =_L$$
$$Sup_{L\ \forall k,l\in\mathbb{N}}\{R^{(k+l)_\mathcal{T}}\} =_L Sup_{L\ \forall m\in\mathbb{N}\setminus\{0,1\}}\{R^{(m)_\mathcal{T}}\} \subseteq S$$

By lemma 4.1, it follows that S is T-transitive.

Moreover, if $R \subseteq S'$ and S' is T-transitive, then again by lemma 4.1 and 4.4 it holds that:

$$R^{(2)_\mathcal{T}} \subseteq_L S'^{(2)_\mathcal{T}} \subseteq_L S', ..., R^{(k)_\mathcal{T}} \subseteq_L S'^{(k)_\mathcal{T}} \subseteq_L S'$$

hence $S \subseteq_L S'$ and $R^\mathcal{T} =_L S\Box$.

Lemma 4.5 *Let $\mathcal{T}$ be any generalized t-norm, then:*

$$\mathcal{T}([x_1,x_2],[y_1,y_2]) \leq_L Inf_L([x_1,x_2],[y_1,y_2])$$
$$\forall [x_1,x_2],[y_1,y_2] \in L$$

Proof.
$\mathcal{T}([x_1,x_2],[y_1,y_2]) \leq_L \underline{\mathcal{T}([x_1,x_2],1_L) =_L [x_1,x_2]}$
$\Rightarrow [\underline{\mathcal{T}([x_1,x_2],[y_1,y_2])}, \overline{\mathcal{T}([x_1,x_2],[y_1,y_2])}] \leq_L [x_1,x_2]$
$\underline{\mathcal{T}([x_1,x_2],[y_1,y_2])} =_L \overline{\mathcal{T}([y_1,y_2],[x_1,x_2])} \leq_L$
$\underline{\mathcal{T}([y_1,y_2],1_L) =_L [y_1,y_2]}$
$\Rightarrow [\underline{\mathcal{T}([x_1,x_2],[y_1,y_2])}, \overline{\mathcal{T}([x_1,x_2],[y_1,y_2])}] \leq_L [y_1,y_2]$
Therefore:
$\underline{\mathcal{T}([x_1,x_2],[y_1,y_2])} \leq minimum[x_1,y_1]$
$\overline{\mathcal{T}([x_1,x_2],[y_1,y_2])} \leq minimum[x_2,y_2] \Rightarrow$
$\mathcal{T}([x_1,x_2],[y_1,y_2]) \leq_L Inf_L([x_1,x_2],[y_1,y_2])\Box$

Lemma 4.6 *Let $\mathcal{S}$ be any generalized t-conorm, then:*

$$\mathcal{S}([x_1,x_2],[y_1,y_2]) \geq_L Sup_L([x_1,x_2],[y_1,y_2])$$
$$\forall [x_1,x_2],[y_1,y_2] \in L$$

Proof.
$\mathcal{S}([x_1,x_2],[y_1,y_2]) \geq_L \underline{\mathcal{S}([x_1,x_2],1_L) =_L [x_1,x_2]}$
$\Rightarrow [\underline{\mathcal{S}([x_1,x_2],[y_1,y_2])}, \overline{\mathcal{S}([x_1,x_2],[y_1,y_2])}] \geq_L [x_1,x_2]$
$\underline{\mathcal{S}([x_1,x_2],[y_1,y_2])} =_L \overline{\mathcal{S}([y_1,y_2],[x_1,x_2])} \geq_L$
$\underline{\mathcal{S}([y_1,y_2],1_L) =_L [y_1,y_2]}$
$\Rightarrow [\underline{\mathcal{S}([x_1,x_2],[y_1,y_2])}, \overline{\mathcal{S}([x_1,x_2],[y_1,y_2])}] \geq_L [y_1,y_2]$
Therefore:
$\underline{\mathcal{S}([x_1,x_2],[y_1,y_2])} \geq maximun[x_1,y_1]$
$\overline{\mathcal{S}([x_1,x_2],[y_1,y_2])} \geq maximun[x_2,y_2] \Rightarrow$
$\mathcal{S}([x_1,x_2],[y_1,y_2]) \geq_L Sup_L([x_1,x_2],[y_1,y_2])\Box$

Due associativity of $\mathcal{T}$:

$$\mathcal{T}([x_1,x_2],\mathcal{T}([y_1,y_2],[z_1,z_2])) =_L$$
$$\mathcal{T}(\mathcal{T}([x_1,x_2],[y_1,y_2]),[z_1,z_2]) =_L [x_1,x_2] \triangle [y_1,y_2] \triangle [z_1,z_2]$$
$$Inf_L([x_1,x_2],Inf_L([y_1,y_2],[z_1,z_2])) =_L$$
$$Inf_L(Inf_L([x_1,x_2],[y_1,y_2]),[z_1,z_2])) =_L [x_1,x_2] \triangle_{L-inf} [y_1,y_2] \triangle_{L-inf} [z_1,z_2]$$

Lemma 4.7 *Let $\mathcal{T}$ be an arbitrary generalized t-norm. Let P be a path with a cycle:*

$$P \equiv R(a,a_1) \triangle R(a_1,a_2) \triangle ... \triangle R(a_{t-1},a_t) \triangle R(a_t,c_1) \triangle R(c_1,c_2) \triangle ... \triangle R(c_q,a_t) \triangle R(a_t,a_{t+1}) \triangle ... \triangle R(a_k,b)$$

then

$$P \equiv R(a,a_1) \triangle R(a_1,a_2) \triangle ... \triangle R(a_{t-1},a_t) \triangle R(a_t,c_1) \triangle R(c_1,c_2) \triangle ... \triangle R(c_q,a_t) \triangle R(a_t,a_{t+1}) \triangle ... \triangle R(a_k,b) \leq_L R(a,a_1) \triangle R(a_1,a_2) \triangle ... \triangle R(a_{t-1},a_t) \triangle R(a_t,a_{t+1}) \triangle ... \triangle R(a_k,b)$$

Proof.
Trivial due to $\mathcal{T}([x_1,x_2],[y_1,y_2]) \leq_L [y_1,y_2]$ forall $[x_1,x_2],[y_1,y_2]$ in L and associativity of $\mathcal{T}\Box$.

Theorem 4.2 *Let X be a finite universe with cardinality n. The transitive closure of R, $R^\mathcal{T}$, is:*

$$R^\mathcal{T} =_L Sup_{L\ k=1..n}\{R^{(k)_\mathcal{T}}\}$$

Proof.

$R^{(k)_\mathcal{T}}(a,b) =_L$

$=_L Sup_{L\ a_1,a_2,...,a_{k-1}} R(a,a_1) \triangle R(a_1,a_2) \triangle ... \triangle R(a_{k-1},b) \leq_L$
$=_L Sup_{L\ a_1,a_2,...,a_{k-1}} R(a,a_1) \triangle_{Inf_L} R(a_1,a_2) \triangle_{Inf_L} ... \triangle_{Inf_L} R(a_{k-1},b)$ according to lemma 4.5.

Now suppose $k > n$. Then $R^{(k)} \subseteq_L R^{(n)}$ because it exists a cycle (lemma 4.7). Therefore, the cases $k > n$ need not be considered $\Box$.

Corollary 4.1 *Let X be a finite universe with cardinality n. If R is a locally reflexive relation, the transitive closure of R is:*

$$R^\mathcal{T} =_L Sup_{L\ k=1..n-1}\{R^{(k)_\mathcal{T}}\}$$

Proof
If R is a locally reflexive relation then $R(a,b) \leq_L R(a,a)$. Consider a cycle of length n, then its strength is:
$R^{(n)_\mathcal{T}}(a,a) =_L$

$=_L Sup_{L\ a_1,a_2,...,a_{k-1}} R(a,a_1) \triangle R(a_1,a_2) \triangle ..., \triangle R(a_{k-1},a) \leq_L$
$Sup_{L\ a_1,a_2,...,a_{k-1}} R(a,a_1) \triangle_{Inf_L} R(a_1,a_2) \triangle_{Inf_L} ... \triangle_{Inf_L} R(a_{k-1},a) \leq_L R(a,a)\Box$

Theorem 4.3 *Let X be a finite universe with cardinality n. If R is a reflexive fuzzy relation on X then the transitive closure of R is:*

$$R^\mathcal{T} =_L R^{(n-1)_\mathcal{T}}$$

Proof.
$R(a,b)^{(k)_\mathcal{T}} =_L$

$=_L \mathcal{T}(R(a,a), R^{(k)_\mathcal{T}}(a,b))$
$\leq_L Sup_{L\ c\in X}\{T(R(a,c), R^{(k)_\mathcal{T}}(c,b))\}$
$=_L R^{(k+1)_\mathcal{T}}(a,b)$

Therefore $R(a,b)^{(k)_\mathcal{T}} \leq_L R^{(k+1)_\mathcal{T}}(a,b) \Leftrightarrow R^{(k)_\mathcal{T}} \subseteq R^{(k+1)_\mathcal{T}}\Box$

Theorem 4.4 *Let $\mathcal{T}$ be a t-representable t-norm ($\mathcal{T} = [T_1,T_2]$) and let $R = [\underline{R},\overline{R}]$ be a interval-valued relation. Then $R^\mathcal{T} = [\underline{R}^{T_1}, \overline{R}^{T_2}]$.*

Proof.

$R^{(k)_\mathcal{T}}(a,b) =_L$

$=_L \ Sup_{L\ a_1,a_2,...,a_{k-1}} R(a,a_1) \ \triangle \ R(a_1,a_2) \ \triangle$
$..., \triangle \ (a_{k-1},b)$
$=_L \ Sup_{L\ a_1,a_2,...,a_{k-1}} [\underline{R(a,a_1)} \ \triangle_1 \ \underline{R(a_1,a_2)} \ \triangle_1$
$..., \triangle_1 \ \underline{R(a_{k-1},b)}, \overline{R(a,a_1)} \ \triangle_2 \ \overline{R(a_1,a_2)} \ \triangle_2$
$..., \triangle_2 \ \overline{R(a_{k-1},b)}]$
$=_L \ [max_{a_1,a_2,...,a_{k-1}} \underline{R(a,a_1)} \ \triangle_1 \ \underline{R(a_1,a_2)} \ \triangle_1$
$..., \triangle_1 \ \underline{R(a_{k-1},b)}$
$, max_{a_1,a_2,...,a_{k-1}} \overline{R(a,a_1)} \ \triangle_2 \ \overline{R(a_1,a_2)} \ \triangle_2 ..., \triangle_2$
$\overline{R(a_{k-1},b)}]$
$=_[\ \underline{R^{(k)_{T_1}}(a,b)}, \overline{R^{(k)_{T_2}}(a,b)}]$

where $\triangle = [\triangle_1, \triangle_2]$. Then $\mathcal{T} = [T_1, T_2] \Rightarrow R^{\mathcal{T}} = [\underline{R}^{T_1}, \overline{R}^{T_2}]\Box$.

Definition 4.3 *The distance d of the supremum on L is defined for all $x_1, x_2, y_1, y_2 \in [0,1]$ by $d([x_1,x_2],[y_1,y_2]) = max(|x_1 - y_1|, |x_2 - y_2|)$.*

Definition 4.4 *Let R,S be two interval-valued fuzzy relatons on a set X. The distance d between R and S us defined by:*

$$d(R,S) = sup_{x,y \in X} d(R(x,y), S(x,y))$$

Lemma 4.8 *d is a distance on the set R_X of interval-valued fuzzy relations on X $\Box$.*

Proof.
The supremum of distances is a distance.

Theorem 4.5 *Let T be a generalized continuous t-norm and R_X the set of interval-valued fuzzy relations on X. R_X with the sup-$\mathcal{T}$ product is an ordered topological semigroup.*

Proof.

- Associativity is a straightforward exercise.
- The interval-valued relation $E(x,y) = \begin{cases} [1,1], & x=y; \\ [0,0], & \text{otherwise.} \end{cases}$ is the identity element of R_X.
- Continuity: Since T is defined on a compact set, it is uniformly continuous. Therefore:

$$\forall \epsilon > 0, \exists \delta > 0 \text{ such that}$$
$$\forall m, n, m', n', a, b, a', b' \in [0,1]$$
$$max(|m-a|, |m'-a'|, |n-b|, |n'-b'|) \Rightarrow$$
$$|\mathcal{T}([m,m'],[n,n']) - \mathcal{T}([a,a'],[b,b'])| < \epsilon (*)$$

We want to prove that given two interval-valued fuzzy relations $A, B \in R_X$:

$$\forall \epsilon > 0, \exists \delta > 0 \text{ such that } \forall M, N \in R_X$$
$$max(d(M,A), d(N,B)) < \delta \Rightarrow d(M \circ N, A \circ B) < \epsilon$$

Given ϵ, we take $\delta > 0$ satisfying (*). Then:

$$d(M \circ N, A \circ B) =$$
$$sup_{x,y \in X} |sup_{z \in X} \mathcal{T}(M(x,z), N(z,y)) -$$
$$sup_{z \in X} \mathcal{T}(A(x,z), B(z,y))| \leq$$
$$sup_{x,y \in X} sup_{z \in X} |\mathcal{T}(M(x,z), N(z,y)) -$$
$$\mathcal{T}(A(x,z), B(z,y))| \leq \epsilon$$

- Monotonicity is an immediate consequence of the monotonicity of $\mathcal{T}\Box$.

Corollary 4.2 *For any positive integer n the map assigning to an interval-values fuzzy relation M on a finite set its n^{th} power is non-decreasing and continuous if the corresponding generalized t-norm $\mathcal{T}$ is continuous.*

Corollary 4.3 *If a generalized t-norm $\mathcal{T}$ is continuous, the map that assigns the $\mathcal{T}$-transitive closure to the interval-valued fuzzy relations on a finite set is non-decreasing and continuous.*

Theorem 4.6 *Given a method to calculate the transitive closure of a fuzzy relation. Nevertheless, there are more effective methods and algorithms. For example it is possible to use an extension of the Floyd-Warshall algorithm [16]*

4.1 Algorithm to compute the transitive closure of an interval-valued fuzzy relation

Let R be an interval-valued relation on a finite universe X with cardinality n and let $\mathcal{T}$ be a generalized t-norm. It is possible to compute $R^{\mathcal{T}}$ using the next algorithm:

for k $\leftarrow$ 1 **until** n **do**
  for i $\leftarrow$ 1 **until** n **do**
    for j $\leftarrow$ 1 **until** n **do**
      R(i,j)$\leftarrow Sup_L(R(i,j), \mathcal{T}(R(i,k), R(k,j)))$
    end for
  end for
end for

Example 4.1 *Let $\mathcal{T}$ be a generalized t-norm $T_w([x_1,x_2],[y_1,y_2]) =_L [max(0, x_1 + y_1 - 1), max(0, x_2 + y_2 - 1)]$ and let $R : X \times X \rightarrow L$ be the following interval-valued relation:*

$$R = \begin{pmatrix} [1,1] & [0.6,0.8] & [0.6,0.9] & [0,0] \\ & [1,1] & [0.4,0.9] & [0.6,0.9] \\ & & [1,1] & [0.6,0.9] \\ & & & [1,1] \end{pmatrix}$$

The computation of the $\mathcal{T}$-transitive closure of R is the following:

$$R^{\mathcal{T}} = \begin{pmatrix} [1,1] & [0.6,0.8] & [0.6,0.9] & [0.2,0.8] \\ & [1,1] & [0.4,0.9] & [0.6,0.9] \\ & & [1,1] & [0.6,0.9] \\ & & & [1,1] \end{pmatrix}$$

Note that by theorem 4.4, $R^{\mathcal{T}} = [\underline{R^W}, \overline{R^W}]$

Example 4.2 *Let $\mathcal{T}$ be a generalized t-norm $T([x_1,x_2],[y_1,y_2]) =_L [W(x_1,y_1), Min(x_2,y_2)]$ and let $R : X \times X \rightarrow L$ be the following interval-valued relation:*

$$R = \begin{pmatrix} [1,1] & [0.6,0.8] & [0.6,0.9] & [0,0] \\ & [1,1] & [0.4,0.9] & [0.6,0.9] \\ & & [1,1] & [0.6,0.9] \\ & & & [1,1] \end{pmatrix}$$

The computation of the $\mathcal{T}$-transitive closure of R is the following:

ISBN: 978-989-95079-6-8

$$R^{\mathcal{T}} = \begin{pmatrix} [1,1] & [0.6,0.8] & [0.6,0.9] & [0.2,0.9] \\ & [1,1] & [0.4,0.9] & [0.6,0.9] \\ & & [1,1] & [0.6,0.9] \\ & & & [1,1] \end{pmatrix}$$

Note that by theorem 4.4, $R^{\mathcal{T}} = [\underline{R^W}, \overline{R^{Min}}]$

Example 4.3 *Let* $\mathcal{T}$ *be a generalized t-norm* $T_W([x_1,x_2],[y_1,y_2]) =_L [max(0, x_1 + y_1 - 1), max(0, x_1 + y_2 - 1, x_2 + y_1 - 1)]$ *and let* $R : X \times X \to L$ *be the following interval-valued relation:*

$$R = \begin{pmatrix} [1,1] & [0.6,0.8] & [0.6,0.9] & [0,0] \\ & [1,1] & [0.4,0.9] & [0.6,0.9] \\ & & [1,1] & [0.6,0.9] \\ & & & [1,1] \end{pmatrix}$$

The computation of the $\mathcal{T}$*-transitive closure of* R *is the following:*

$$R^{\mathcal{T}} = \begin{pmatrix} [1,1] & [0.6,0.8] & [0.6,0.9] & [0.2,0.5] \\ & [1,1] & [0.4,0.9] & [0.6,0.9] \\ & & [1,1] & [0.6,0.9] \\ & & & [1,1] \end{pmatrix}$$

Note that the generalized $\mathcal{T}$*-norm in this example is not t-representable.*

5 An Application

It is possible to model social relations using Interval-valued Fuzzy Sets. For example, it is possible to model "friendship". Let R be a relation which describes the degree of friendship between pairs of people. If A and B are friends and B and C are friends, then it will probably exist some degree of friendship between A and C. In this context, the $\mathcal{T}$-transitive closure can be used to learn friendship of individuals from a chain of friends, for example, from common friends.

Example 5.1 *Let* R *be an interval-valued relation which models the degree of friendship between four people* $\{a, b, c, d\}$*. It is possible to model the transitivity of friendship using the generalized t-norm* $*_L([x_1,x_2],[y_1,y_2]) =_L [x_1 * y_1, x_2 * y_2]$*:*

$$R = \begin{pmatrix} [1,1] & [0.6,0.9] & [0.6,0.7] & [0,0] \\ & [1,1] & [0,0] & [0.2,0.8] \\ & & [1,1] & [0,0] \\ & & & [1,1] \end{pmatrix}$$

Initially, d *has only friendship with* b*. In addition, the uncertainty of degree of friendship between* b *and* d *is wide. It is interesting to see how friendship is propagated.*

$$R^{\mathcal{T}} = \begin{pmatrix} [1,1] & [0.6,0.9] & [0.6,0.7] & [0.16,0.72] \\ & [1,1] & [0.48,0.63] & [0.2,0.8] \\ & & [1,1] & [0.096,0.5] \\ & & & [1,1] \end{pmatrix}$$

In $R^{\mathcal{T}}$ d *has some degree of friendship with* a, c *learned by a common friend* b*. The uncertainty of degree of friendship of* d *with the rest of them has been propagated as well.*

6 Conclusions

Traditionally, the study of interval-valued fuzzy sets has been reduced to t-representable t-norms. We have generalized the main properties of interval-valued fuzzy relations: reflexivity, symmetry, composition, local reflexivity and t-transitivity. We also have generalized the T-transitive closure for any t-norm and an algorithm to compute it is provided. It is analyzed the case of t-representable t-norms and proved that in this case the T-transitive closure of R can be obtained from the two classical T-transitive closures of the lower and higher relations (theorem 4.4). Some examples are provided.

References

[1] L.A. Zadeh. Fuzzy sets. *Information And Control*, 8:338–353, 1965.

[2] H. Bustince, F. Herrera, and J. Montero. *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models*, volume 220 of *Studies in Fuzziness and Soft Computing*. Hardcover, 2008.

[3] D. Gomez J. Montero and Bustince H. On the relevance of some families of fuzzy sets. *Fuzzy Sets And Systems*, 158:2429–2442, 2007.

[4] J. Goguen. L-fuzzy sets. *Journal Of Mathematical Analysis And Applications*, 18:145–174, 1967.

[5] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning i. *Information Sciences*, 8:199–249, 1975.

[6] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning ii. *Information Sciences*, 8:301–357, 1975.

[7] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning iii. *Information Sciences*, 9:43–80, 1975.

[8] I. Grattan-Guiness. Fuzzy membership mapped onto interval and many-valued quantities. *Math. Logik. Grundladen Math*, 22:149–160, 1975.

[9] K.U. Jahn. Intervall-wertige mengen. *Math. Nach.*, 68:115–132, 1975.

[10] E. Sanchez and R. Sambuc. Fuzzy relationships. phi -fuzzy functions. application to diagnostic aid in thyroid pathology. *Proceedings of an International Symposium on Medical Data Processing*, pages 513–524, 1976.

[11] K.T. Atanassov. *Intuitionistic Fuzzy Sets*. Physica-Verlag, Heidelberg, New York,, 1999.

[12] C. Cornelis, G. Deschrijver, and E. Kerre. Advances and challenges in interval-valued fuzzy logic. *Fuzzy Sets and Systems*, 157(5):622–627, 2006.

[13] C. Cornelis, G.Deschrijver, and E. Kerre. Implication in intuitionistic fuzzy and interval-valued fuzzy set theory: construction, classification, application. *Int. J. Approx. Reasoning*, 35(1):55–95, 2004.

[14] B. De Baets and H. De Meyer. On the existence and construction of t-transitive closures. *Inf. Sci.*, 152(1):167–179, 2003.

[15] H. Bustince and P. Burillo. Mathematical analysis of interval-valued fuzzy relations: application to approximate reasoning. *Fuzzy Sets Syst.*, 113(2):205–219, 2000.

[16] H. Naessens, H. De Meyer, and De B. Baets. Algorithms for the computation of t-transitive closures. *IEEE Transactions on Fuzzy Systems*, 10(1):541–551, 2002.

ISBN: 978-989-95079-6-8

Choquet-integral-based Evaluations by Fuzzy Rules

Eiichiro Takahagi

School of Commerce, Senshu University,
Tamaku, Kawasaki 214-8580, Japan
Email: takahagi@isc.senshu-u.ac.jp

Abstract— *Choquet-integral-based evaluation models are proposed. The evaluation parameters – fuzzy measures – are assigned from a fuzzy rule table. There are three variations in this model: TF-, BP-, and AV-type models. The TF-type model is a natural extension of ordinal Choquet integrals. The BP-type model involves an evaluation using a reference point. The AV-type model involves a neutral evaluation method. These methods have a number of advantages: Output values are continuous and piecewise linear; If the fuzzy rules have monotonicity, output values are also monotone. We compare these models with some fuzzy reasoning models.*

Keywords— Choquet Integral, Fuzzy rule, Simplified fuzzy reasoning, Cumulative prospect theory, Choquet integral with respect to bi-capacities

1 Introduction

Fuzzy reasoning models are very useful tools for developing fuzzy control models. A part of a fuzzy control model involves sensing status, evaluation using fuzzy reasoning models, and producing output. By repeatedly applying the feedback process, fuzzy control models help approach the control goal (Approach Principal [1]). However, to be applied in the social sciences, evaluation models must be able to determine one global evaluation value by a single calculation process. It is important that the global evaluation values adhere obsequiously to fuzzy rules. For example, if a fuzzy rule is monotone with respect to the inputs, the output of the evaluation model must satisfy the monotonicity property with respect to the inputs. Almost all the fuzzy reasoning models do not satisfy the monotonicity property [2].

In this paper, we propose Choquet-integral-based evaluation methods by using fuzzy rules. The fuzzy rule table is of the same form as ordinal simplified fuzzy reasoning models, but the calculations use (extended) Choquet integrals instead of min-max calculations or product-sum calculations. The global evaluation values satisfy continuous and piecewise linear outputs. Moreover, if fuzzy rules are monotone respect to the inputs, the output values also satisfy the monotonicity property.

2 Definitions

2.1 Notations

X**:** Set of evaluation items (n: number of evaluation items)

x_i**:** Input value of the i^{th} item

y**:** Global evaluation value

2.2 Fuzzy Space Division Constraint

Each input i is divided into m_i fuzzy sets in which the membership functions are p_i^j for $i = 1, \ldots, n, j = 1, \ldots, m_i$. For each input item, all the membership functions satisfy the following conditions:

1. All fuzzy sets are normal and convex. The vertex of a membership function is unique $\forall i, j$, that is, there is a unique point v_i^j, where $p_i^j(v_i^j) = 1\ \forall i$ and j.

2. The sum of the membership values is 1:

$$\sum_j p_i^j(x_i) = 1, \forall x_i, i. \tag{1}$$

3. There are one or two active membership functions – $p_i^j(x_i) > 0 - \forall x_i, \forall i$.

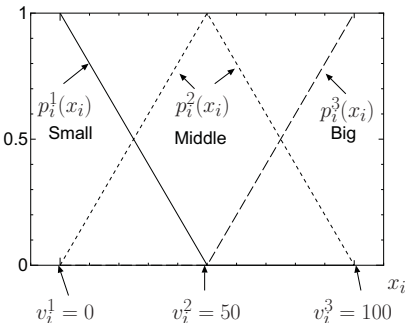


Figure 1: Fuzzy space division ($i = 1, 2$)

An example of the fuzzy space division is shown in figure 1. In this example, there are two inputs, $X = \{1, 2\}$, the membership functions are of the triangle type, and the membership functions are the same for all inputs, $v_i^1 = 0, v_i^2 = 50, v_i^3 = 100, i = 1, 2$.

2.3 Fuzzy Rule Table

Representative points $(k_1, \ldots, k_n)$ are defined as pairs of vertex numbers. For example, the representative point $(2, 3)$ $(n = 2)$ is the point at which input 1 is "Middle" and input 2 is "Big." The fuzzy rule table c is defined as a function from the representative points to the output values:

$$c : \{1, \ldots, m_1\} \times \ldots \times \{1, \ldots, m_n\} \to \mathbb{R}. \tag{2}$$

Table 1 is an example of the fuzzy rule tables,

- If input 1 is "Big" and input 2 is "Big," the output is 100.

ISBN: 978-989-95079-6-8

- If input 1 is "Big" and input 2 is "Middle," the output is 80.
- ...

Table 1: Fuzzy rule table

Input 2 \ 1	Small (1)	Middle (2)	Big (3)
Small (1)	0 (= $c(1,1)$)	50 (= $c(2,1)$)	80 (= $c(3,1)$)
Middle (2)	30 (= $c(1,2)$)	60 (= $c(2,2)$)	90 (= $c(3,2)$)
Big (3)	70 (= $c(1,3)$)	80 (= $c(2,3)$)	100 (= $c(3,3)$)

2.4 Choquet Integral and Extended Choquet Integral

The Choquet integral [3] is a useful tool for use in global evaluation methods. The integral can be used to represent complementary and substitute relations among evaluation items [4]. A fuzzy measure $\mu^\dagger$ is defined as

$$\mu^\dagger : 2^X \to \mathbb{R} \tag{3}$$

$$\mu^\dagger(\emptyset) = 0. \tag{4}$$

In ordinal fuzzy measure definitions, this definition is known as a non-monotonicity fuzzy measure, but in this paper, the monotonicity of the fuzzy measure is not assumed.

The Choquet integral is defined as

$$y = (\mathrm{C}) \int h d\mu \equiv \int_0^\infty \mu(\{x \mid h(x) > r\}) dr. \tag{5}$$

The extended fuzzy measure and extended Choquet integral [5, 6] are proposed for handling the cases in which $\mu(\emptyset) \neq 0$. An extended fuzzy measure μ is defined as

$$\mu : 2^X \to \mathbb{R}. \tag{6}$$

Since it is not assumed by definition that $\mu(\emptyset) = 0$, the integration interval is limited to $[0, 1]$. The extended Choquet integral is defined as

$$y = (\mathrm{EC}) \int h d\mu \equiv \int_0^1 \mu(\{x \mid h(x) > r\}) dr. \tag{7}$$

The extended Choquet integral can be calculated by using an ordinal Choquet integral:

$$(\mathrm{EC}) \int h d\mu = (\mathrm{C}) \int h d\mu^\dagger + \mu(\emptyset) \tag{8}$$

$$\text{where } \mu^\dagger(A) = \mu(A) - \mu(\emptyset), \forall A \in 2^X.$$

3 Segment Division Calculation Method

3.1 Segmentation and Segment Selection

Segmentation is performed at vertex points v_i^j for all inputs. Figure 2 shows an example of the segmentation. Segment $S_{(k_1,\ldots,k_n)}(k_i < m_i, \forall i)$ is the n-th rectangle whose vertices are the representative points $(k_1 + l_1, \ldots, k_2 + l_n), \forall l_i \in \{0, 1\}, i = 1, \ldots, n$. In the example, there are four segments, $S_{(1,1)}$, $S_{(2,1)}$, $S_{(1,2)}$ and $S_{(2,2)}$.

First, we select the segment that includes the input values $((x_1, \ldots, x_n))$. If $x_1 = 40$ and $x_2 = 20$, $S_{(1,1)}$ is selected.

In this model, each segment has a different extended fuzzy measure $\mu^{(k_1,\ldots,k_n)}$ and integrand $h^{(k_1,\ldots,k_n)}$, but the Choquet integral is caclculated only for the selected segment.

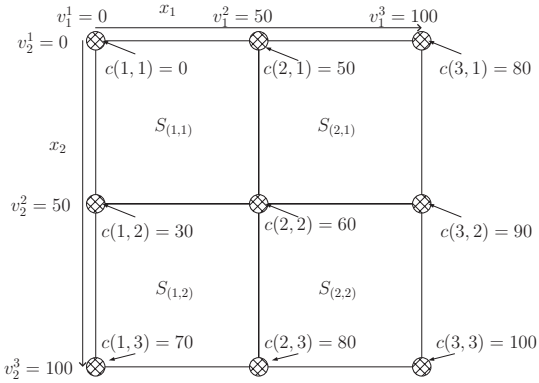


Figure 2: Segmentation

3.2 Segment Calculation

Figure 3 shows segment $S_{(1,1)}$. The output is calculated by interpolating the output values of four representative points; those values are $c(1,1), c(2,1), c(1,2)$, and $c(2,2)$.

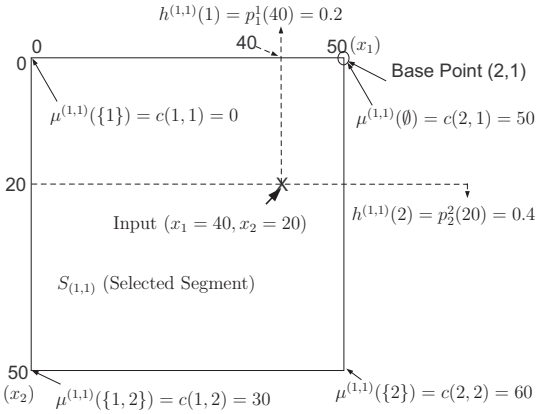


Figure 3: Calculation process (Base Point: $(2,1)$)

Base point selection The base point is the origin of the fuzzy measure and integrand for the segment. First, the base point is selected from among the representative points in the segment. Alternative base points of segment $S_{(k_1,\ldots,k_n)}$ are $(k_1 + l_1, \ldots, k_n + l_n), \forall l_i \in \{0, 1\}, i = 1, \ldots, n$.

Alternative base points of segment $S_{(1,1)}$ in figure 3 are $(1,1)$,$(2,1)$,$(1,2)$, and $(2,2)$.

Extended fuzzy measure $\mu^{(k_1,\ldots,k_n)}$ assignment When the segment $S_{(k_1,\ldots,k_n)}$ and the base point $(q_1, \ldots, q_n)$ are selected, extended fuzzy measures $\mu^{(k_1,\ldots,k_n)}$ are assigned as follows:

$$\mu^{(k_1,\ldots,k_n)}(A) = c(l_1, \ldots, l_n), \forall A \in 2^X \tag{9}$$

$$\text{where } l_i = \begin{cases} q_i & \text{if } i \notin A \\ k_i & \text{if } i \in A \text{ and } k_i \neq q_i \\ k_i + 1 & \text{if } i \in A \text{ and } k_i = q_i. \end{cases}$$

If $(2,1)$ is selected as the base point in figure 3,

$$\mu^{(1,1)}(\emptyset) = c(2,1),\ \mu^{(1,1)}(\{1\}) = c(1,1)$$
$$\mu^{(1,1)}(\{2\}) = c(2,2),\ \mu^{(1,1)}(\{1,2\}) = c(1,2)$$

ISBN: 978-989-95079-6-8

Integrand $h^{(k_1,\ldots,k_n)}$ assignment When the segment $S_{(k_1,\ldots,k_n)}$ and the base point $(q_1,\ldots,q_n)$ are selected, integrands $h^{(k_1,\ldots,k_n)}$ are assigned as follows:

$$h^{(k_1,\ldots,k_n)}(i) = p_i^j(x_i), i = 1,\ldots,n \quad (10)$$
$$\text{where } j = \begin{cases} k_i & \text{if } k_i \neq q_i \\ k_i + 1 & \text{otherwise.} \end{cases}$$

If $(2,1)$ is selected as the base point of segment $S_{(1,1)}$,

$$h^{(1,1)}(1) = p_1^1(x_1), \;\; h^{(1,1)}(2) = p_2^2(x_2) \quad (11)$$

Extended Choquet integral The output value y is calculated from the extended Choquet integral by using the fuzzy measure $\mu^{(k_1,\ldots,k_n)}$ and integrand $h^{(k_1,\ldots,k_n)}$ of the selected segment.

$$y = (\mathrm{EC}) \int h^{(k_1,\ldots,k_n)} d\mu^{(k_1,\ldots,k_n)} \quad (12)$$

The output values change with the methods for selection the base points. We propose the following type of Choquet-integral-based evaluation methods: TF-, BP- and BP-type.

3.3 TF-type

In TF-type models, the smallest (upper left) representative points are selected as base points for all segments, that is, the base point of a segment $S_{(k_1,\ldots,k_n)}$ is $(k_1,\ldots,k_n)$.

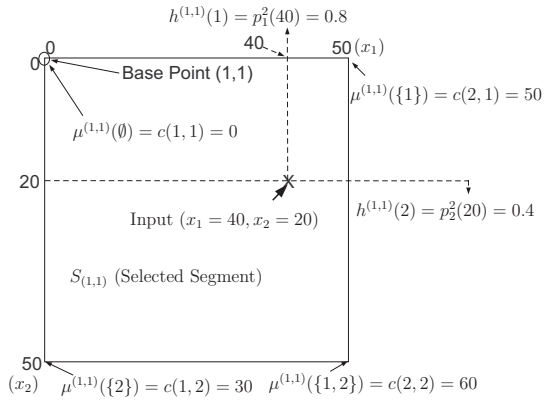


Figure 4: Extended fuzzy measure assignment in TF-type models

When $x_1 = 40$ and $x_2 = 20$, from figure 2, segment $S_{(1,1)}$ is selected. In the TF-type model, $(1,1)$ is selected as the base point. Figure 4 shows an example of the calculation process. Since the origin is $(1,1)$, the membership functions "Middle" ($p_1^2(x_1)$ and $p_2^2(x_2)$) are used. Therefore, the integrand $h^{(1,1)}(1) = p_1^2(x_1) = 0.8$ and $h^{(1,1)}(2) = p_2^2(x_2) = 0.4$. Since the origin is $(1,1)$, $\mu^{(1,1)}(\emptyset) = c(1,1) = 0$, $\mu^{(1,1)}(\{1\}) = c(2,1) = 50$, $\mu^{(1,1)}(\{2\}) = c(1,2) = 30$, and $\mu^{(1,1)}(\{1,2\}) = c(2,2) = 60$ (figure 4). The output is calculated as

$$y = (\mathrm{EC}) \int h^{(1,1)} d\mu^{(1,1)}. \quad (13)$$

Since $\mu^{11}(\emptyset) = 0$, this equation is calculated by using the ordinal Choquet integral.

$$y = [h^{(1,1)}(1) - h^{(1,1)}(2)]\mu^{(1,1)}(\{1\}) + h^{(1,1)}(2)\mu^{(1,1)}(\{1,2\}) = 44. \quad (14)$$

Table 2 lists integrands and extended fuzzy measures for all segments. Figure 5 shows a graph corresponding to Table1.

Table 2: Integrand and extended fuzzy measure for TF ($n = 2$)

		Integrand		μ			
Segment	Base Point	$h(1)$	$h(2)$	$\emptyset$	$\{1\}$	$\{2\}$	$\{1,2\}$
$S_{(1,1)}$	$(1,1)$	$p_1^2(x_1)$	$p_2^2(x_2)$	$c(1,1)$	$c(2,1)$	$c(1,2)$	$c(2,2)$
$S_{(2,1)}$	$(2,1)$	$p_1^3(x_1)$	$p_2^2(x_2)$	$c(2,1)$	$c(3,1)$	$c(1,2)$	$c(3,2)$
$S_{(1,2)}$	$(1,2)$	$p_1^2(x_1)$	$p_2^3(x_2)$	$c(1,2)$	$c(2,2)$	$c(1,3)$	$c(2,3)$
$S_{(2,2)}$	$(2,2)$	$p_1^3(x_1)$	$p_2^3(x_2)$	$c(2,2)$	$c(3,2)$	$c(2,3)$	$c(3,3)$

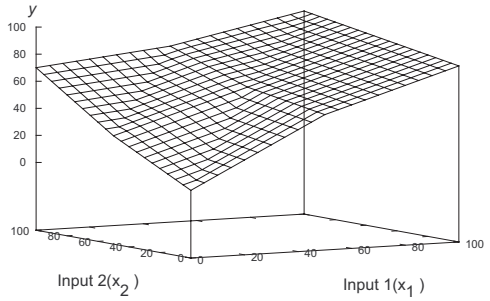


Figure 5: 3D graph corresponding to Table1 (TF-Type)

3.4 Base Point Change

The output values y are dependent on the selection of the base point. If the base point $(2,1)$ is selected for segment $S_{(1,1)}$, the extended fuzzy measure $\mu^{(1,1)}$ is

$$\mu^{(1,1)}(\emptyset) = c(2,1) = 50, \;\; \mu^{(1,1)}(\{1\}) = c(1,1) = 0,$$
$$\mu^{(1,1)}(\{2\}) = c(2,2) = 60, \;\; \text{and } \mu^{(1,1)}(\{1,2\}) = c(1,2) = 30.$$

If the input values are $x_1 = 40, x_2 = 20$, $h^{(1,1)}(1) = p_1^1(x_1) = 0.2$ and $h^{(1,1)}(2) = p_2^2(x_2) = 0.4$. The output value is

$$y = (\mathrm{EC}) \int h^{(1,1)} d\mu^{(1,1)} = 48,$$

which is not the equal to the output value 44 given by the TF-type models (14).

3.5 BP-type

In cumulative prospect theory [7], the evaluation methods differ for values above and below the reference point. The model has been developed for analyzing such situations. BP-type calculations can represent the generalized Choquet integral with respect to bi-capacity [8, 9], and the generalized Choquet integral is an extension of cumulative prospect theory.

In the BP-type model, each input has three membership functions: negative, reference, and positive; $m_i = 3, \forall i$. In prospect theory, it is important to evaluate the distance from the reference point, and therefore, the base point for the all segments is the reference point (figure 6 and table 4), namely $(2,\ldots,2)$.

ISBN: 978-989-95079-6-8

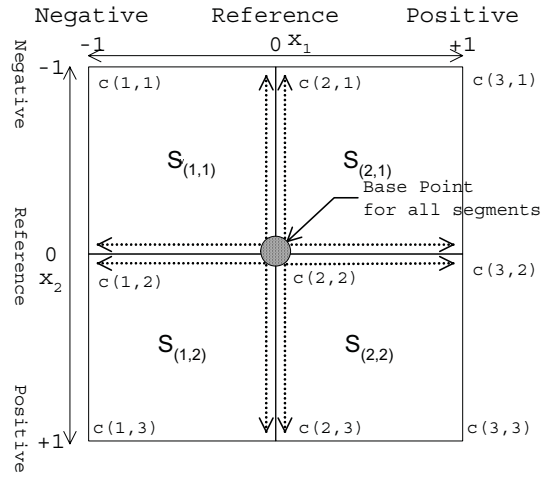


Figure 6: BP-type

BP type model Table 3 is an example of a BP-type evaluation rule table. In thoose BP-type models, membership functions are of the triangle type, such as figure 1; the base points for all segments is the reference point $(2,\ldots,2)$, and the output of the reference point is 0, that is, $c(2,\ldots,2) = 0$ and $x_i \in [-1,1]$. The integrands $h^{(k_1,\ldots,k_n)}$ are

$$h^{(k_1,\ldots,k_n)}(i) = \mid x_i \mid, \forall(k_1,\ldots,k_n), i = 1,\ldots,n. \quad (15)$$

Table 4 lists the integrands and extended fuzzy measures for all segments when $n = 2$. If $x_1 = 0.2$ and $x_2 = -0.6$, $S_{(2,1)}$ is selected and $h^{(2,1)}(1) = 0.2$ and $h^{(2,1)}(2) = 0.6$;

$$y = (\mathrm{EC}) \int h^{(2,1)} d\mu^{(2,1)} = -0.44.$$

Figure 7 shows a graph corresponding to table 3.

Table 3: BP-type fuzzy rule table

Input 2 \ 1	Negative	Reference	Positive
Negative	-1.0	-0.8	-0.6
Reference	-0.8	0	$+0.6$
Positive	-0.6	$+0.8$	$+1.0$

Table 4: Integrands and extended fuzzy measures for BP

		Integrands		μ			
Segments	Base Points	$h(1)$	$h(2)$	$\emptyset$	$\{1\}$	$\{2\}$	$\{1,2\}$
$S_{(1,1)}$	$(2,2)$	$-x_1$	$-x_2$	$c(2,2)$	$c(1,2)$	$c(2,1)$	$c(1,1)$
$S_{(2,1)}$	$(2,2)$	x_1	$-x_2$	$c(2,2)$	$c(3,2)$	$c(2,1)$	$c(3,1)$
$S_{(1,2)}$	$(2,2)$	$-x_1$	x_2	$c(2,2)$	$c(1,2)$	$c(2,3)$	$c(1,3)$
$S_{(2,2)}$	$(2,2)$	x_1	x_2	$c(2,2)$	$c(3,2)$	$c(2,3)$	$c(3,3)$

Bi-capacity Model Let $Q(X) = \{(A,B) \in 2^X \times 2^X \mid A \cap B = \emptyset\}$. Bi-capacity [8] is defined as a function $v : Q(X) \to [-1,1]$ that satisfies the following conditions:

- If $A \subset A'$, $v(A,B) \leq v(A',B)$ and if $B \subset B'$, $v(A,B) \geq v(A,B')$,
- $v(\emptyset,\emptyset) = 0$ and

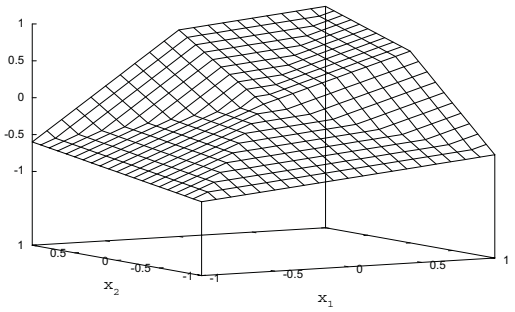


Figure 7: Outputs corresponding to Table 3 (BP-type))

- $v(X,\emptyset) = 1, v(\emptyset,X) = -1$.

The Choquet integral with respect to bi-capacity[9] is defined as

$$C_v(x) = \sum_{i=1}^{n} \mid x_{\sigma(i)} \mid [v(A_{\sigma(i)} \cap X^+, A_{\sigma(i)} \cap X^-) - v(A_{\sigma(i+1)} \cap X^+, A_{\sigma(i+1)} \cap X^-)] \quad (16)$$

where $X^+ \equiv \{i \in X \mid x_i \geq 0\}$, $X^- = X \setminus X^+$ and σ is a permutation on X satisfying $\mid x_{\sigma(1)} \mid \leq \ldots \leq \mid x_{\sigma(n)} \mid$ and $A_{\sigma(i)} \equiv \{\sigma(i),\ldots,\sigma(n)\}$.

BP type and bi-capacity model If a fuzzy rule table c is assigned from a bi-capacity v in the following manner:

$$c(k_1,\ldots,k_n) = v(A,B), \quad (17)$$
$$\text{where } A = \{i \mid k_i = 3\}, B = \{i \mid k_i = 1\},$$

the output y is equal to $C_v(x)$. Table 5 shows the correspondence when $n = 2$.From (17), the extended fuzzy measures are as follows:

$$\mu^{(k_1,\ldots,k_n)}(A) = v(E,F), \forall(k_1,\ldots,k_n), \forall A \in 2^X \quad (18)$$

where $E = \{i \mid k_i = 2\} \cap A$ and $F = \{i \mid k_i = 1\} \cap A$

For the input $x_1,\ldots,x_n$, the selected segment is $(k_1,\ldots,k_n)$, where

$$k_i = \begin{cases} 1 & \text{if } x_k < 0 \\ 2 & \text{otherwise} \end{cases} \quad (19)$$

Since $h^{(k_1,\ldots,k_n)}(i) = \mid x_i \mid$, $\mu^{(k_1,\ldots,k_n)}(\emptyset) = 0$, and from (18) and (19), $\mu^{(k_1,\ldots,k_n)}(A_{\sigma(i)}) = v(A_{\sigma(i)} \cap X^+, A_{\sigma(i)} \cap X^-)$,

$$\begin{aligned} y &= (\mathrm{C}) \int h^{(k_1,\ldots,k_n)} d\mu^{(k_1,\ldots,k_n)} \\ &= \sum_{i=1}^{n} \mid x_{\sigma(i)} \mid [\mu^{(k_1,\ldots,k_n)}(A_{\sigma(i)}) - \mu^{(k_1,\ldots,k_n)}(A_{\sigma(i+1)})] \\ &= C_v(x). \end{aligned}$$

3.6 AV-type

TF-type and BP-type calculations depend on the selection of the base points. Since the output value in the AV-type model is the average value for all base point selections, this output

ISBN: 978-989-95079-6-8

Table 5: Fuzzy rule table by a bi-capacity ($n = 2$)

Input 2 \ 1	Negative	Reference	Positive
Negative	$v(\emptyset, \{1,2\})$	$v(\emptyset, \{2\})$	$v(\{1\}, \{2\})$
Reference	$v(\emptyset, \{1\})$	$v(\emptyset, \emptyset)$	$v(\{1\}, \emptyset)$
Positive	$v(\{2\}, \{1\}))$	$v(\{2\}, \emptyset)$	$v(\{1,2\}, \emptyset)$

does not depend on the base point selection. Therefore, the AV-type method is a neutral evaluation method in terms of the selection of base points.

If $n = 2$, there are four options for selections the base point for each segment. For $S_{(1,1)}$, $(1,1)$, $(2,1)$, $(1,2)$, and $(2,2)$ are the options for selecting the base points. Table 6 lists the integrands and extended fuzzy measures for each base point selection.

Table 6: Integrand and extended fuzzy measure for AV ($S_{(1,1)}$)

	Integrand		μ			
Base Point	$h(1)$	$h(2)$	$\emptyset$	$\{1\}$	$\{2\}$	$\{1,2\}$
$(1,1)$	$p_1^2(x_1)$	$p_2^2(x_2)$	$c(1,1)$	$c(2,1)$	$c(1,2)$	$c(2,2)$
$(2,1)$	$p_1^1(x_1)$	$p_2^2(x_2)$	$c(2,1)$	$c(1,1)$	$c(2,2)$	$c(1,2)$
$(1,2)$	$p_1^2(x_1)$	$p_2^1(x_2)$	$c(1,2)$	$c(2,2)$	$c(1,1)$	$c(2,1)$
$(2,2)$	$p_1^1(x_1)$	$p_2^1(x_2)$	$c(2,2)$	$c(1,2)$	$c(2,1)$	$c(1,1)$

The calculation process when $x_1 = 30$, $x_2 = 40$, and the rule in table 1 is followed is given in Table 7. The output value of AV-type model, 44, is the average value for four selected base points. Figure 8 shows the graph of the AV-type outputs.

Table 7: Calculation process for $x_1 = 30$ and $x_2 = 40$

	Integrand		μ				output
Base Point	$h(1)$	$h(2)$	$\emptyset$	$\{1\}$	$\{2\}$	$\{1,2\}$	y
$(1,1)$	0.6	0.8	0	50	30	60	42
$(2,1)$	0.4	0.8	50	0	60	30	46
$(1,2)$	0.6	0.2	30	60	0	50	46
$(2,2)$	0.4	0.2	60	30	50	0	42
			Average (y)				44

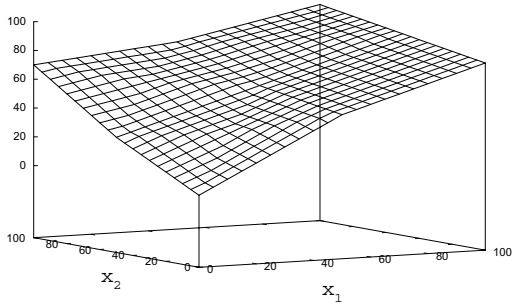


Figure 8: AV-type corresponding to Table 1

4 Properties

Continuity Choquet integral models satisfy the continuity property. Therefore, the proposed models satisfy the continuity property in each segment. For calculating the output values of a point on a bordering hyperplane, there are two or more options for selecting the calculation segment. However, in TF-, BP- and AV-type models, the output values remains the same regardless of the selected segment, that is, the output value on a bordering hyperplane does not depend on the selection of segments (The proof is omitted.) Therefore, Choquet integral based evaluations satisfy the continuity property.

Monotonicity It is known that almost all fuzzy reasoning models do not satisfy the monotonicity property when the fuzzy rule table is monotone. Choquet-integral-based evaluations satisfy the monotonicity property when the fuzzy rule table is monotone. A monotone fuzzy rule table is defined as $c(k_1, \ldots, k_n) \geq c(k'_1, \ldots, k'_n), \forall k_i \geq k'_i$. Examples of monotone rule tables are Table 1 and 3. When the fuzzy rule table is monotone, it is easy to confirm that the values of the extended Choquet integral are monotone and continuous for each segment. From the continuity of the proposed model, it follows that the Choquet-integral-based evaluation models have monotonicity if the fuzzy rule table is monotone.

Piecewise linear If the all membership functions are linear in that case in figure 1, the outputs are piecewise linear. This property is derived from the piecewise linearity of the Choquet integral.

5 Comparisons with Fuzzy Reasoning models

Fuzzy-rule-based global evaluation methods have been developed by using fuzzy reasoning models [10] such as "Min-Max Gravity Method," "Product-Sum Gravity Method," and "Simplified Fuzzy Reasoning." The outputs of the fuzzy rules for the min-max gravity method and product-sum gravity method are fuzzy sets. Therefore, to obtain the output (non fuzzy) value, a defuzzfication (gravity) calculation is performed. However, these calculations are complex. To analyze properties, the simplified fuzzy reasoning is utilized.

Simplified fuzzy reasoning is based on the product-sum method [11]. In the simplified method, it is possible to use min-max calculations instead of product-sum calculations. To demonstrate the properties of min-max and product-sum calculations, we utilize the two types of simplified fuzzy reasoning models.

5.1 Product-sum-type Simplified Fuzzy Reasoning

Simplified fuzzy reasoning is the most popular type of fuzzy reasoning. If the membership function shown in figure 1 and monotone fuzzy rules are assumed, product-sum-type simplified fuzzy reasoning satisfies the monotonicity property.

Product-sum-type simplified fuzzy reasoning can also be perfoermed using the segment division method. If $S_{(1,1)}$ in figure 2 is selected, the output given by the method is

$$\begin{aligned} y = & [1 - p_1^2(x_1)][1 - p_2^2(x_2)]c(1,1) \\ + & p_1^2(x_1) \times [1 - p_2^2(x_2)]c(2,1) \\ + & [1 - p_1^2(x_1)]p_2^2(x_2)c(1,2) + p_1^2(x_1)p_2^2(x_2)c(2,2) \end{aligned} \quad (20)$$

This equation involves an expectation calculation, that is, $p_1^2(x_1)$ and $p_2^2(x_2)$ are random variables, and the expected value of four events ($c(1,1), c(2,1), c(1,2)$, and $c(2,2)$) is calculated by this eqation.

In the segment, if the fuzzy rules are monotone, the outputs satisfy the monotonicity and continuities. Therefore, product-sum methods with monotone fuzzy rules satisfy the monotonicity property.

ISBN: 978-989-95079-6-8

Since product-sum-type calculations involve random variables calculation, output values vary unpredictably. Table 8 lists the AND-type fuzzy rule "If Input 1 and 2 are big, the output is 1." Figure 9 shows the output values for the rule table. Product-sum-type calculations are nonlinear because of product calculations. The input variation from $x_1 = x_2 = 0$ to $x_1 = x_2 = 0.01$ increases the output value by only 0.001, but the variation from $x_1 = x_2 = 0.99$ to $x_1 = x_2 = 1$ increases the output value by 0.0199. These results are inconsistent with the rule.

Table 8: AND-type rule table

Input 2 \ Input 1	Small	Big
Small	0.0	0.0
Big	0.0	1.0

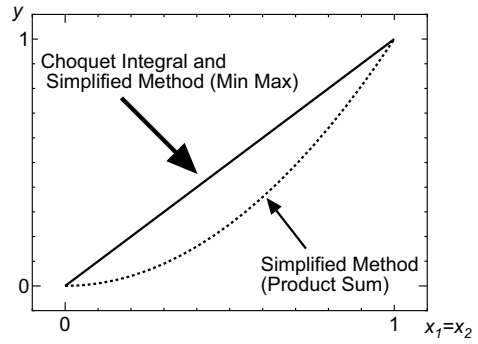


Figure 9: Comparison with product-sum-type method

Since $x_1 = x_2 = 0.5$ only satisfies half of the rule condition, it is expected that the output value is 0.5, but while the product-sum-type is 0.25. Moreover, if the number of inputs for the condition is increased, the output given by the product-sum method decreases. For example, if the number of inputs is 3 or 4, the product-sum output is $0.5^3 = 0.125$ or $0.5^4 = 0.0625$, respectively.

5.2 Min-max type Simplified Fuzzy Reasoning

In ordinal simplified fuzzy reasoning, product-sum calculation is performed, but it is possible to use min and max operators instead of product and sum operators.

Table 9: Monotone fuzzy rule table

Input 2 \ Input 1	Small	Big
Small	0.6 ($c(1,1)$)	0.7 ($c(2,1)$)
Big	0.8 ($c(1,2)$)	1.0 ($c(2,2)$)

Table 9 is represented by the fuzzy switching function

$$y = (c(1,1) \wedge x_1^1 \wedge x_2^1) \vee (c(2,1) \wedge x_1^2 \wedge x_2^1) \vee (c(1,2) \wedge x_1^1 \wedge x_2^2) \vee (c(2,2) \wedge x_1^2 \wedge x_2^2),$$

where $x_1^1 + x_1^2 = 1$ and $x_2^1 + x_2^2 = 1$. The equation does not satisfy the monotonicity property because fuzzy switching functions do not satisfy the complementary laws. For example, if $x_1^2 = 0.5$ and $x_2^2 = 1.0$, $y = 0.5$. However, if $x_1^2 = 0.4$ and $x_2^2 = 1.0$, $y = 0.6$.

Therefore, if fuzzy-value calculations are necessary, it is preferable to use Choquet-integral-based calculations.

6 Conclusions

This model is very useful for applications in social science, such as performance evaluation and educational evaluations. Since social science applications require strict validity of the output values, the global evaluation functions should not exhibit non-monotonicity and unexpected nonlinearity, such as that seen in figure 9. The Choquet integral models can be used to ensure the validity of the outputs. However, the attitude of evaluators might change according to the situation or the input values, and this might change the weights and the relationships among the evaluation parameters. The Choquet integral models keep those validities. Fuzzy rule tables can be used to represent these attitudes. Therefore, the proposed model inherits the merits of both the Choquet integral model and the fuzzy rule table.

References

[1] N. Nakajima, Inference, Reasoning, and Fuzzy Reasoning, *Journal of Japan Society for Fuzzy Theory and Systems*, Vol.6, No.6, 1105-1114, 1994.

[2] Y. Okuda et al, Monotonicity Analysis of Fuzzy Reasoning, *Waseda journal of mathematics education*, Vol.21, No.1, 28-33, 2003.

[3] G. Choquet, Theory of Capacities, Annales de l'Institut Fourier 5, 131-295, 1954.

[4] M. Grabisch et al, Fundamentals of Uncertainty Calculi with Applications to Fuzzy Inference, Kluwer, 1995.

[5] E.Takahagi, Fuzzy Integral Based Fuzzy Switching Functions, in *Transactions on Rough Sets II. Lecture Notes in Computer Science 3135 edited by James F. Peters et al*, pp.129-150, Springer, 2004.

[6] E. Takahagi, Fuzzy three-valued switching functions using Choquet integral, *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol.7 No.1, 47-52, 2003.

[7] A. Tversky and D. Kahneman, Advances in Prospect Theory: Cumulative Representation of Uncertainty, *J. of Risk and Uncertainty*, 5, 297-323, 1992.

[8] M. Grabisch and Ch. Labreuche, Bi-capacities — Part I: definition, Möbius transform and interaction, *Fuzzy Sets and Systems*, 151, 211-236, 2005.

[9] M. Grabisch and Ch. Labreuche, Bi-capacities — Part II: the Choquet integral, *Fuzzy Sets and Systems*, 151, 237-259, 2005.

[10] E.H. Mamdani, Applications of fuzzy algorithm for control of a simple dynamic plant, *Proc. of IEEE* ,121,1585-1588, 1974.

[11] M. Mizumoto, Product-Sum-Gravity Method = Fuzzy Singleton-type Reasoning Method = Simplified Fuzzy Reasoning Method, *Proc. of Fifth IEEE International Conference on Fuzzy Systems*, 2098-2102, 1996.

ISBN: 978-989-95079-6-8

Fuzzy modelling of sensor data for the estimation of an origin-destination matrix

Krystyna Biletska[124] Sophie Midenet[4] Marie-Hélène Masson[23] Thierry Denœux[12]

1. Université de Technologie de Compiègne
2. UMR CNRS 6599 Heudiasyc,
BP 20529, 60205 Compiègne cedex, France
3. Université de Picardie Jules Verne, IUT de l'Oise
4. Université Paris-Est, Institut National de Recherche sur les Transports et leur Sécurité (INRETS), Laboratoire GRETIA
2 rue de la Butte Verte, 93166 Noisy le Grand cedex, France
Email: krystyna.biletska@inrets.fr (corresponding author)

Abstract— *This paper examines a short-time estimation problem of an origin-destination (OD) matrix, where each element is a volume of vehicle flow between one of the OD pair of zones of a signalised junction. The estimation is based on the use of traffic measurements provided by video sensors and on the knowledge of the traffic lights. This data is subject to redundancy, imprecision and uncertainty. The main purpose of this paper is to obtain the best estimates of the OD matrix by modelling the data imperfection, using a two-step method. First, relationships between the observed data are built in real-time using High-Level Petri Nets. Due to the imperfection of data the system obtained is underdetermined and inconsistent. Second, the fuzzy sets theory is used to model this imperfection and to overcome the inconsistency of the system.*

Keywords— Fuzzy least squares, Fuzzy linear programming, Fuzzy modelling, Origin-destination matrix

1 Introduction

Knowledge about origins and destinations (OD) of vehicles in a road network is important in most transport systems. In the case of a junction, its mathematical representation is OD matrix $\mathbf{B}$, each element b_{ij} of which is a proportion of the flow of vehicles that come from entrance (origin) i and go to exit (destination) j. Such a proportion is called the *OD flow rate*. Since the OD matrix changes in time following the changes in traffic demand, the estimation period of the OD matrix has to be as short as possible. In particular, the period should be equal to a traffic light cycle (duration of green-amber-red sequence) when we deal with a signalised junction. At INRETS/GRETIA we use such a short-time estimation of the OD matrix as part of a diagnostic system for signalised junctions [1]. This system compares the impacts of different traffic control strategies expressed in terms of CO_2 and pollutant emissions.

The OD matrix is generally deduced from vehicle counts made on each entrance and each exit of the junction during a given time interval. These counts are usually provided by magnetic loops embedded in the road surfaces and sensitive to metallic masses. The estimation can be obtained from a conservation law of vehicles which is a set of relationships between exit and entrance flow counts. In general, when loops are installed on every entrance and exit of the junction, the estimation problem is underdetermined. Thus a solution is not unique and additional information such as a *prior* OD matrix is used to choose the OD matrix which corresponds best to the actual matrix. Some of the existing methods are based on the information minimisation principle [2], on maximisation of likelihood [3] or on Bayesian inference [4]. Other methods propose a recursive estimation of the OD matrix [5, 6]. Furthermore, the estimation problem can be represented as a constraint optimisation problem [7, 8]. Methods which use traffic lights are described in [9, 10].

This paper considers the problem of reconstituting the origins and destinations of vehicle flows crossing a signalised junction, at each traffic light cycle. OD flow volumes are estimated using traffic lights and traffic measurements from video cameras installed at the junction. These measurements, provided every second, are the vehicle counts made on each entrance and exit of the junction and the number of vehicles stopped at each inner section of the junction. The data is subject to redundancy, uncertainty and inaccuracy. Such an imperfection is linked to the reliability of video sensors, measurement conditions, traffic characteristics and drivers' behaviour.

In order to take into account all available information it is necessary to consider the nature and possible interdependence of the data. None of the cited methods takes into account the lake of imprecision of vehicle counts, the possible physical complexity of the junction and the traffic lights at the same time. Moreover they cannot be applied to the problem because the period of estimation is quite short.

The main purpose of this paper is to obtain the best estimates of the OD matrix by modelling the data imperfection, using a two-stage method. First, a conservation law of vehicles, which is represented by an inconsistent and underdetermined system of equations, is built by High-Level Petri Nets at each traffic light cycle. Second, the fuzzy sets theory is used to overcome the inconsistency of the system and to model the data imperfection. Three different approaches have been analysed to solve this system of equations: ordinary least squares, fuzzy least squares and fuzzy linear programming. A numerical study of the proposed methods has been done using the data collected in a real experimental junction fitted out with video cameras and a traffic light controller.

The rest of the paper is organised as follows. Section 2

ISBN: 978-989-95079-6-8

presents the problem and describes the real data. Sections 3 and 4 introduce the fundamental principles of High-Level Petri Nets and show how they are applied to model the OD flows through the experimental site. Section 5 proposes three methods for OD matrix estimation and the paper concludes by proposing further lines of research.

2 The problem

The experimental site is an isolated signalised junction of two double-lane roads situated in the south suburb of Paris (Fig. 1). The main road B-D which connects the suburbs to Paris has a high traffic volume, whereas the road A-C has lower traffic volume. Traffic lights control four incoming links and four inner zones. Note that right-turning vehicles use special lanes and are not taken into account in this study. Only eight OD flows are statistically significant: $AC, AD, BD, BB, CB, DB, DC, DD$.

Eight video cameras are installed at the junction in order to capture all the entrance and exit links and the inner zones. The location, height and angle of each camera depend on the geometry of the junction and are chosen to favour the measurement of space traffic parameters such as queue length on incoming links. The camera views are analysed in real time using image processing techniques developed at INRETS [11]. They provide several measurements every second:

- $\mathcal{X}_i(\tau)$ vehicle counts measured at the end of an entrance i at second τ ($i = 1, \ldots, n$),
- $\mathcal{Y}_j(\tau)$ number of vehicles that have passed through the beginning of the exit j at second τ ($j = 1, \ldots, m$),
- $\mathcal{Z}_k(\tau)$ number of stopped vehicles at inner zone k at second τ ($k = 1, \ldots, p$),

where n, m and p are the numbers of entrances, exits and inner zones respectively. Here, a *traffic light cycle* is a period of time between two sequential onsets of the red light on the main road B-D .

Many phenomena influence the quality measurement of traffic parameters. Traffic conditions (peak or off-peak periods) are the reason for many traffic count errors. If the traffic flow is heavy, the gaps between vehicles are small and it is difficult to distinguish these gaps on the video images. Thus the number of vehicles measured is lower than the actual number.

The characteristics of vehicles are also a source of measurement errors. High vehicles passing in front of camera will hide the smaller vehicles or the whole camera field, i.e. they will produce a *masking effect*. Two-wheeled vehicles are only seldom counted because they are small. The heterogeneous colours of vehicle roofs also add to the problem of detection.

Meteorological conditions inevitably have an influence on all types of traffic measurements and the video are blurred: the wind shakes the posts the cameras are fixed to, the sun's rays cause the reflections on the vehicle surfaces and camera lenses, rain, snow and fog obscure a camera field. Changes in brightness caused by the position of the sun, clouds and headlights at night also determine the reliability of the measurements.

Let $x_i(c)$ be the flow volume at entrance i during a traffic light cycle c and $y_j(c)$ be the flow volume which entered the junction during cycle c and leaves it by exit j. *OD flow rate*

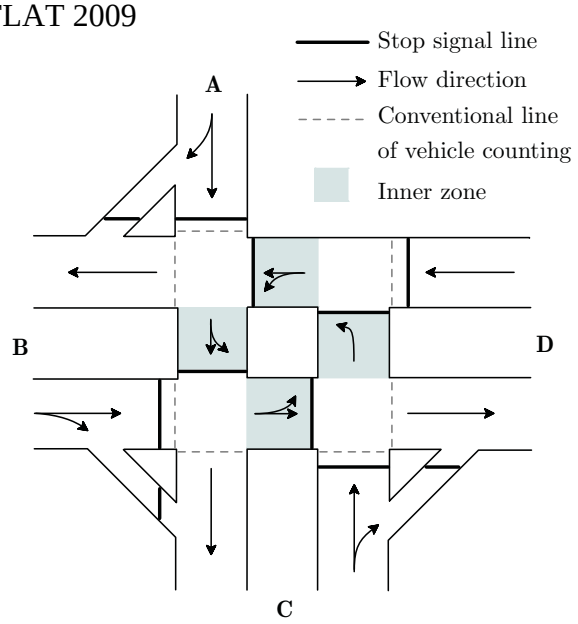


Figure 1: The experimental junction

b_{ij} is the proportion of the flow of vehicles that come from entrance i and go to exit j. The problem is to estimate OD flow rates b_{ij} ($\forall i \in [\![1, n]\!], \forall j \in [\![1, m]\!]$) at the end of each traffic light cycle c, such that

$$y_j(c) = \sum_{i=1}^{n} b_{ij}(c) x_i(c), \tag{1a}$$

$$x_i(c) b_{ij}(c) \geq z_{kij}(c) \quad \forall k \text{ s.t. } \delta_{kij} = 1, \tag{1b}$$

$$\sum_{j=1}^{m} b_{ij}(c) = 1, \tag{1c}$$

$$b_{ij}(c) \geq 0, \tag{1d}$$

where $z_{kij}(c)$ is the number of vehicles which cross the junction from i to j and stop at inner zone k during cycle c, $\delta_{kij} = 1$ if OD flow from i to j can pass through inner zone k and is 0 otherwise. For a given cycle c the value of variable $x_i(c)$ can be obtained from instantaneous vehicle counts

$$x_i(c) = \sum_{\tau=1}^{\mathcal{G}_i(c)} \mathcal{X}_i(\tau),$$

where $\mathcal{G}_i(c)$ is a duration of the green light of cycle c in entrance i. The values of $y_j(c)$ and $z_{kij}(c)$ cannot be obtained directly from traffic measurements, because it is impossible to know the period of time when the vehicle flow $x_i(c)$ leaves the junction or stops at inner zones.

This paper proposes a method to obtain the values of $y_j(c)$ and $z_{kij}(c)$ from $\mathcal{Y}_j(\tau)$ and $\mathcal{Z}_k(\tau)$ respectively and then to estimate the OD flow rates. First, a tool for vehicle flow segmentation is built via two High-Level Petri Nets (HLPN). The first net indicates the set of vehicle flows which may be present in each zone at any given second. It makes possible to associate these flows to the measurements taken in the corresponding zones and provides the beginnings of the flows. The second net provides the ends of the flows. For a given cycle we thus know the duration of the presence of the flows in each zone

ISBN: 978-989-95079-6-8

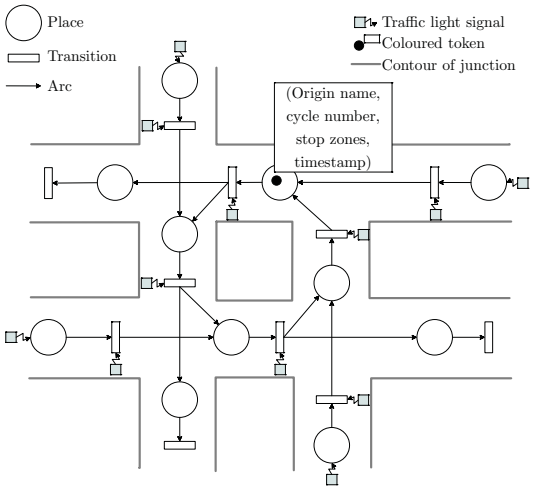


Figure 2: High Level Petri Net 1

and can collect the corresponding measurements. A consistent underdetermined system of equations, whose unknowns are the OD flow rates, can thus be built dynamically and solved at each traffic light cycle.

3 Fundamentals of High-Level Petri Nets

A High-Level Petri Net (HLPN) is a graphical and mathematical tool used to model discrete-event systems [12, 13]. In this paper, the HLPN is a classical Petri Net [14] complemented by the notions of time and colour. In graphic representation, an HLPN is a directed graph composed of two kinds of nodes: *places*, drawn by circles, and *transitions* represented as bars. The arcs connect either a place to a transition or, inversely, a transition to a place. A place may be *marked* by coloured *tokens* endowed with timestamps and characterises the system state. A transition represent an event that can change the state of the system modelled. The set of input and output places of a transition is also interpreted as a set of pre- and post-conditions of an event. When the event occurs the transition is *enabled* and the marking of input and output places changes. Thereby, the dynamic behaviour of the system is expressed by means of time-varying marking.

A 7-tuple $N = (\Sigma, \mathcal{P}, \mathcal{T}, \mathcal{A}, \mathcal{C}, \mathcal{W}, \mathcal{F})$ is a formal representation of HLPN, where:

- Σ is a finite non-empty colour set,
- $\mathcal{P}$ and $\mathcal{T}$ are the finite sets of places and transitions respectively, such that $\mathcal{P} \cap \mathcal{T} = \emptyset$ and $\mathcal{P} \cup \mathcal{T} \neq \emptyset$,
- $\mathcal{A} \subset (\mathcal{P} \times \mathcal{T}) \cup (\mathcal{T} \times \mathcal{P})$ is the finite set of arcs,
- $\mathcal{C} : \mathcal{P} \rightarrow 2^{\Sigma}$ is a *colour function* which associates a set of colours to a place,
- $\mathcal{W} : \mathcal{A} \rightarrow 2^{\Sigma}$ is an *arc function*.
- $\mathcal{F}$ is a delay associated to an arc.

A token is represented by a triple (p, s, θ), where $p \in \mathcal{P}$ is the place marked by the token, $s \subseteq \mathcal{C}(p)$ is a colour set which describes the system state in p, called *colour*, and θ is the time at which the token becomes available and the system state changes, called *timestamp*. Marking $\mathcal{M}(p, \tau)$ of place $p \in \mathcal{P}$ at time τ is a multiset represented by the pair $(S, \mathbf{n})$ where S is the set of colours of tokens arrived in p at $\tau = \theta$ and $\mathbf{n}$ is the vector every element of which is the number of occurrences of the token's colour s in set S.

Function $\mathcal{W}(p, t)$, corresponding to the arc from output place p to t, is a multiset of colours that are elements of the set $\mathcal{C}(p)$. The function $\mathcal{W}(t, p)$ is defined similarly.

Transition $t \in \mathcal{T}$ is enabled if $\mathcal{M}(p, \tau) \supseteq \mathcal{W}(p, t)$. It removes $\mathcal{W}(p, t)$ tokens from each of its input places p of t and adds $\mathcal{W}(t, p)$ tokens to its output places p with delay $\mathcal{F}(t, p)$ associated to an arc from t to p.

4 Modelling the traffic flows crossing a signalised junction

In this paper two HLPNs are used to model and segment the OD flows of vehicles crossing the junction in order to built the vehicle conservation law in a dynamic way. In the first net, a place stands for a zone of the experimental junction (Fig. 2). A token in a place is represented by a colour set $s = (s_1, s_2, s_3)$, where $s1 \in \{A, B, C, D\}$ is the name of flow origin, $s2 \in \mathbb{N}$ is a cycle number and $s3$ is a set of stop zones. In addition, it has a timestamp θ. A token in place p means the possible presence of a flow of origin $s1$ in the zone corresponding to p. Some transitions are related to the change of traffic lights in entrances and inner zones, other transitions represent the flow departures. An arc function $\mathcal{W}(t, p)$ is defined over a set of colours $s1$ representing the origins of the flows that pass through this arc. Delay $\mathcal{F}(t, p)$ is an average vehicle travel time between two successive zones of the junction corresponding to the input and output places of transition t.

The rules for changing colours of tokens, enabling transitions and marking places are defined as follows. During the green light at entrance i of the junction, the corresponding place is marked by one token with colour $s = (s1, s2, s3)$, $s_3 = \emptyset$, every second. Transition t associated to entrance i becomes enabled and the token is transmitted to output places of t. The transition associated with an inner traffic light is enabled if its input place contains at least one token and the corresponding traffic light is green. The enabled transition transmits the token downstream according to $\mathcal{W}(t, p)$ and $\mathcal{F}(t, p)$. If the light in inner zone k is red, tokens are stacked in the corresponding place. Their colours are changed by adding the name of inner zone k to the colour set s_3. At the beginning of the green light these tokens are shifted to the exit place p with predefined delay $\mathcal{F}(t, p)$. The transition related to the exit of junction is enabled when each of its input places contains at least one token.

Marking $\mathcal{M}(p, \tau)$ of place p characterizes a set of entrance flows that can be present at time τ in a junction zone corresponding to p. When the marking of p does not change during certain period of time $\Delta\tau$, all measurements taken at the corresponding zone during $\Delta\tau$ are associated to the set of flows. Thus, for a given cycle c, each exit flow $y_j(c)$ can be seg-

ISBN: 978-989-95079-6-8

mented into L_j vehicle platoons, such that

$$y_j^l(c) = \sum_{\tau \in \Delta\tau_l(c)} \mathcal{Y}_j(\tau) \qquad \forall l \in [\![1, L_j(c)]\!],$$

$$z_{kij}^l(c) = \max_{\tau \in \mathcal{R}_k(c)} \mathcal{Z}_k(\tau)\delta_{kij} \quad \begin{array}{l} \forall l \in [\![1, L_j(c)]\!], \\ \forall i \in \mathcal{M}(p, \Delta\tau_l), \\ \forall k \text{ s.t. } \delta_{kij} = 1, \end{array}$$

where $\Delta\tau_l(c)$ is a period of time during which platoon l has left the junction, $z_{kij}^l(c)$ is a maximum number of vehicles that belong to platoon l and have stopped at inner zone k, $\mathcal{R}_k(c)$ is a duration of red light in inner zone k, δ_{kij} indicates if zone k is situated between OD pair of zones (i, j), $\mathcal{M}(p, \Delta\tau_l)$ is a marking of place $p \in \mathcal{P}$ corresponding to exit j. The number of platoons $L_j(c)$ depends on the number of entrances of the junction, and on traffic volume and traffic light command.

Note that the first occurrence of token $(p, \{s1, s2, s3\}, \theta)$ denotes that the beginning of the flow with origin $s1$ crosses the zone corresponding to place p. The ends of the flows are provided by the second HLPN, which has the identical topology and the same representation of places and transitions as the first HLPN. However, the meaning of the tokens is different. A token stands for the end of a flow and is represented by a set of two colours without a timestamp: the name of flow origin $s1 \in \{A, B, C, D\}$ and a cycle number $s2 \in \mathbb{N}$.

The onsets of the beginning and the end of a flow in place p allow us to determine the duration of the flow presence in zone related to p. Thus we can collect the measurements made in this zone and, at each cycle c, can built the following set of equations

$$y_j^l = \sum_{i \in \mathcal{M}(p_j, \Delta\tau_l)} b_{ij}^l x_i \qquad \forall j \in [\![1, m]\!], \forall l \in [\![1, L_j]\!], \quad (2a)$$

$$x_i b_{ij}^l \geq z_{kij}^l \quad \begin{array}{l} \forall j \in [\![1, m]\!], \forall l \in [\![1, L_j]\!], \\ \forall i \in \mathcal{M}(p_j, \Delta\tau_l), \\ \forall k \text{ s.t. } \delta_{kij} = 1, \end{array}, \quad (2b)$$

$$\sum_{j,l} b_{ij}^l = 1 \qquad \forall i \in [\![1, n]\!], \quad (2c)$$

$$b_{ij}^l \geq 0 \quad \begin{array}{l} \forall j \in [\![1, m]\!], \forall l \in [\![1, L_j]\!], \\ i \in \mathcal{M}(p_j, \Delta\tau_l), \end{array} \quad (2d)$$

where the cycle c has been omitted to simplify the notations.

Let J and K be the numbers of constraints (2a) and (2b) respectively, and I be the number of unknowns b_{ij}^l, such that

$$\begin{aligned} J &= \sum_{j=1}^{m} jL_j, \\ K &= \sum_{j=1}^{m} \sum_{l=1}^{L_j} \sum_{i \in \mathcal{M}(p_j, \tau_l)} \sum_{k=1}^{p} \delta_{kij}, \\ I &= \sum_{j=1}^{m} \sum_{l=1}^{L_j} \sum_{i \in \mathcal{M}(p_j, \tau_l)} i. \end{aligned}$$

Since in general $I \geq J$ (in this paper $I \approx 10$), the obtained system of equations (2a) is underdetermined.

5 OD matrix estimation

The estimation problem of OD flow rates b_{ij}^l from the relationship (2a) can be viewed as an estimation of coefficients of the regression model

$$\mathbf{y} = \mathbf{X_1 b}, \quad (3)$$

where $\mathbf{y}$ is a vector composed of J output variables y_j^l, $\mathbf{b}$ is a vector containing the I unknowns b_{ij}^l, $\mathbf{X_1}$ is an $J \times I$ matrix rearranged in such a way that (2a) corresponds to (3).

The elements of model (3) must satisfy the constraints:

$$\mathbf{X_2 b} \geq \mathbf{z}, \quad (4a)$$

$$\mathbf{I b} = \mathbb{1}, \quad (4b)$$

$$\mathbf{b} \geq 0, \quad (4c)$$

where $\mathbf{z}$ is a vector composed of K variables z_{ijk}^l, $\mathbf{X_2}$ is an $K \times I$ matrix built so that (2b) is equal to (4a), $\mathbf{I}$ is an indicator $n \times I$ matrix organised in such a manner that (4b) is equivalent to (2c) and $\mathbb{1}$ is an identity n-vector.

Considering $\mathbf{b}$ to be an m-vector of crisp elements, three situations are studied in this section of the paper:

A) values of $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{y}$ and $\mathbf{z}$ are crisp

B) values of $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{y}$ and $\mathbf{z}$ are fuzzy

C) values of $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{y}$ and $\mathbf{z}$ are crisp, relationships (3) and (4a) are fuzzy, where the strict mathematical relations are replaced by fuzzy equivalents $\lessapprox$ and $\gtrapprox$ that are read "essentially smaller or equal to" and "essentially greater than or equal to".

In the first case, we use the least squares method to estimate the crisp regression coefficients $\mathbf{b}$. The fuzzy least squares (FLS) model based on commonly used Diamond ρ_2-metric [15] has been applied to the second situation. For the last case, we formulate a fuzzy linear programming problem (FLP) following the approach proposed by Zimmernann [16]. We present in the sequel the three methods and their results on our application.

5.1 Least squares method

Let the input and output variables $\mathbf{X}_1$ and $\mathbf{y}$ of the regression model (3) be non-negative crisp numbers. The unknown coefficients $\mathbf{b}$ of the model and the elements of $\mathbf{X}_2$ and $\mathbf{z}$ are also considered to be crisp. A first approach to estimate $\mathbf{b}$ is to use a least squares method. To insure the existence of feasible solutions, we introduce slack variables, e_i ($\forall i = 1, \ldots, K$), and we propose to solve the following problem:

$$\min_{\mathbf{b},\mathbf{e}} \| \mathbf{y} - \mathbf{X_1 b} \|^2 + \sum_{i=1}^{K} e_i \quad (5)$$

subject to

$$\begin{cases} \mathbf{X_2 b} + \mathbf{e} \geq \mathbf{z}, \\ \mathbf{I b} = \mathbb{1}, \\ \mathbf{b} \geq 0, \\ \mathbf{e} \geq 0. \end{cases} \quad (6)$$

5.2 Fuzzy least squares method

To take into account model errors and the inherent imprecision of data measurements, we choose to represent the input and output variables of the regression model (3) by triangular fuzzy numbers $\tilde{\mathbf{X}}_1 = (\underline{\mathbf{X}}_1, \mathbf{X}_1^m, \overline{\mathbf{X}}_1)$, $\tilde{\mathbf{y}} = (\underline{\mathbf{y}}, \mathbf{y}^m, \overline{\mathbf{y}})$ and

ISBN: 978-989-95079-6-8

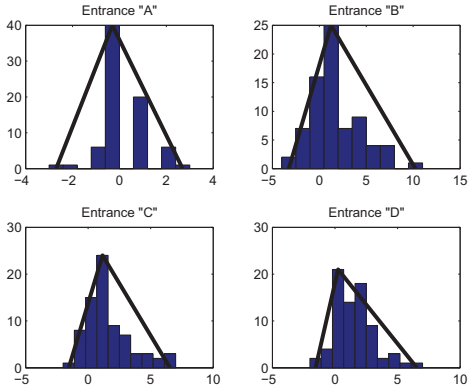


Figure 3: Empirical distributions of error counts for the entrances of the junction

$\tilde{\mathbf{z}} = (\underline{\mathbf{z}}, \mathbf{z}^m, \overline{\mathbf{z}})$, where $[\underline{a}, \overline{a}]$ is the support and a^m is the mode of fuzzy number $\tilde{a}$. The form of fuzzy numbers $\tilde{\mathbf{X}}_1$ was derived from empirical distributions of the error counts as shown in Fig. 3. In most cases, counts of vehicles at the entrances are smaller than the true value of $\mathbf{X}_1$. The fuzzy number $\tilde{\mathbf{X}}_2$ is determined similarly to $\tilde{\mathbf{X}}_1$. Since there is no available histograms of error counts for $\mathbf{y}$ and $\mathbf{z}$, the fuzzy numbers $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{z}}$ are supposed to be symmetrical with the spreads chosen experimentally. Note that $\underline{\mathbf{X}}_1, \underline{\mathbf{X}}_2, \underline{\mathbf{y}}, \underline{\mathbf{z}} \in \mathbb{R}^+$ because of the nature of the data.

According to the fuzzy least squares (FLS) method proposed by Diamond [15, 17] and supposing the unknown coefficients $\mathbf{b}$ to be crisp, the following minimisation problem is written:

$$\min_{\mathbf{b}} \left(\| \underline{\mathbf{y}} - \underline{\mathbf{X}}_\mathbf{1}\mathbf{b} \|^2 + \| \mathbf{y}^m - \mathbf{X}_1^m\mathbf{b} \|^2 + \right. \\ \left. + \| \overline{\mathbf{y}} - \overline{\mathbf{X}}_\mathbf{1}\mathbf{b} \|^2 \right) \quad (7)$$

subject to

$$\begin{cases} \overline{\mathbf{X}}_\mathbf{2}\mathbf{b} \geq \underline{\mathbf{z}}, \\ \mathbf{I}\mathbf{b} = \mathbb{1}, \\ \mathbf{b} \geq 0. \end{cases} \quad (8)$$

Note that the first constraint in (8) is the least conservative translation of the strong relation (4a).

5.3 Fuzzy linear programming approach

Another approach to estimate the OD flow rates is to model the constraints (3) and (4a) by fuzzy sets. Assuming that the values of $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{y}$ and $\mathbf{z}$ are strict numbers, we consider the estimation problem of $\mathbf{b}$ that satisfies:

$$\mathbf{X_1 b} \gtrsim \mathbf{y} \quad -\mathbf{X_1 b} \gtrsim -\mathbf{y}, \quad (9a)$$
$$\mathbf{X_2 b} \gtrsim \mathbf{z}, \quad (9b)$$
$$\mathbf{I b} = \mathbb{1}, \quad (9c)$$
$$\mathbf{b} \geq 0, \quad (9d)$$

where the sign $\gtrsim$ means that we accept a small violation of the strict relation $\geq$ in a sense that is described below..

Denoting

$$\mathbf{H} = (-\mathbf{X}_1, \mathbf{X}_1, -\mathbf{X}_2)^\top,$$
$$\mathbf{h} = (-\mathbf{y}, \mathbf{y}, -\mathbf{z})^\top,$$

we write (9a)-(9b) as

$$\mathbf{H b} \lesssim \mathbf{h} \quad (10)$$

Each row i of (10) is assumed to be a fuzzy set with a membership function $\mu_i(\mathbf{b})$:

$$\mu_i(\mathbf{b}) = \begin{cases} 1 & H_i\mathbf{b} \leq h_i \\ 1 - \frac{H_i\mathbf{b} - h_i}{\xi_i} & h_i < H_i\mathbf{b} \leq h_i + \xi_i \\ 0 & H_i\mathbf{b} > h_i + \xi_i \end{cases} \quad (11)$$

where each $\xi_i > 0$ is a given constant reflecting an acceptable degree of constraint violation, $i = 1, \ldots, 2J + K$. Here $\mu_i(\mathbf{b})$ is defined as the degree to which $\mathbf{b}$ satisfies the relationship i.

With respect to Bellman-Zadeh decision-making rule for fuzzy sets [18], the membership function of the fuzzy decision set $\tilde{B}$ is defined as follows

$$\mu_{\tilde{B}}(\mathbf{b}) = \min_i \left(\mu_i(\mathbf{b})\right) = \min_i \left(1 - \frac{H_i\mathbf{b} - h_i}{\xi_i}\right), \\ \forall i = 1, \ldots, 2J + K. \quad (12)$$

According to the symmetric FLP method proposed by Zimmermann [16, 19], the crisp optimal solution can be given by

$$\max_{\mathbf{b} \geq 0, \mathbf{I}\mathbf{b} = \mathbb{1}} \mu_{\tilde{B}}(\mathbf{b}) \quad (13)$$

which can be obtained by solving the following problem of linear programming:

$$\max_{\lambda, \mathbf{b}} \lambda \quad (14)$$

subject to

$$\begin{cases} \lambda < 1 - \frac{H_i\mathbf{b} - h_i}{\xi_i} & \forall i = 1, \ldots, 2J + K, \\ \mathbf{I}\mathbf{b} = \mathbb{1}, \\ \mathbf{b} \geq 0. \end{cases} \quad (15)$$

5.4 Results

These methods have been tested using the *raw data* collected at the experimental junction. The estimation is made on 25 consecutive traffic light cycles, equivalent to 30 minutes. In addition, the actual values of OD flow rates b_i^* ($\forall i = 1, \ldots, 8$), derived manually from video images, are available for all cycles.

Since the measure of the number of stopped vehicles is more accurate than the measure of exit flow volume, the following parameters of the FLS method where chosen: $\mathbf{y}^m - \underline{\mathbf{y}} = \overline{\mathbf{y}} - \mathbf{y}^m = 3$, $\mathbf{z}^m - \underline{\mathbf{z}} = \overline{\mathbf{z}} - \mathbf{z}^m = 2$. The acceptable degree of constraint violation ξ_i ($\forall i = 1, \ldots, 2J + K$) in FLP method has been fixed to 5 for constraints (9a) and to 3 for constraints (9b).

The estimation error has been calculated for the OD flow rates (Fig. 4): $E = \mathbf{b}\mathbf{X}^* - \mathbf{b}^*\mathbf{X}^*$, where $\mathbf{X}^*$ is a vector of actual vehicle counts at the entrances of the experimental junction. The best results are obtained with the FLS method for which the median error is almost zero for all OD flows. The results of the least squares method are less accurate than those of the FLS method. Surprisingly, the FLP method, although it seems to be a more natural fuzzy translation of the crisp initial problem, does not provide very good results. Note that the estimation error E of all methods is higher, if the flow volume is lower, like for the OD flows "AD" and "BB".

ISBN: 978-989-95079-6-8

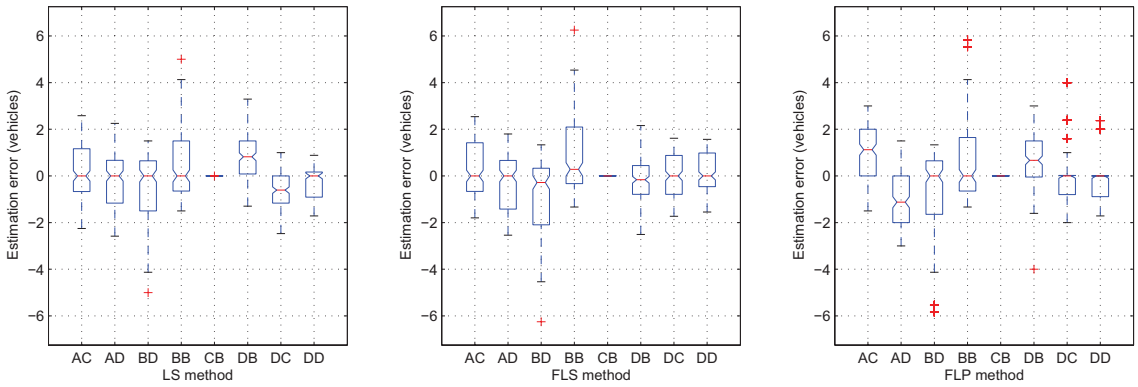


Figure 4: Estimation errors for 8 OD flow rates

6 Conclusion and future lines of research

In this paper we have proposed a new short-time estimation method of the OD matrix for a signalised junction. We have built a model of the segmentation of traffic flows crossing the junction in order to draw up a vehicle conservation law, represented by a underdetermined system of equations, for each traffic light cycle. The model, made using High-Level Petri Nets, is based on the use of traffic lights.

Real traffic measurements collected at the real experimental junction are used to estimate the OD flow rates. The experimental comparison shows that a small gain is obtained when the data imprecision and uncertainty is taken into account using a fuzzy modelling. Among the two proposed fuzzy approaches, the fuzzy least squares yields the most accurate results. In spite of the important imprecision of the real data, the first results are very encouraging to continue the research on improvement of the estimation accuracy.

The main future line of research is to fuzzify the token timestamps of the first HLPN in order to model the temporal imprecision of data measurements. The application of our method should be also extended to a sequence of junctions.

References

[1] S. Midenet *et al.* Signalized intersection with real-time adaptive control: on-field assessment of CO_2 and pollutant emission reduction. *Transportation Research Part D*, 9:29–47, 2004.

[2] H. J. Van Zuylen. The estimation of turning flows on a junction. *Traffic Engineering Control*, 20(11):539–541, 1979.

[3] E. Hauer, E. Pagitsas, and B. T. Shin. Estimation of turning flows from automatic counts. *Transportation Research Record*, (795):1–7, 1981.

[4] M. J. Maher. Inferences on trip matrices from observations on link volumes: a Bayesian statistical approach. *Transportation Research Part B*, 17(6):435–447, 1983.

[5] M. Cremer and H. Keller. A new class of dynamic methods for the identification of origin-destination flows. *Transportation Research Part B*, 21(2):117–132, 1987.

[6] N. L. Nihan and G. A. Davis. Recursive estimation of origin-destination matrices from input/output counts. *Transportation Research Part B*, 21(2):149–163, 1987.

[7] B. Li and B. De Moor. Recursive estimation based on the equality-constrained optimization for intersection origin-destination matrices. *Transportation Research Part B*, 33(3):203–214, 1999.

[8] H. D. Sherali *et al.* Parameter optimization methods for estimating dynamic origin-destination trip-tables. *Transportation Research Part B*, 31(2):141–157, 1997.

[9] H. Keller and G. Ploss. Real-time identification of O-D network flows from counts for urban traffic control. In N.H.Gartner and N.H.M.Wilson, editors, *Proceedings of the 10th International Symposium on Transportation and Traffic Theory*, pages 267–289. Elsevier Science Publishing Co., 1987.

[10] I. Matschke and B. Friedrich. Dynamic OD estimation using additional information from traffic signal lights timing. *TRISTAN IV Conference*, 2001.

[11] D. Aubert *et al.* Automatic vehicle queue measurement at intersections using image-processing. *8th IEE International Conference on Road Traffic Monitoring and Control, London*, 422:100–104, 1996.

[12] M. Dotoli and M. P. Fanti. An urban traffic network model via coloured timed Petri nets. *Control Engineering Practice*, 14:1213 – 1229, 2006.

[13] K. Jensen *et al.* Coloured Petri nets and CPN tools for modelling and validation of concurrent systems. In *International Journal on Software Tools for Technology Transfer*, pages 98–132. Springer Verlag, 1998.

[14] T. Murata. Petri nets: properties, analysis and applications. In *Proceedings of the IEEE*, volume 77, pages 541–580, April 1989.

[15] P. Diamond and R. Korner. Extended fuzzy linear models and least squares estimates. *Computers and Mathematics with Applications*, 33(9):15–32, May 1997.

[16] H.-J. Zimmermann. Description and optimization of fuzzy systems. *International Journal of General Systems*, 2:209–215, 1976.

[17] R. Slowinski, editor. *Fuzzy sets in decision analysis, operation research and statistics*. The handbooks of fuzzy sets. Kluwer Academic Publishers, 1998.

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

[19] H.-J. Zimmermann, editor. *Fuzzy set theory and its applications*. Kluwer Academic Publishers, second edition, 1991.

ISBN: 978-989-95079-6-8

Classification with reject options in a logical framework: a fuzzy residual implication approach

Hoel Le Capitaine Carl Frélicot

MIA Laboratory, University of La Rochelle
La Rochelle, France
Emails: {hoel.le_capitaine, carl.frelicot}@univ-lr.fr

Abstract— *In many classification problems, overlapping classes and outliers can significantly decrease a classifier performance. In this paper, we introduce the possibility of a given classifier to reject patterns either for ambiguity or for distance. From a set of typicality degrees for a pattern to be classified, we propose to use fuzzy implications to quantify the similarity of the degrees. A class-selective scheme based on this new family is presented, and experimental results showing the efficiency of the proposed algorithm are given.*

Keywords— Ambiguity measures, Dombi and Hamacher implications, fuzzy r-implications, pattern classification, reject options

1 Introduction

Usual classification of objects, or patterns, consists in assigning incoming samples to one class belonging to a set of known ones. Since classes may overlap in the feature space and all of them do not appear during the learning step of most real classification problems, some samples may have to be not classified but rejected. In practice, two situations can lead to rejection, either the incoming sample could belong to unknown class(es) (*distance rejection*) or it should be assigned to several classes (*ambiguity rejection*) [1]. In this paper, we focus our attention on this latter situation. Initially introduced by Chow [2], it consists in rejecting the most unreliable patterns that have a low posterior probability, or more generally a low membership degree value, to its closest class. Such a reject classification rule is optimal if posterior probabilities are estimated without error, unfortunately this is not the case in practice.

We investigate some new measures that aim at quantifying to what extent such a pattern is ambiguous from its membership degrees, or more generally from its multi-class classifier soft outputs, using fuzzy implications. These measures assess the relationships between membership values with various combination of triangular norm operators (t-norms for short, see [3] for an overview).

The remainder of this paper is organized as follows. In section 2, we start by presenting distance-based class models that we will use in the sequel, and then discuss about standard approaches to pattern rejection. Next, we give a brief recall on some basic definitions of triangular norms and fuzzy residual implications, as well as some new parametrical implications (section 3). Section 4 is devoted to the proposal of using implications as ambiguity measures. In section 5, an experimental evaluation and comparisons to well known class-selective pattern recognition schemes are provided. We finally conclude and give some perspectives in section 6.

2 Classification with reject options

2.1 Classification

Let $\mathbf{x}$ be a pattern described by a p (real) features and let $\Omega = \{\omega_1, \cdots, \omega_c\}$ be a set of classes of cardinality c. The objective of classification is to assign any pattern $\mathbf{x} \in \mathbb{R}^p$ to one of the c classes of Ω. It generally consists of two steps L (*labeling*) and H (*hardening*):

- $L : \mathbf{x} \mapsto \mathbf{u}(\mathbf{x}) = {}^t(u_1(\mathbf{x}), \ldots, u_c(\mathbf{x})) \in \mathcal{L}_{\bullet c}$
- $H : \mathbf{u}(\mathbf{x}) \mapsto \mathbf{h}(\mathbf{x}) = {}^t(h_1(\mathbf{x}), \ldots, h_c(\mathbf{x})) \in \mathcal{L}_{hc}$.

where $\mathcal{L}_{hc} = \{\mathbf{h}(\mathbf{x}) \in \mathcal{L}_{fc} : h_i(\mathbf{x}) \in \{0,1\}\}$ and $\mathcal{L}_{\bullet c}$ depends on the mathematical framework the classifier relies on, e.g. $\mathcal{L}_{pc} = [0,1]^c$ for degrees of typicality or $\mathcal{L}_{fc} = \{\mathbf{u}(\mathbf{x}) \in \mathcal{L}_{pc} : \sum_{i=1}^{c} u_i(\mathbf{x}) = 1\}$ for posterior probabilities and membership degrees. Posterior probabilities can be obtained either from (known) class-conditional densities whose parameters are estimated using a learning set X of patterns, i.e. patterns for which the assignment is known, or from class-density estimates using the classes of their neighbors in X. However, for high dimensional feature spaces, many patterns are required to perform good class-density estimates (curse of dimensionality). When the learning set is small, distances to prototypes of each class are used, e.g.

$$d^2(\mathbf{x}, \mathbf{v}_i) = (\mathbf{x} - \mathbf{v}_i)^t \Sigma_i^{-1} (\mathbf{x} - \mathbf{v}_i) \tag{1}$$

where $\mathbf{v}_i$ and Σ_i are the mean vector and covariance matrix of the class ω_i estimated from X. In order to have a common scaling, label values are often obtained in the unit interval, e.g.

$$u_i(\mathbf{x}) = \frac{\alpha_i}{\alpha_i + d^2(\mathbf{x}, \mathbf{v}_i)} \tag{2}$$

where α_i are user-defined parameters.

We focus on the $H-$step because it is in charge of the classification. This step often reduces to the class of maximum label value selection, resulting in an exclusive classification rule which is not efficient in practice because it supposes that Ω is exhaustively defined (closed-world assumption) and classes do not overlap (separability assumption). Such untrue assumptions can lead to very undesired decisions. In many real applications, it is more convenient to with-hold making a decision than making a wrong assignment, e.g. in medical diagnosis where a false negative outcome can be much more costly than a false positive. Reject options have been proposed to overcome these difficulties and to reduce misclassification risk. The first one, called *distance rejection* [1], is dedicated

ISBN: 978-989-95079-6-8

to outlying patterns. If $\mathbf{x}$ is far from all the class prototypes, this option allows to assign it to no class. The second one, called *ambiguity rejection*, allows to assign inlying patterns to several or all classes [2, 4]. If $\mathbf{x}$ is close to two or more class prototypes, it is associated with the corresponding classes. Including reject options leads to partition the feature space into as many regions as subsets of classes, i.e. at most 2^c ones, to which a pattern can be assigned. Formally, it consists in modifying H such that $\mathbf{h}(\mathbf{x})$ can take values in the set of vertices of the unit hypercube $\mathcal{L}_{hc}^c = \{0,1\}^c$ instead of the exclusive subset $\mathcal{L}_{hc} \subset \mathcal{L}_{hc}^c$. Different strategies can be adopted to handle these options at hand, but they all lead to a three types decision system: distance rejection when $\mathbf{h}(\mathbf{x}) = {}^t(0,\ldots,0) = \underline{0}$, exclusive classification when $\mathbf{h}(\mathbf{x}) \in \mathcal{L}_{hc}$, ambiguity rejection when $\mathbf{h}(\mathbf{x}) \in \mathcal{L}_{hc}^c \backslash \{\mathcal{L}_{hc} \cup \underline{0}\}$. The resulting classification rule is then a matter of selecting the appropriate number of classes varying from zero (distance rejection) to c (total ambiguity rejection).

2.2 Usual reject schemes

Since the work by Chow [2], many rejection schemes have been proposed. In its general form, a class-selective procedure is defined by

$$n^\star(\mathbf{x},t) = \min_{k\in[1,c]} \{k : \mathcal{A}_k(\mathbf{x}) \leq t\} \tag{3}$$

where $\mathcal{A}_k$ is a given ambiguity measure on membership degrees, $n^\star(\mathbf{x},t)$ is the number of selected classes for the pattern $\mathbf{x}$ to be classified, and t is a user-defined threshold which can be class dependent (e.g. t_k). This threshold can be set conditionally to cost functions relative to error, reject and correct classification rates. Propositions from the literature mainly consist in new definitions of ambiguity measures $\mathcal{A}_k$. For Chow, $\mathcal{A}_k$ was one minus the maximum value of the membership degrees. In the paper by Ha [4], the second highest value was tested to decide wether one or several classes are selected. Since this scheme is leading to unnatural decisions, Horiuchi [5] proposed a new measure defined by the difference of the membership degrees which is actually a disambiguity measure. In [6], Frélicot proposed to use the ratio of membership degrees. As usual, we use the convention that if $\mathcal{A}_k(x) > t$ for all k, then we set $n^\star(\mathbf{x},t) = c$. Table 1 summarizes the existing ambiguity measures used for pattern rejection, where the membership degrees are assumed to be sorted in decreasing order for writing convenience, i.e. $u_1(\mathbf{x}) \geq \cdots \geq u_c(\mathbf{x})$.

Table 1: Existing ambiguity measures related to Eq. (3).

Scheme	**Ambiguity Measure** $\mathcal{A}_k(\mathbf{x})$
Chow [2]	$1 - u_1(\mathbf{x})$
Ha [4]	$u_{k+1}(\mathbf{x})$
Horiuchi [5]	$1 - (u_k(\mathbf{x}) - u_{k+1}(\mathbf{x}))$
Frélicot [6]	$u_{k+1}(\mathbf{x})/u_k(\mathbf{x})$

3 Fuzzy residual implications

3.1 Basic definitions

Let us recall basic definitions of fuzzy operators that will be used to combine the values of interest, i.e. the pattern class-degrees of typicality. Depending on properties, aggregation functions can be classified into several categories: conjunctive, disjunctive, compensatory, and so on. We restrict on conjunctive and disjunctive functions. By definition, the output of a conjunctive operator is lower or equal than the minimum value, whereas the output of a disjunctive operator is greater or equal than the maximum value. Beyond these operators, we choose to use the triangular norms because of their ability to generalize the logical AND and OR crisp operators to fuzzy sets, see [3] for a survey. Briefly, a triangular norm (or t-norm) is a binary operation on the unit interval $\top : [0,1]^2 \to [0,1]$ which is commutative, associative, non decreasing and has 1 for neutral element. Thus, a t-norm $\top$ is conjunctive and the minimum operator $\wedge$ is the greatest t-norm. Alternatively, a triangular conorm (or t-conorm) is the dual binary operation $\perp : [0,1]^2 \to [0,1]$ having the same properties except that its neutral element is 0. Thus, a t-conorm $\perp$ is disjunctive and the maximum operator $\vee$ is the lowest t-conorm. Typical examples of dual couples (t-norm, t-conorm) that will be used in the sequel are given in Table 2.

Table 2: Examples of dual couples, including parametrical ones.

Standard	$a\top_S b = \min(a,b)$ $a\perp_S b = \max(a,b)$
Algebraic	$a\top_A b = a\,b$ $a\perp_A b = a + b - a\,b$
Łukasiewicz	$a\top_L b = \max(a+b-1,0)$ $a\perp_L b = \min(a+b,1)$
Hamacher	$a\top_{H_\gamma} b = \frac{a\,b}{\gamma+(1-\gamma)\,(a+b-a\,b)}$ $a\perp_{H_\gamma} b = \frac{a+b+(\gamma-2)\,a\,b}{1+(\gamma-1)\,a\,b}$
Dombi	$a\top_{D_\gamma} b = \left(1+\left(\left(\frac{1-a}{a}\right)^\gamma + \left(\frac{1-b}{b}\right)^\gamma\right)^{1/\gamma}\right)^{-1}$ $a\perp_{D_\gamma} b = 1-\left(1+\left(\left(\frac{a}{1-a}\right)^\gamma + \left(\frac{b}{1-b}\right)^\gamma\right)^{1/\gamma}\right)^{-1}$

A fuzzy residual implication, denoted R-implication (or $\rightarrow$) is defined by:

$$I(a,b) = \sup_t \{t \in [0,1] : \top(a,t) \leq b\} \tag{4}$$

Note that if $\top$ is a left-continuous t-norm, the supremum operation can be substituted by maximum. If we use additive generating functions, i.e. a strictly decreasing function $f : [0,1] \to [0,+\infty]$ with $f(1) = 0$, and admitting an inverse (or pseudo-inverse) function f^{-1}, Eq. (4) can be written as

$$I(a,b) = \max(f^{-1}(f(b) - f(a)), 0) \tag{5}$$

because f is strictly monotonic. We generally speak about an implication function if I is non-increasing in the first variable, non-decreasing in the second variable and $I(0,0) = I(1,1) = 1$, and $I(1,0) = 0$, see [7] for a large survey on fuzzy implication functions. Within theses implications, the well-known *Gödel* and *Goguen* ones are respectively given by

$$I(a,b) = \begin{cases} 1 & \text{if } b \geq a \\ b & \text{if } b < a \end{cases} \tag{6}$$

and

$$I(a,b) = \begin{cases} 1 & \text{if } b \geq a \\ \frac{b}{a} & \text{if } b \leq a \end{cases} \tag{7}$$

ISBN: 978-989-95079-6-8

which are obtained with the minimum and algebraic (or product) triangular norms, respectively. In the sequel, we will assume for writing convenience that the fuzzy values are sorted in decreasing order, e.g. $a \geq b$.

3.2 Parametrical implications

Proposition 1. *Let $(\top_{H_\gamma})$, $\gamma \in [0, +\infty[$, be the family of Hamacher t-norms. The residual Hamacher implication is given by*

$$I_{H_\gamma}(a,b) = \begin{cases} 1 & \text{if } b \geq a \\ \frac{b\,(\gamma+a-\gamma a)}{b\,(\gamma+a-\gamma a)+a-b} & \text{if } b \leq a \end{cases} \tag{8}$$

Proof. By definition of R-implications (4), we can write $I_{H_\gamma}(a,b) = \sup_t\{t \in [0,1] : \top_{H_\gamma}(a,t) \leq b\}$. We also have $I_{H_\gamma}(a,b) = \max_t\{t \in [0,1] : \top_{H_\gamma}(a,t) \leq b\}$ because $\top_{H_\gamma}$ is a left-continuous t-norm. Then, solving

$$\frac{a\,t}{\gamma + (1-\gamma)\,(a+t-a\,t)} \leq b \tag{9}$$

gives

$$t \leq \frac{b\,(\gamma+a-\gamma a)}{b\,(\gamma+a-\gamma a)+a-b}. \tag{10}$$

Since $a \geq b$, it is easy to show that the right part of Eq. (10) is in $[0,1]$, hence we obtain Eq.(8). □

Proposition 2. *Let $(\top_{D_\gamma})$, $\gamma \in]0, +\infty[$, be the family of Dombi t-norms. The residual Dombi implication is given by*

$$I_{D_\gamma}(a,b) = \begin{cases} 1 & \text{if } b \geq a \\ \left(1 + \left(\left(\frac{1-b}{b}\right)^\gamma - \left(\frac{1-a}{a}\right)^\gamma\right)^{1/\gamma}\right)^{-1} & \text{if } b < a \end{cases} \tag{11}$$

Proof. $I_{D_\gamma}(a,b) = \sup_t\{t \in [0,1] : \top_{H_\gamma}(a,t) \leq b\}$ by (4). Since $\top_{D_\gamma}$ is a left-continuous t-norm, we can write $I_{D_\gamma}(a,b) = \max_t\{t \in [0,1] : \top_{D_\gamma}(a,t) \leq b\}$. Then, solving

$$\left(1 + \left(\left(\frac{1-a}{a}\right)^\gamma + \left(\frac{1-t}{t}\right)^\gamma\right)^{1/\gamma}\right)^{-1} \leq b \tag{12}$$

gives

$$t \leq \left(1 + \left(\left(\frac{1-b}{b}\right)^\gamma - \left(\frac{1-a}{a}\right)^\gamma\right)^{1/\gamma}\right)^{-1}. \tag{13}$$

Since, $a \geq b$, it is easy to show that $\left(\left(\frac{1-b}{b}\right)^\gamma - \left(\frac{1-a}{a}\right)^\gamma\right)^{1/\gamma} \geq 0$, hence the right part of Eq. (13) is in $[0,1]$ and (11) is obtained. □

Note that R-implications are mostly used in fuzzy inference systems, see [8] for a large overview on the use of parametrical R-implications in fuzzy rule based systems.

4 Some parametrical measures of ambiguity

In this section, the concept of similarity measure and its relationship with ambiguity measures are described. Then we propose to use fuzzy parametrical implications as new families of ambiguity measures to be used for pattern classification with reject options. The resulting class-selection algorithm (H-step) is presented and we finally discuss the choice of the parameter for parametrical implications with the help of numerical examples.

4.1 Proposition and properties

A similarity measure $\mathcal{S}$ generally satisfies the following properties:

(P1) $\mathcal{S}(a,b) = \mathcal{S}(b,a)$, *(symmetry)*

(P2) $\mathcal{S}(a,a) \geq \mathcal{S}(a,b)$, *(minimality)*

(P3) $\mathcal{S}(a,b) = 1 \Leftrightarrow a = b$, *(identity)*

However, the symmetry property (P1) is still subject to experimental investigation: if $\mathcal{S}(a,b)$ is the answer to the question *how is a similar to b?*, then, when making comparisons, subjects focus more on the feature a than on b. This corresponds to the notion of *saliency of a and b* [9]: if b is more salient than a, then a is more similar to b than vice versa, which is experimentally confirmed. Property (P2) is also open because it can be violated experimentally, see [9] for details.

There are several ways to deal with the comparison of fuzzy values, or fuzzy quantities. The first one is based on a broad class of measures of equality based on a distance measure which is specified for membership functions of fuzzy sets. This approach takes its roots from studies on how to measure the distance between two real functions and do not refer to any specific interpretation. The general form of a Minkowski r-metric is usually taken and leads to well known distance functions (Hamming, Euclidean, Chebyshev). A second way to compare fuzzy values comes from some basic set-theoretic considerations where union, intersection and complement are defined for fuzzy sets. Cardinal and possibility based measures belong to this category. In this paper, we focus on the third way to deal with fuzzy values comparison: the logical framework. Accordingly to [10], for a certain universe of discourse $\mathcal{D}$, the degree of equality of two fuzzy elements a and b can be defined by implications as follows:

$$(a \equiv b) = \frac{1}{2}\big((a \to b) \wedge (b \to a) + (\bar{a} \to \bar{b}) \wedge (\bar{b} \to \bar{a})\big) \tag{14}$$

where $\wedge$ stands for minimum, $\to$ is an implication and $\bar{a}$ is the strict negation $\bar{a} = 1 - a$.

Since 1 is the neutral element of t-norms, applying Eq. (4) with $a \geq b$ gives

$$(a \equiv b) = \frac{1}{2}\big((a \to b) + (\bar{b} \to \bar{a})\big). \tag{15}$$

A convenient way to define an ambiguity measure is to quantify to which extent two fuzzy membership degrees are similar, so that it is closely related to the problem of matching fuzzy quantities, or fuzzy sets similarity. So we propose to use fuzzy R-implications as generalized ambiguity measures.

ISBN: 978-989-95079-6-8

Given a set of c truth values assumed to be sorted in decreasing order, i.e. $u_1(\mathbf{x}) \geq \cdots \geq u_c(\mathbf{x})$, with no loss of generality, let us have two predicates ($\mathbf{x}$ *is* ω_i), with the truth value $u_i(\mathbf{x})$, and ($\mathbf{x}$ *is* ω_k), with the truth value $u_k(\mathbf{x})$. The truth value of the implication *if the pattern* $\mathbf{x}$ *is* ω_i*, then the pattern* $\mathbf{x}$ *is also* ω_j*,* $\forall j$ *varying from* $i+1$ *to* k is an ambiguity measure given by $I(u_i(\mathbf{x}), u_k(\mathbf{x}))$. By convention, we assume that the assignment of $\mathbf{x}$ to ω_i is more probable than the assignment of $\mathbf{x}$ to ω_k when using this implication because $u_i(\mathbf{x}) \geq u_k(\mathbf{x})$ and obviously we have $I(u_i(\mathbf{x}), u_{i+1}(\mathbf{x})) \geq I(u_i(\mathbf{x}), u_{k>i+1}(\mathbf{x}))$.

Proposition 3. *Given* $t \in [0,1]$*, the optimum cardinality of the generalized class-selective rejection rule is given by*

$$n^\star(\mathbf{x}, t) = \min_{k \in [1,c]} \left\{ k : I\Big(u_k(\mathbf{x}), u_{k+1}(\mathbf{x})\Big) \leq t \right\} \qquad (16)$$

with $u_1(\mathbf{x}) \geq \cdots \geq u_c(\mathbf{x})$*, and the convention* $u_{c+1}(\mathbf{x}) = 0$*.*

Since $I(a,0) = 0$ if $a \neq 0$, c classes will be selected if none were previously selected.

Property 1. *If we use the Standard triangular norm* min*, we obtain the Ha class-selective rejection scheme [4].*

Property 2. *If we use the Łukasiewicz triangular norm, we obtain the Horiuchi class-selective rejection scheme [5].*

Property 3. *If we use the Algebraic triangular norm, we obtain the Frélicot class-selective rejection scheme [6].*

Proofs are straightforward and will be given in a longer forthcoming paper. Note that modifying Eq.(16) such as

$$n^\star(\mathbf{x}, t) = \min_{k \in [0,c]} \left\{ k : I\Big(u_k(\mathbf{x}), u_{k+1}(\mathbf{x})\Big) \leq t \right\} \qquad (17)$$

with the convention $u_0(\mathbf{x}) = 1$ allows to select none of the classes, i.e. to proceed to distance rejection, since $I(1,a) = a$ whatever the triangular norm.
The resulting generalized pattern classification rule with reject options ($H-$step) is presented in Algorithm 1. It can be used to compare various schemes, depending on the choice of the triangular norm.

Algorithm 1: $H-$step classification algorithm.

Data: a vector of soft class-labels $\mathbf{u}(\mathbf{x}) \in \mathcal{L}_{pc}$ and a reject threshold t
Result: a vector of class-selective assignments $\mathbf{h}(\mathbf{x}) \in \mathcal{L}_{hc}^c$
begin
  set $\mathbf{h}(\mathbf{x})$ to $\underline{0}$
  find $n^\star(\mathbf{x}, t)$ − Eq.(16) or Eq.(17)
  foreach $j = 1 : n^\star(\mathbf{x}, t)$
  in decreasing order of $u_j(\mathbf{x})'s$ **do**
    set $h_j(\mathbf{x}) = 1$
  end
end

4.2 Discussion and examples

One of the main difficulties when using t-norms is to choose the dual couple and if needed to set the parameter value. Let us study how the choice of γ for Hamacher and Dombi implications modifies the resulting implication strength.
• *Hamacher*: increasing the value of γ will make two fuzzy values more similar because $I_{H_\gamma}(a,b) = ab/(a-b+ab)$ if $\gamma = 0$ while $\lim_{\gamma \to +\infty} I_{H_\gamma}(a,b) = 1$ whatever $(a,b) \in [0,1]^2$. Indeed, I_{H_γ} is non-decreasing with γ since

$$\frac{\partial I_{H_\gamma}}{\partial \gamma} = \frac{(b-ab)(a-b)}{(b(\gamma + a - \gamma a) + a - b)^2} \geq 0.$$

The influence of γ is much more significant for low values of a and b than for high ones because $b(\gamma + a - \gamma a)$ appears to be of order $b\,a$ (respectively $b\,(\gamma + a)$) for high (respectively low) values of a and b. It follows that if $\gamma_1 >> \gamma_2$, we have $\frac{\gamma_1}{\gamma_1+\varepsilon} >> \frac{\gamma_2}{\gamma_2+\varepsilon}$, so that large values of γ associated to low values of (a,b) will result in a high value of I_{H_γ}, see Table 3 for examples.

Table 3: Hamacher implications examples for $a = 0.9$, $b = 0.8$, $c = 0.1$ and $d = 0.05$.

γ	0	2	10
$I_{H_\gamma}(a,b)$	0.87	0.89	0.93
$I_{H_\gamma}(a,c)$	0.10	0.12	0.19
$I_{H_\gamma}(c,d)$	0.09	0.65	0.90

• *Dombi*: decreasing the value of γ will make two fuzzy values more similar because $\lim_{\gamma \to +\infty} I_{D_\gamma}(a,b) = b$ while $\lim_{\gamma \to 0} I_{D_\gamma}(a,b) = 1$ whatever $(a,b) \in [0,1]^2$. Analogously to Hamacher family, I_{D_γ} is non-increasing with γ since

$$\frac{\partial I_{D_\gamma}}{\partial \gamma} \leq 0.$$

Thus, decreasing γ for the Dombi family has the same impact as increasing γ for the Hamacher family, so that we expect the opposite tendency, see Table 4 for examples. The appropriate tuning of γ can be achieved using a gradient procedure, as in [11].

Table 4: Dombi implications examples for $a = 0.9$, $b = 0.8$, $c = 0.1$ and $d = 0.05$.

γ	0.5	2	10
$I_{D_\gamma}(a,b)$	0.97	0.81	0.80
$I_{D_\gamma}(a,c)$	0.12	0.10	0.10
$I_{D_\gamma}(c,d)$	0.35	0.06	0.05

The simple graphical example shown in Fig.1 illustrates how the different implications behave. The top-plot shows a one-dimensional dataset composed of two classes (□ and ×) described by a distance-based model (Eqs.(1-2)) with $\alpha_i = 1$ ($\forall i = 1, c$). In the middle, implication truth values for the standard, algebraic, and Łukasiewicz triangular norms, corresponding to Ha, Frélicot and Horiuchi class-selective schemes respectively, as well as Chow's scheme are shown. Truth values when using parametrical implications are shown in the bottom-plot. Observing the data points in the top-plot allows to obtain the bounds of intervals of x values where the classes do not overlap, therefore for which no misclassification should

ISBN: 978-989-95079-6-8

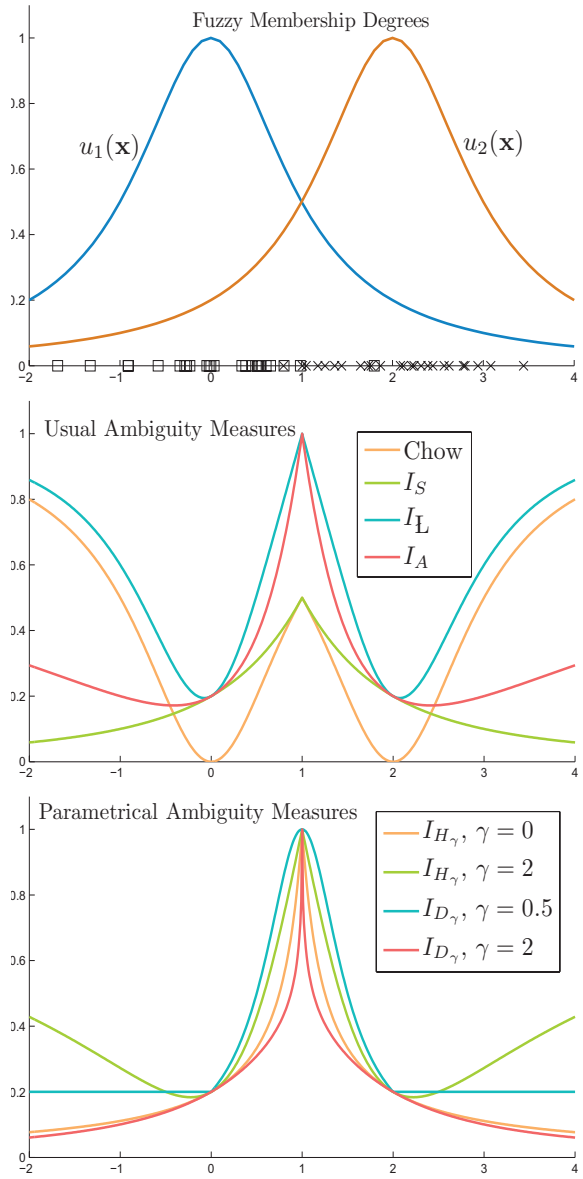


Figure 1: Original outputs using Eqs.(1-2) for two classes with normal distributions (*top*), implication truth values with usual triangular norms (*middle*) and implication truth values with Hamacher and Dombi families (*bottom*).

happen, here $\mathbf{x}_{low} \approx 0.8$ and $\mathbf{x}_{up} \approx 1.9$. Then, the corresponding $u_1(\mathbf{x}_{low})$ and $u_2(\mathbf{x}_{up})$ ordinate values can be used to set the classification threshold t. Then, referring to measures of ambiguity in the middle - and bottom - plots, the behaviour of the different schemes can be analyzed. In such a situation, Chow's and Horiuchi's (Łukasiewicz implication) schemes would lead to reject too many points in low density areas, while other schemes do not show these drawbacks. On the other hand, Ha's scheme (standard implication) would not allow to reject simultaneously highly ambiguous and outlying patterns, whereas schemes based on algebraic or parametrical implications do (provided an appropriate value for the parameter), so a better classification performance can be expected. One can also see in the bottom-plot that tuning the parameter allows to reject patterns for ambiguity as well as for distance (a higher value of γ for the Hamacher family and a lower value for the Dombi family) as pointed out in the discussion.

5 Experimental results

In this section, we report experiments carried out on both artificial and real benchmark datasets for which it is beneficial to introduce the ambiguity reject option because the classes overlap in the feature space. The proposed class-selective rejection scheme is compared to the usual rejection schemes presented in section 2. For additional comparison, results obtained using the two very recent classification rules proposed by Tax and Duin [12] called *Outlier-norm* and *Target-norm* are given, the former being especially designed for distance rejection. Note we did not detailed these (not usual) rules here because they do not rely on an ambiguity measure, therefore they cannot be derived using fuzzy R-implications, by contrast to the previous ones. In all cases, the $L-$step was performed using Eqs.(1-2) with $\alpha_i = 1$ $(\forall i = 1, c)$.

5.1 The datasets

Two synthetic datasets are used: D which contains 2000 points drawn from two normal seven-dimensional distributions of 1000 points each with means $\mathbf{v}_1 = {}^t(1 \;\; 0 \;\cdots\; 0)$ and $\mathbf{v}_2 = {}^t(-1 \;\; 0 \;\cdots\; 0)$, and equal covariance matrices $\Sigma_1 = \Sigma_2 = I$, and DH which consists of two overlapping gaussian classes with different covariance matrices according to the Highleyman distribution, each composed of 800 observations in $\mathbb{R}^2$, see [13]. Real datasets are taken from the UCI Machine Learning Repository [14]. The characteristics (number n of patterns, number p of features, number c of classes, degree of overlap) are reported in Table 5.

Table 5: The datasets and their characteristics.

Dataset	n	p	c	**Overlap**
D	2000	7	2	slight
DH	1600	2	2	very slight
Ionosphere	351	34	2	very strong
Forest	495411	10	2	moderate
Vowel	528	10	11	very slight
Digits	10992	16	10	very slight
Thyroid	215	5	3	very slight
Pima	768	9	2	strong
Statlog	6435	36	6	slight
Glass	214	9	6	moderate

5.2 Results

Table 6 shows the classification performance of the different rejection schemes obtained by a 10-fold cross-validation procedure on the different datasets. In all cases, the threshold t was set to reject 10% of the data, so that 90% is the best achievable correct classification rate performance, then the error rate is $(90 - \text{correct})\%$. The best results are indicated in bold. Note that there are no outliers in the datasets, so that a part of the rejected points are considered as outliers whereas they belong to classes.

It appears from these results that parametrical implications generally outperform usual rejection schemes as well as non usual ones (Outlier-norm, Target-norm), e.g. $I_{H_{\gamma=2}}$ and $I_{D_{\gamma=0.5}}$. As expected in the previous section, these schemes enable to reject both ambiguous and outlying patterns. Furthermore, the tradeoff to be found between a scheme which

ISBN: 978-989-95079-6-8

Table 6: Classification performance (%) on synthetic and real datasets.

Scheme	D	DH	Ionosphere	Forest	Vowel	Digits	Thyroid	Pima	Statlog	Glass
Chow	75.15	83.75	54.42	67.44	87.88	88.03	84.19	59.51	75.66	66.82
Ha	77.35	86.63	54.99	68.35	88.18	87.38	86.98	60.55	75.56	69.63
Horiuchi	77.55	84.31	56.13	69.27	89.09	88.76	86.98	62.89	77.68	68.22
Frélicot	79.65	87.12	**58.12**	69.73	89.09	89.13	**87.91**	63.28	77.48	71.03
Hamacher_0	**80.00**	87.12	55.56	69.76	88.99	88.26	**87.91**	63.15	77.26	70.09
Hamacher_2	79.75	**87.30**	**58.12**	**70.01**	**89.49**	**89.17**	**87.91**	63.28	77.78	**71.33**
$\text{Dombi}_{0.5}$	79.20	86.68	56.41	69.76	89.09	88.96	**87.91**	**63.40**	**78.12**	71.03
Dombi_2	78.85	86.38	55.56	69.65	88.89	87.91	**87.91**	62.50	76.94	70.09
Outlier-norm	76.85	83.31	57.54	67.44	87.88	86.19	84.65	61.33	75.66	69.16
Target-norm	78.80	87.19	56.83	69.69	87.27	86.39	84.19	63.15	75.86	66.82

does not reject outliers (Ha) and others which reject too much patterns (Chow, Horuichi) appears to favour the choice of implications based on a t-norm which is lower than $\top_S$ (the highest t-norm) and greater than $\top_L$, e.g. the algebraic or the parametrical implications (provided an appropriate value for γ). More generally, rejection schemes that take into account relationship between fuzzy membership degrees (Horiuchi, Frélicot, Hamacher, Dombi) perform better than all the others schemes we tested. Looking at the degrees of overlap, $I_{D_{\gamma=0.5}}$ (respectively $I_{H_{\gamma=2}}$) is more efficient for datasets presenting a slight/very slight (respectively strong/very strong) overlap.

6 Conclusion and perspectives

In this paper, a generalized class-selective rejection scheme based on a logical approach to pattern assignation is presented allowing to either reject only ambiguous patterns or ambiguous and outlying patterns. For this purpose, we propose to design a family of ambiguity measures based on fuzzy residual implication functions, which indicate to which extent a pattern should be assigned to $n^\star$ classes depending on its membership degrees from the truth value of the implication. These measures assess the relationships between membership values with various combination of triangular norm operators, including parametrical families. It is shown that the proposed scheme generalizes well-known ones of the literature on pattern classification with reject options. It appears from experiments on both synthetic and real datasets that using, as the basis for residual implication computation, triangular norms greater than the Łukasiewicz triangular norm and lower than the standard one, gives better classification performance. Furthermore, measures taking in consideration several membership degrees, so that interactions between classes are taken into account, also give better results.

A future work will consist in defining new class-selective rejection schemes based on other parametrical triangular families (Sklar, Frank, Yager and so on) and compare their classification performance. We also think that, depending on the context of the pattern recognition problem, other implications functions than the residual implication ones would be suitable. We plan to make an extensive study on the behavior of *S-implications* which are an immediate generalization of the usual boolean implication, *QL-implications* coming from quantum mechanic logic, and *D-implications* which are the contraposition with respect to a negation of QL-implications, see [7] for definitions.

References

[1] B. Dubuisson and M. Masson. A statistical decision rule with incomplete knowledge about classes. *Pattern Recognition*, 26:155–165, 1993.

[2] C.K. Chow. On optimum error and reject tradeoff. *IEEE Transactions on Information Theory*, 16:41–46, 1970.

[3] E.P Klement and R. Mesiar. *Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms*. Elsevier, 2005.

[4] T. Ha. The optimum class-selective rejection rule. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(6):608–615, 1997.

[5] T. Horiuchi. Class-selective rejection rule to minimize the maximum distance between selected classes. *Pattern Recognition*, 31:1579–1588, 1998.

[6] C. Frélicot and B. Dubuisson. A multi-step predictor of membership function as an ambiguity reject solver in pattern recognition. In *4th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-based Systems*, pages 709–715. IPMU'92, July 1992.

[7] M. Mas, M. Monserrat, J. Torrens, and E. Trillas. A survey on fuzzy implication functions. *IEEE Transactions on Fuzzy Systems*, 15(6):1107–1121, 2007.

[8] Th. Whalen. Parameterized r-implications. *Fuzzy Sets and Systems*, 134(2):231–281, 2003.

[9] A. Tversky. Features of similarity. *Psychological review*, 84(4):327–352, 1977.

[10] K. Hirota and W. Pedrycz. Matching fuzzy quantities. *IEEE Transactions on Systems, Man, and Cybernetics*, 21(6):1580–1586, 1991.

[11] L. Rutkowski and K. Cpalka. Flexible neuro-fuzzy systems. *IEEE Transactions on Neural Networks*, 14(3):554–574, 2003.

[12] D.M.J. Tax and R.P.W. Duin. Growing a multi-class classifier with a reject option. *Pattern Recognition Letters*, 29(10):1565–1570, 2008.

[13] W. Highleyman. Linear decision functions, with application to pattern recognition. *Proc. IRE 50*, 50(6):1501–1514, 1962.

[14] C. Blake and C. Merz. Uci repository of machine learning databases, 1998.

ISBN: 978-989-95079-6-8

Fuzzy c-Lines for Data with Tolerance

Yuchi Kanzawa[1] Yasunori Endo[2] Sadaaki Miyamoto[2]

1.Shibaura Institute of Technology, Japan
2.University of Tsukuba, Japan
Email: kanzawa@sic.shibaura-it.ac.jp

Abstract— *This paper presents a new clustering algorithm ,which is based on fuzzy c-lines, can treat data with some errors. First, the tolerance is formulated and introduce into optimization problem of clustering. Next, the problem is solved using Karush-Kuhn-Tucker conditions. Last, the algorithm is constructed based on the results of solving the problem. Some numerical examples for the proposed method are shown.*

Keywords— Data with Tolerance, Fuzzy c-Lines, Fuzzy Clustering

1 Introduction

Fuzzy c-means (FCM) [1] is one of the well-known fuzzy clustering and many FCM variants have been proposed after FCM. In these variants, an algorithm allowing cluster prototypes to be linear varieties instead of points has been proposed [1]. This algorithm is called fuzzy c-varieties (FCV) and regarded as the combination of clustering and principal component analysis. If linear varieties of the prototypes are limited to one dimensional, FCV is called fuzzy c-lines(FCL).

In general, any data that is represented by numeric have some errors. In order to classify such data, an algorithm has been proposed by some of the authors [2]. This algorithm is called the FCM for data with tolerance (FCM-T). In FCM-T, the tolerance is regarded as another decision variable composing the optimization problem like membership grades and the cluster centers, and is determined as minimizing the objective function.

In this paper, on the analogy of the story from FCM to FCM-T, we propose fuzzy c-lines for data with tolerance (FCL-T).

The contents of this paper are the followings. In the second section, we define some notations and introduce FCM-T and FCL. In the third section, we propose FCL-T. In the fourth section, some numerical examples are shown. In the last section, we conclude this paper.

2 Preliminaries

In this section, we define some notations, introduce fuzzy c-lines (FCL), and also introduce fuzzy c-means with tolerance (FCM-T) derived from conventional fuzzy c-means (FCM). In the first subsection, we define some notations which are the data for clustering, the membership by which the each data belongs to the each cluster, the cluster prototypes of one dimensional linear manifold. In the second subsection, we introduce FCL, whose prototypes are one dimensional linear manifolds. In the third subsection, we define the tolerance for the data and the maximum tolerance for the data, and introduce FCM-T, which is the story that clustering algorithms were oriented for data with some errors.

2.1 Notations

In this subsection, we define some notations which are the data for clustering, the membership by which the each data belongs to the each cluster, the cluster prototypes of one dimensional linear manifold.

The data set $x = \{x_i \mid x_i \in \mathbf{R}^p, i \in \{1, \ldots, N\}\}$ is given. The membership by which x_i belongs to the j-th cluster is denoted by $u_{i,j}$ $(i \in \{1, \cdots, N\}, j \in \{1, \cdots, C\})$ and the set of $u_{i,j}$ is denoted by $u \in \mathbf{R}^{N \times C}$ called the partition matrix. The constraint for u is

$$\sum_{j=1}^{C} u_{i,j} = 1 \quad (0 \leq u_{i,j} \leq 1).$$

The cluster prototype set is denoted by $V = \{V_1, \cdots, V_C\}$, where

$$V_j = \{y_j \in \mathbf{R}^p \mid y_j = v_j + t_j s_j, t_j \in \mathbf{R}\}$$

with $v = \{v_j \mid v_j \in \mathbf{R}^p, j \in \{1, \ldots, C\}\}$ and $s = \{s_j \in \mathbf{R}^p \mid \|s_j\| = 1, j \in \{1, \cdots, C\}\}$.

2.2 Fuzzy c-Lines

In this subsection, we introduce FCL (FCV), whose prototypes are one dimensional linear manifolds.

FCL is the algorithm obtained by solving the following optimization problem:

$$\underset{u,v,s}{\text{minimize}}\, J_\ell(u, v, s) \tag{1}$$

$$\text{subject to} \sum_{j=1}^{C} u_{i,j} = 1, \quad \|s_j\| = 1, \tag{2}$$

where

$$J_\ell(u, v, s) = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{i,j}^m \left(\|x_i - v_j\|^2 - ((x_i - v_j) s_{j,\ell})^2 \right). \tag{3}$$

The optimal solutions u, v and s are obtained by the following algorithm.

Algorithm 1 (FCL)

Step 1 *Give the number of clusters C and the fuzzifier parameter m. Set the initial values of v and s.*

Step 2 *Calculate u such that*

$$u_{i,j} = 1 / \sum_{k=1}^{C} \left(\frac{d_{i,j}}{d_{i,k}} \right)^{1/(m-1)}, \tag{4}$$

where

$$d_{i,j} = \|x_i - v_j\|^2 - ((x_i - v_j) s_{j,\ell})^2. \tag{5}$$

ISBN: 978-989-95079-6-8

Step 3 *Calculate v such that*

$$v_j = U_j^{-1} \sum_{i=1}^{N} u_{i,j}^m x_i, \tag{6}$$

where

$$U_j = \sum_{i=1}^{N} u_{i,j}^m. \tag{7}$$

Step 4 *Calculate s as the eigenvector corresponding to the maximal eigenvalue of the following matrix:*

$$\sum_{i=1}^{N} u_{i,j}^m (x_i - v_j)(x_i - v_j)^{\mathrm{T}}. \tag{8}$$

Step 5 *Check the stopping criterion for (u, v, s). If the criterion is not satisfied, go back to* Step 2.

2.3 Fuzzy c-Means for Data with Tolerance

In this subsection, we define the tolerance for the data and the maximum tolerance for the data, and introduce fuzzy c-means for data with tolerance (FCM-T) [2]. This algorithm is oriented toward data with error. The tolerance for the data x is denoted by $\varepsilon = \{\varepsilon_i \mid \varepsilon_i \in \mathbf{R}^p, i \in \{1, \ldots, N\}\}$. The maximum tolerance is denoted by $\kappa = \{\kappa_i \mid \kappa_i \in \mathbf{R}_+, i \in \{1, \ldots, N\}\}$.

FCM-T is the algorithm obtained by solving the following optimization problem:

$$\underset{u,\varepsilon,v}{\text{minimize}}\, J_{\mathrm{t}}(u, \varepsilon, v)$$

$$\text{subject to} \quad \begin{cases} \displaystyle\sum_{j=1}^{C} u_{i,j} = 1, \\ \|\varepsilon_i\|^2 \le \kappa_i^2 \quad (\kappa_i > 0), \end{cases}$$

where

$$J_{\mathrm{t}}(u, \varepsilon, v) = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{i,j}^m \|x_i + \varepsilon_i - v_j\|^2. \tag{9}$$

The optimal solutions u, ε and v are obtained by the following algorithm.

Algorithm 2 (FCM-T)

Step 1 *Give the number of clusters C, the fuzzifier parameter m and the maximum tolerance set κ for data x. Set the initial values of ε and v.*

Step 2 *Calculate u such that*

$$u_{i,j} = 1 / \sum_{k=1}^{C} \left(\frac{d_{i,j}}{d_{i,k}} \right)^{1/(m-1)}, \tag{10}$$

where

$$d_{i,j} = \|x_i + \varepsilon_i - v_j\|^2. \tag{11}$$

Step 3 *Calculate ε such that*

$$\varepsilon_i = -\alpha_i \left(x_i - \sum_{j=1}^{C} u_{i,j}^m v_j \right), \tag{12}$$

where

$$\alpha_i = \min \left\{ \kappa_i \left\| x_i - \sum_{j=1}^{C} u_{i,j}^m v_j \right\|^{-1}, \left(\sum_{j=1}^{C} u_{i,j}^m \right)^{-1} \right\}. \tag{13}$$

Step 4 *Calculate v such that*

$$v_j = U_j^{-1} \sum_{i=1}^{N} u_{i,j}^m (x_i + \varepsilon_i), \tag{14}$$

where

$$U_j = \sum_{i=1}^{N} u_{i,j}^m. \tag{15}$$

Step 5 *Check the stopping criterion for (u, ε, v). If the criterion is not satisfied, go back to* Step 2.

If the maximal tolerance κ_i is set to zero, this algorithm is coincides with FCM without tolerance. This algorithm is executing FCM without tolerance for data moving within a region as shown in Fig. 1.

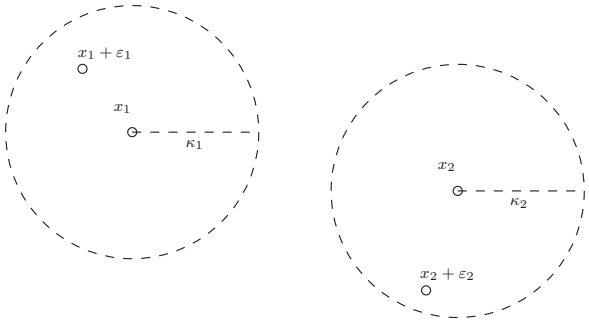


Figure 1: Data with Tolerance

3 Fuzzy c-Lines for Data with Tolerance

In this section, fuzzy c-lines for data with tolerance (FCL-T) is proposed. First, the tolerance is formulated and introduce into optimization problem of FCL. Next, the problem is solved using Karush-Kuhn-Tucker conditions. Last, the algorithm is constructed based on the results of solving the problem.

3.1 Optimization Problem and Its KKT conditions

In this subsection, a new optimization problem is proposed such that the tolerance is formulated and introduce into optimization problem of FCL, and its Karush-Kuhn-Tucker conditions are considered.

Let us consider the following optimization problem:

$$\underset{u,\varepsilon,v,s}{\text{minimize}}\, J_{\ell,\mathrm{t}}(u, \varepsilon, v, s)$$

$$\text{subject to} \quad \begin{cases} \displaystyle\sum_{j=1}^{C} u_{i,j} = 1, \\ \|s_j\| = 1, \\ \|\varepsilon_i\|^2 \le \kappa_i^2, \end{cases} \tag{16}$$

ISBN: 978-989-95079-6-8

where

$$J_{\ell,\mathrm{t}}(u,\varepsilon,v,s) = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{i,j}^{m}\left(\|x_i+\varepsilon_i-v_j\|^2-((x_i+\varepsilon_i-v_j)s_j)^2\right). \tag{17}$$

Lagrange function of this problem $L_{\ell,\mathrm{t}}(u,\varepsilon,v,s)$ is described as below:

$$\begin{aligned} L_{\ell,\mathrm{t}}(u,\varepsilon,v,s) =& J_{\mathrm{s},\ell,\mathrm{t}}(u,\varepsilon,v,s) + \sum_{i=1}^{N}\gamma_i\left(\sum_{j=1}^{} u_{i,j}-1\right) \\ &+\sum_{j=1}^{C}\xi_j\left(s_j^{\mathrm{T}}s_j-1\right) \\ &+\sum_{i=1}^{N}\delta_i\left(\|\varepsilon_i\|^2-\kappa_i^2\right). \end{aligned} \tag{18}$$

where $\gamma = (\gamma_1,\cdots,\gamma_N)$, $\xi = (\xi_1,\cdots,\xi_C)$ and $\delta = (\delta_1,\cdots,\delta_N)$ are Karush-Kuhn-Tucker vectors. Karush-Kuhn-Tucker (KKT) conditions are described as below:

$$\frac{\partial L_{\ell,\mathrm{t}}}{\partial u_{i,j}} =0, \tag{19}$$
$$\frac{\partial L_{\ell,\mathrm{t}}}{\partial v_j} =0, \tag{20}$$
$$\frac{\partial L_{\ell,\mathrm{t}}}{\partial s_j} =0, \tag{21}$$
$$\frac{\partial L_{\ell,\mathrm{t}}}{\partial \varepsilon_i} =0, \tag{22}$$
$$\frac{\partial L_{\ell,\mathrm{t}}}{\partial \gamma_i} =0, \tag{23}$$
$$\frac{\partial L_{\ell,\mathrm{t}}}{\partial \xi_j} =0, \tag{24}$$
$$\frac{\partial L_{\ell,\mathrm{t}}}{\partial \delta_i} =0, \tag{25}$$
$$\delta_i\frac{\partial L_{\ell,\mathrm{t}}}{\partial \delta_i} =0, \tag{26}$$
$$\delta_i \geq 0. \tag{27}$$

3.2 Optimal Solution of $u_{i,j}$, v_j and s_j

In this subsection, we derive the optimal solution of $u_{i,j}$, v_j and s_j from the above KKT conditions, and we show that these solutions are just substituting $x_i+\varepsilon_i$ to x_i in FCL (Algorithm 1).

From KKT conditions (19) and (23), we have the optimal solution of $u_{i,j}$ as

$$u_{i,j} = 1/\sum_{k=1}\left(\frac{d_{i,j}}{d_{i,k}}\right)^{1/(m-1)}, \tag{28}$$

where

$$d_{i,j} = \|x_i+\varepsilon-v_j\|^2-(s_j^{\mathrm{T}}(x_i+\varepsilon_i-v_j))^2. \tag{29}$$

From KKT conditions (20), we have

$$(E-s_js_j^{\mathrm{T}})\sum_{i=1}^{N}u_{i,j}^{m}(x_i+\varepsilon_i-v_j)=0, \tag{30}$$

where E is the p-dimensional unit matrix. The null space of the matrix $E-s_js_j^{\mathrm{T}}$ include s_j since

$$(E-s_js_j^{\mathrm{T}})s_j = s_j-(s_j^{\mathrm{T}}s_j)s_j=0. \tag{31}$$

For the p-dimensional vector w such as $w^{\mathrm{T}}s_j=0$ and as $w^{\mathrm{T}}w=1$ and for a scalar value α, we have

$$\begin{aligned}(E-s_js_j^{\mathrm{T}})(\alpha w) =& \alpha(w-(s^{\mathrm{T}}w)s_j)\\ =&\alpha w,\end{aligned} \tag{32}$$

which is equal to 0 if and only if $\alpha=0$, hence, the null space of $E-s_js_j^{\mathrm{T}}$ is just s_j. Therefore, Eq.(30) indicates that

$$\sum_{i=1}^{N}u_{i,j}^{m}(x_i+\varepsilon_i-v_j)=as_j \tag{33}$$

for an arbitrary value of a. From the above, we have the optimal solution of v_j as

$$v_j=\frac{\sum_{i=1}^{N}u_{i,j}^{m}(x_i+\varepsilon_i)-as_j}{\sum_{i=1}^{N}u_{i,j}^{m}}, \tag{34}$$

in which we may set $a=0$ and have

$$v_j=\frac{\sum_{i=1}^{N}u_{i,j}^{m}(x_i+\varepsilon_i)}{\sum_{i=1}^{N}u_{i,j}^{m}}. \tag{35}$$

From KKT conditions (21) and (24), we have

$$\sum_{i=1}^{N}u_{i,j}^{m}(x_i+\varepsilon_i-v_j)(x_i+\varepsilon_i-v_j)^{\mathrm{T}}s_j=-\xi_js_j, \tag{36}$$

from which we can obtain the optimal solution of s_j as the eigenvector for the maximal eigenvalue of the matrix

$$\sum_{i=1}^{N}u_{i,j}^{m}(x_i+\varepsilon_i-v_j)(x_i+\varepsilon_i-v_j)^{\mathrm{T}}. \tag{37}$$

Comparing between Eq. (28) and (4), Eq. (29) and (5), Eq. (35) and (6), and Eq. (37) and (8), we can see that optimal solutions in FCL-T is obtained just by substituting $x_i+\varepsilon_i$ to x_i in FCL (Algorithm 1).

3.3 Optimal Solution of ε_i

In this subsection, we obtain the optimal solution of ε_i from the KKT conditions.

From KKT conditions (22), (25), (26) and (27), we have

$$\left(\delta_iE+\sum_{j=1}^{C}u_{i,j}^{m}(E-s_js_j^{\mathrm{T}})\right)\varepsilon_i+ \sum_{j=1}^{C}u_{i,j}^{m}(E-s_js_j^{\mathrm{T}})(x_i-v_j)=0, \tag{38}$$

$$\delta_i(\|\varepsilon_i\|^2-\kappa_i^2)=0. \tag{39}$$

ISBN: 978-989-95079-6-8

In order to solve this equation, we discuss in each case that the matrix $\left(\sum_{j=1}^{C} u_{i,j}^m (E - s_j s_j^{\mathrm{T}})\right)$ is invertible or not as the followings.

The matrix $\left(\sum_{j=1}^{C} u_{i,j}^m \left(E - s_j s_j^{\mathrm{T}}\right)\right)$ is not invertible if $s_j = s_{\tilde{j}}$ for any $j \neq \tilde{j}$ or if there exists j^* such as $u_{i,j^*} = 1$, though the proof is omitted because of the sake of the pages. Here, we write $s = s_j$ in the case of $s_j = s_{\tilde{j}}$ for any $j \neq \tilde{j}$ and also write $s = s_{j^*}$ in the case that there exists j^* such as $u_{i,j^*} = 1$. In these cases, Eq. (38) is rewritten by setting $\varepsilon_i = \alpha s + \beta w$ with certain values α and β, and a p-dimensional vector w such as $s^{\mathrm{T}} w = 0$ and as $w^{\mathrm{T}} w = 1$, as

$$\beta(\sum_{j=1}^{C} u_{i,j}^m) w + \delta_i(\alpha s + \beta w) + (E - ss^{\mathrm{T}}) \sum_{j=1}^{C} u_{i,j}^m (x_i - v_j) = 0. \quad (40)$$

Multiplying s^{T} to Eq.(40), we have $\delta_i \alpha = 0$, which indicates $\delta_i = 0$ or $\alpha = 0$. $\alpha = 0$ means that ε_i is orthogonal to s. In the case of $\delta_i = 0$, we have

$$(E - ss^{\mathrm{T}}) \sum_{j=1}^{C} u_{i,j}^m (x_i + \varepsilon_i - v_j) = 0. \quad (41)$$

which implies

$$\sum_{j=1}^{C} u_{i,j}^m (x_i + \varepsilon_i - v_j) = as \quad (42)$$

with an arbitrary value a. Hence, we have the optimal solution of ε_i in the case that $\left(\sum_{j=1}^{C} u_{i,j}^m (E - s_j s_j^{\mathrm{T}})\right)$ is not invertible with $\delta_i = 0$, as

$$\varepsilon_i = \frac{as - \sum_{j=1}^{C} u_{i,j}^m (x_i - v_j)}{\sum_{j=1}^{C} u_{i,j}}. \quad (43)$$

If we set a such that ε_i is orthogonal to s, we can discuss both the case of $\delta_i \alpha = 0$ and the base of $\alpha = 0$ as the followings. Since $s^{\mathrm{T}} \varepsilon_i = 0$, Eq.(38) is rewritten as

$$\left(\sum_{j=1}^{C} u_{i,j}^m\right) \varepsilon_i + \delta_i \varepsilon_i + (E - ss^{\mathrm{T}}) \sum_{j=1}^{C} u_{i,j}^m (x_i - v_j) = 0, \quad (44)$$

from which we have

$$\varepsilon_i = -\frac{(E - ss^{\mathrm{T}}) \sum_{j=1}^{C} u_{i,j}^m (x_i - v_j)}{\sum_{j=1}^{C} u_{i,j}^m + \delta_i}. \quad (45)$$

If $\delta_i = 0$, we have

$$\varepsilon_i = -\frac{(E - ss^{\mathrm{T}}) \sum_{j=1}^{C} u_{i,j}^m (x_i - v_j)}{\sum_{j=1}^{C} u_{i,j}^m}. \quad (46)$$

If $\delta_i \neq 0$, then $\|\varepsilon_i\| = \kappa_i^2$ from Eq. (39), by which we have

$$\|\varepsilon_i\|^2 = \left\| \frac{(E - ss^{\mathrm{T}}) \sum_{j=1}^{C} u_{i,j}^m (x_i - v_j)}{\sum_{j=1}^{C} u_{i,j}^m + \delta_i} \right\|^2 = \kappa_i^2. \quad (47)$$

By solving Eq.(47) as

$$\sum_{j=1}^{C} u_{i,j}^m + \delta_i = \frac{\left\| (E - ss^{\mathrm{T}}) \sum_{j=1}^{C} u_{i,j}^m (x_i - v_j) \right\|}{\kappa_i}, \quad (48)$$

we have

$$\varepsilon_i = -\kappa_i \left\| (E - ss^{\mathrm{T}}) \sum_{j=1}^{C} u_{i,j}^m (x_i - v_j) \right\|^{-1} \cdot (E - ss^{\mathrm{T}}) \sum_{j=1}^{C} u_{i,j}^m (x_i - v_j). \quad (49)$$

Getting Eq.(46) and (49) together, we have the optimal solution of ε_i in the case that the matrix $\left(\sum_{j=1}^{C} u_{i,j}^m \left(E - s_j s_j^{\mathrm{T}}\right)\right)$ is not invertible

$$\varepsilon_i = -\min \left\{ \kappa_i \left\| (E - ss^{\mathrm{T}}) \sum_{j=1}^{C} u_{i,j}^m (x_i - v_j) \right\|^{-1}, \left(\sum_{j=1}^{C} u_{i,j}^m\right)^{-1} \right\} (E - ss^{\mathrm{T}}) \sum_{j=1}^{C} u_{i,j}^m (x_i - v_j). \quad (50)$$

In the case that $\left(\sum_{j=1}^{C} u_{i,j}^m (E - s_j s_j^{\mathrm{T}})\right)$ is invertible, we cannot obtain the explicit form of the optimal solution ε_i, so, we apply some numerical method for Eq. (38) and (39) as the followings. First, we set $\delta_i = 0$ and solve Eq. (38) as

$$\varepsilon_i = -\left(\sum_{j=1}^{C} u_{i,j}^m (E - s_j s_j^{\mathrm{T}})\right)^{-1} \cdot \sum_{j=1}^{C} u_{i,j}^m (E - s_j s_j^{\mathrm{T}})(x_i - v_j). \quad (51)$$

If the obtained value of ε_i satisfies $\|\varepsilon_i\|^2 \leq \kappa_i^2$, we adopt this value as the optimal solution of ε_i. Otherwise, we solve Eq. (38) with $\|\varepsilon_i\|^2 = \kappa_i^2$ numerically.

ISBN: 978-989-95079-6-8

3.4 Algorithm

In this subsection, we propose an iterative algorithm of fuzzy c-lines for data with tolerance from the above discussion.

Algorithm 3 (FCL-T)

Step 1 *Give the values of the fuzzifier parameter m and the maximal tolerance set κ. Set the initial cluster center v, s and ε.*

Step 2 *Calculate $u_{i,j}$ such that*

$$u_{i,j} = 1/\sum_{k=1}^{C}\left(\frac{d_{i,j}}{d_{i,k}}\right)^{1/(m-1)}, \tag{52}$$

where

$$d_{i,j} = \|x_i + \varepsilon - v_j\|^2 - (s_j^{\mathrm{T}}(x_i + \varepsilon_i - v_j))^2. \tag{53}$$

Step 3 *Calculate v_j such that*

$$v_j = \frac{\sum_{i=1}^{N} u_{i,j}^m (x_i + \varepsilon_i)}{\sum_{i=1}^{N} u_{i,j}^m}. \tag{54}$$

Step 4 *Obtain s_j as the eigenvector with the maximal eigenvalue of the matrix*

$$\sum_{i=1}^{N} u_{i,j}^m (x_i + \varepsilon_i - v_j)(x_i + \varepsilon_i - v_j)^{\mathrm{T}}. \tag{55}$$

Step 5 *If $s_j = s_{\tilde{j}}$ for any $j \neq \tilde{j}$ or if there exists j such that $u_{i,j} = 1$, obtain ε_i as*

$$\varepsilon_i = -\min\left\{\kappa_i \left\|(E - ss^{\mathrm{T}})\sum_{j=1}^{C} u_{i,j}^m (x_i - v_j)\right\|^{-1}, \left(\sum_{j=1}^{C} u_{i,j}^m\right)^{-1}\right\}(E - ss^{\mathrm{T}})\sum_{j=1}^{C} u_{i,j}^m (x_i - v_j), \tag{56}$$

where

$$s = \begin{cases} s_j & \text{for} s_j = s_{\tilde{j}} \quad (j \neq \tilde{j}), \\ s_{j^*} & \text{for } u_{i,j^*} = 1, \end{cases} \tag{57}$$

and go to Step 7. *Otherwise, go to* Step 6.

Step 6 *Calculate*

$$\varepsilon_i = -\left(\sum_{j=1}^{C} u_{i,j}^m (E - s_j s_j^{\mathrm{T}})\right)^{-1} \cdot \sum_{j=1}^{C} u_{i,j}^m (E - s_j s_j^{\mathrm{T}})(x_i - v_j). \tag{58}$$

If $\|\varepsilon_i\|^2 \leq \kappa_i^2$ holds, adopt this ε_i and go to Step 7. *Otherwise, solve the equation*

$$\left(\delta_i E + \sum_{j=1}^{C} u_{i,j}^m (E - s_j s_j^{\mathrm{T}})\right)\varepsilon_i + \sum_{j=1}^{C} u_{i,j}^m (E - s_j s_j^{\mathrm{T}})(x_i - v_j) = 0, \tag{59}$$

$$\|\varepsilon_i\|^2 = \kappa_i^2 \tag{60}$$

Step 7 *Check the stopping criterion. If the criterion is not satisfied, go back to* Step 2.

If the maximal tolerance κ_i is set to zero, this algorithm is coincides with FCL without tolerance (Algorithm 1). Remark that Eq. (56) in Algorithm 3 and Eq. (13) for obtaining the tolerance in Algorithm 2 are essentially the same with each other, because Eq. (56) means that the tolerance is obtaining in the orthogonal complementary space of s in the pattern space. Remark also that there is the quite similarity between Eq. (58) in Algorithm 3 and Eq. (9.29) in [3] for dealing with the missing data in linear fuzzy clustering followed by the optimal completion strategy [4], because Eq. (58) means that the optimal solution of the tolerance in a certain condition is not affected by the constraint of the tolerance $\|\varepsilon_i\| \leq \kappa_i^2$, while Eq. (9.29) in [3] does not consider the constraint for missing component of data from the beginning.

4 Numerical Example

In this section, we show some examples of classification by fuzzy c-lines for data with tolerance (Algorithm 3). In each example, 100 trials for Algorithm 3 with different initial cluster centers are tested and the solution with the minimal objective function value is selected as the final result. The classified dataset is Gustaffson's cross [5] shown in Fig. 2. This dataset consists of 20 points ($N = 20$) in $\mathbf{R}^2$ and forms two visually apparent linear clusters with the crossing shape.

We fix $m = 2$ and test four different values of $\kappa_i \in \{0, 0.016, 0.032, 0.064\}$. All cases of κ_i produce the correctly classified results shown in Fig. 3, 4 and 5, respectively. Especially, the cases of $\kappa_i \in \{0.032, 0.064\}$ produce the same results shown in Fig. 5. From Fig. 4, we can see that each data moves to the direction of its own prototype. From Fig. 5, we can see that each data moves just on its own prototype.

From these results, we consider the following hypothesis: if Gustaffson's cross is generated by adding error followed by uniform distribution with a width after sampling from the two crossing exactly linear shaped population, FCL-T (Algorithm 3) produces the shape of the original population by $v_j + ts_j$ and also produces the width of the uniform distribution by κ_i, though we have not proved it yet.

5 Conclusion

In this paper, we proposed fuzzy c-lines for data with tolerance. The proposed method is obtained by introducing the idea of tolerance, which means that each datum moves within a region, into fuzzy c-lines. This is the analogy with that fuzzy c-means for data with tolerance is obtained by introducing the idea of tolerance into fuzzy c-means. In the numerical examples, we can see that each data moves to the direction of its own prototype in setting the positive maximal tolerance, and that each data moves on its own prototype in setting larger maximal tolerance.

ISBN: 978-989-95079-6-8

From this feature of the proposed method, we consider the following hypothesis: if dataset is generated by adding error followed by uniform distribution with a width after sampling from the exactly linear shaped population, FCL-T (Algorithm 3) produces the shape of the original population and also produces the width of the uniform distribution. But we must investigate the proposed method theoretically and through many numerical experiments, which is the first future work.

As another future works, (1) entropy regularlized fuzzy c-means [6] for data with tolerance will be proposed by the similar discussion, (2) the cases of error followed by other probabilistic distribution will be considered, e.g., the case of error followed by normal distribution can be considered by adding the penalty term of squared tolerance, and Laplace one by the penalty term of absolute value of tolerance.

References

[1] Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms, Plenun, New York, (1981).

[2] Endo, Y., Miyamoto, S.: "Fuzzy c-Means for Data with Tolerance", Proc. 2005 International Symposium on Nonlinear Theory and Its Applications, pp.123–127, 2005.

[3] Miyamoto, S., Ichihashi, H., Honda, K.: "Algorithms for Fuzzy Clustering", Springer, Berlin (2008).

[4] Hathaway, R.J., Bezdek, J.C.: "Fuzzy c-means Clustering of Incomplete Data", IEEE Trans. on Systems, Man and Cybernetics Part B 31(5), pp.735–744 (2001).

[5] Gustaffson, D.E., Kessel, W.: "Fuzzy Clustering with a Fuzzy Covariance Matrix", Proc. IEEE–CDC, Vol.2, pp.761–766 (1979).

[6] Miyamoto, S. and Mukaidono, M.: "Fuzzy c-Means as a Regularization and Maximum Entropy Approach", Proc. of the 7th International Fuzzy Systems Association World Congress (IFES'97), Vol. II, pp. 86–92, Prague, 1997.

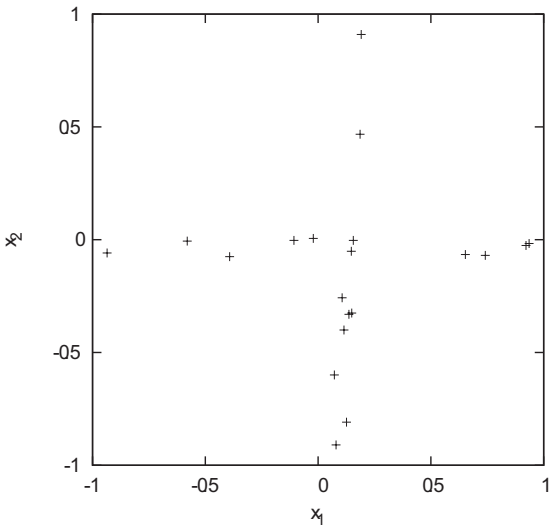


Figure 2: Gustaffson's Cross

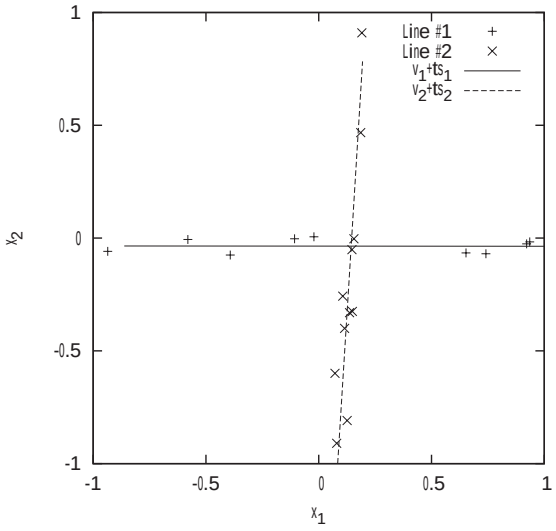


Figure 3: Successful Classification Result of Fig. 2 by FCL-T 3 with $m = 2$ and $\kappa_i = 0$

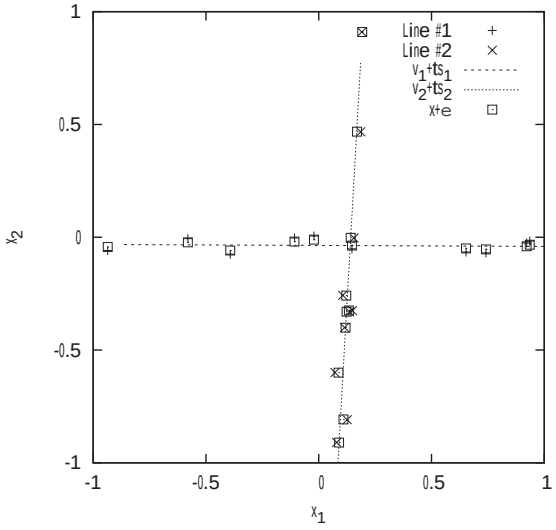


Figure 4: Successful Classification Result of Fig. 2 by FCL-T 3 with $m = 2$ and $\kappa_i = 0.016$

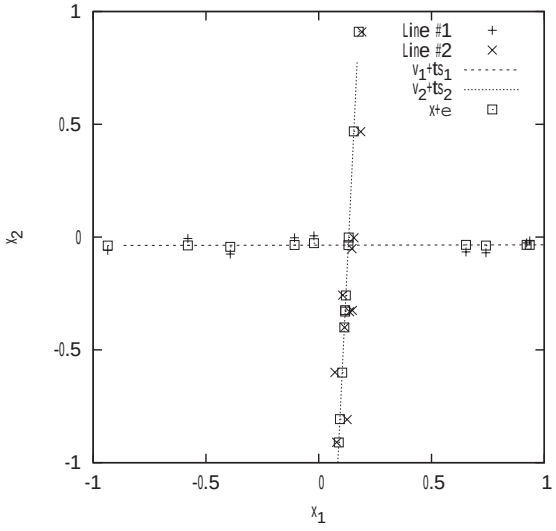


Figure 5: Successful Classification Result of Fig. 2 by FCL-T 3 with $m = 2$ and $\kappa_i = 0.032$

ISBN: 978-989-95079-6-8

Geographic Information Retrieval based on two orthogonal criteria

Gloria Bordogna [1] Giorgio Ghisalberti [2] Marco Pagani [1] Giuseppe Psaila [3]

1 CNR - IDPA, via Pasubio 5, c/o POINT, Dalmine 24044 (BG), Italy
2 Politecnico di Milano, Dip. Di Elettronica, via Bassini 16, 20133 Milano, Italy
3 Università di Bergamo, Facoltà di Ingegneria, via G. Marconi 5, 24044 Dalmine (BG) Italy
Email: gloria.bordogna@idpa.cnr.it, ghisalberti@elet.polimi.it, marco.pagani@idpa.cnr.it, psaila@unibg.it

Abstract— *A Geographic Information Retrieval (GIR) model is defined and the architecture of a GIR system, named Geo-Finder, is described. The GIR system is constituted by two main modules. The geo-indexing module applies bipolar criteria to automatically identify the geo-reference focus (footprint) of textual documents. The geo-retrieval module applies a context dependent matching function to evaluate queries consisting of two orthogonal constraints, a content constraint and a spatial constraint; the spatial constraint is defined by modeling the user's perception of geographic "closeness" between the documents' footprint and the query footprint. For each retrieved document, two relevance scores are computed with respect to the two query conditions, that can be combined to generate an overall ranked list of documents in a flexible way, by allowing users to specify a trade off between them.*

Keywords— bipolar criteria evaluation, geo-footprint, geo-indexing, geographic information retrieval, perception distance, spatial query evaluation.

1 Introduction

The representation and management of Geographic Information is becoming a hot topic in the research area of information retrieval [9]. Given that about 15% of the queries submitted to search engines contain geographic names [20], the effective indexing and retrieval of geographical information poses new challenges to the design of *location-based search services* [18].

Location based search services allow users to search into specialized repositories, and more generally on the Internet, documents describing both resources and specific contents in the *neighbourhood of the user location*, or in the *neighbourhood of a geographic location* that is explicitly specified in the request. "*In the neighbourhood of a location*" specifies a spatial constraint on the geographic content of the retrieved documents, typically a constraint on the geographic distance from the desired location, like in the request "*find Indian restaurants near Bergamo university*".

Nevertheless, actual search engines do not enhance the influence of the geographical information content in evaluating requests containing geographic names. For example, the previous request submitted to *Google* retrieves as first ranked web page "*hotels near Indiana University East*" that does not satisfy the spatial constraint to be close to Bergamo. Generally, search engines are able to find specific resources such as hotels, restaurants, hospitals, in the neighbourhood of a locality, e.g. "*find hospital near Milan*", but this works only for specific resources, and moreover the ranking does not depend on the distance from the specified locality.

Furthermore, the indexing process does not extract most of these places automatically; consequently many web pages related to the query are not displayed as a result. In fact, many of the results obtained by a Google search are relative to resources that have directly informed Google of their geographical position. More specialized services has been developed for automatically mapping documents. For example, *MetaCarta* [11] is a system that geo-tags news, and allows searches in which one can distinctly specify the content constraint and the location of the news. However, this system is unable to retrieve news in the neighbourhood of a desired locality.

The identification of the geographic names, and successively the selection of the *geographic reference focus* (footprint) of web pages have been dealt with in several papers [1][2] [8][15]. Other papers considered the problem of spatial query evaluation [3][13][23]. In [15] the problem of implicit location identification is considered.

In this paper, after introducing the main problems involved in the design of a GIR system, we describe our GIR model and the system, named *Geo-Finder*, implementing it.

The main characteristics of the proposed geo-indexing model is the integration of multiple bipolar criteria satisfaction degrees [4][6][12][17][25], computed based on context dependent rules. Some criteria have a positive influence on the selection of the geographic names as footprints of the document; others have a negative influence. The positive and the negative constraints are heterogeneous and express pieces of information of a different nature [4]. The footprint of a document is represented as a fuzzy set of geographic coordinates (i.e., latitude and longitude, identifying the location and extent on the Earth surface of the geographic names occurring in the document) with membership degrees expressing their strength in defining the document footprint.

Second, we propose a distance measure to model the evaluation of the spatial query constraint, named "user's perception" distance. This distance measure depends on several aspects modelling the user context, such as the *spatial scope* of the query.

Finally, the system user interface maps the bi-dimensional relevance of retrieved documents in a Cartesian plane. The axes correspond to two orthogonal criteria

ISBN: 978-989-95079-6-8

(content constraint, geographic constraint), and the distance from the origin is inversely proportional to the global relevance of the documents. The user can also obtain a unique ranked list by choosing a trade-off of the two criteria satisfaction degrees based on their linear combination.

2 Why Geographic Information Retrieval is difficult

Geographic Information Retrieval (GIR) can be considered a specialized area of Information Retrieval, with an emphasis on the geographic indexing and geographic retrieval. GIR deals with any kind of information, i.e., not just maps or images but also texts, that have some relation to one or more locations on the Earth's surface, i.e., geo-referenced information [22]. Most of the information available on the Internet and in digital libraries is implicitly geo-referenced.

Often the link to the place (geographic footprint) is encoded by a geographic name. Geographic indexing implies the identification of the geographic names in a text and their translation into footprints, which are two operations that imply the management of imprecision, ambiguity, and incompleteness.

Often geographic names are ambiguous [14][21], some of them are homonymous of general terms (e.g. "*Los Angeles"*), some others identify distinct places on the Earth (e.g."*Rome*") or are temporal (e.g."*Leningrad*", "*Petersburg*", "*St Petersburg*"), or even are local names whose recognition relies on the knowledge of the local language.

Further, some pseudo-names are imprecise (e.g. "*around Milan*") or implicitly mentioned (e.g. "*the capital of Italy*"), or depending on the context (e.g. "*highest peak*" implicitly identifies "*Mont Blanc*" in a text describing the Alps). In [19] it has been argued that, given all these characteristics, geographic indexing can be feasibly faced by considering large corpora of geographic knowledge and heuristic rules.

On the other side, geographic retrieval implies being able to retrieve documents whose geographic focus satisfies a spatial constraint specified in a query. Generally, the spatial constraint demands the document footprint to be "*in the neighbourhood of a place*", i.e., it is a constraint on the distance. However, in this context the geographic distance is not merely Euclidean, but is related to the user's spatial context of interest, that can be related to the human perception of the time needed to cover it [5].

When a user searches resources or documents on the internet that are close to his/her current location, the judgment on the distance is related to his/her perception, depending on the fact that he/she is walking, driving or flying. Then, the constraint on the distance depends on the request and users' context.

Besides distance, other topological constraints could be defined, e.g., inclusion, overlapping, at the south/north of a specific region. Also these constraints must be interpreted in a tolerant, approximate way.

Then, geographic retrieval can draw benefits by defining the spatial constraints as soft, context dependent constraints, admitting degrees of satisfaction.

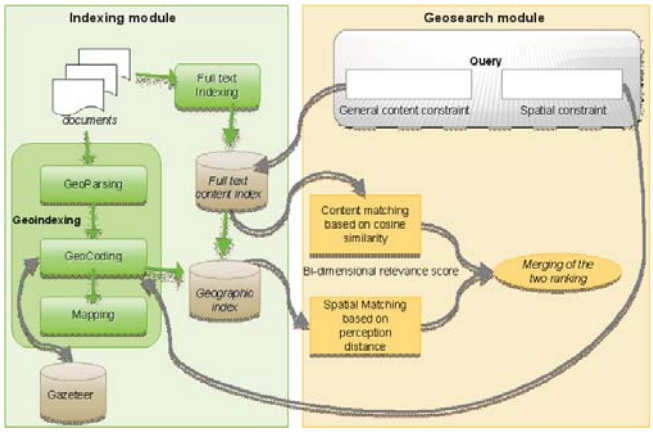


Figure1. Architecture of the Geographic Information Retrieval system *Geo-Finder*

3 The Architecture of the Geographic Information Retrieval System *Geo-Finder*

In this section we present the general architecture of the GIR system named *Geo-Finder*, that we designed and implemented. The architecture is depicted in Figure 1.

The system has the typical structure of an IRS, consisting of two main components: the *Indexing Module* and the *Retrieval Module* (named *GeoSearch*).

The Indexing Module has two main sub-modules: the *Full-Text Indexing sub-module* performs the full text indexing of the documents to represent their generic content, and generates the textual inverted index.

The *GeoIndexing sub-module* is the novel component, specialized in the identification of the footprints of documents, representing their geographic focus. This sub-module makes use of a *gazetteer* stored into a PostgreSQL database, containing the names of geographic entities (both administrative and physical entities) of all over the world, in English language and local languages (*GeoNames*) [10].

Once the *GeoIndexing* module has identified the footprints, it also stores them into the geographic index, that is a posting file containing the documents' identifiers and their footprints.

On the other side, the *GeoSearch* module interprets queries composed of two conditions. On the left hand side of the *GeoSearch* box in Figure 1, the generic content condition can be specified as a set of keywords, which define the *generic content constraint* on the documents' full text content representation. On the right hand side, the spatial condition can be specified as a geographic name, which is used to define the *spatial constraint* on the documents' footprints.

The *GeoIndexing* module parses the terms in the spatial condition to identify their footprint. These two conditions are evaluated by the sub-modules named *Content Matching module* (based on the *Lucene* library) and *Spatial Matching module*, based on perception distance (that is based on the original geo-retrieval model introduced in the next section), respectively.

In order to be retrieved, a document must satisfy, at least a little, the first content based condition (necessary condition), while the spatial condition is used for conditioning the ranking of the documents (optional

ISBN: 978-989-95079-6-8

condition). Then, the two conditions are *merged* by an "*and possibly*" aggregator, applied by the *Merging* sub-module that combines the satisfaction degrees of the two constraints, in order to obtain a unique value that performs the global ranking of documents. In the combination, one can give more preference to one of the two conditions, in order to emphasize the influence on the global ranking of either the document content or the document footprint.

4 The Geo-Retrieval Model

In this section, we describe the geo-retrieval model at the basis of the system implementation. In the first subsection the *GeoIndexing* model is described. In the second subsection, we introduce the geo-retrieval model at the basis of the *GeoSearch* component.

4.1 The GeoIndexing model

The GeoIndexing model is defined to identify for each document its footprint. A footprint of a document *d*, *Foot(d)*, is as a fuzzy set of geographic coordinates gc=(*lat,lon*), *lat*=latitude *lon*=longitude, expressed in degrees, with a membership degree $\mu_{Foot(d)}(gc)=GeoRef(gc) \in [0,1]$ representing the strength by which the geographic location gc, named gw, belongs to the footprint of the document d.

$$Foot(d)=\{GeoRef(gc_1)/gc_1,...,GeoRef(gc_n)/gc_n\}$$

A document is represented as a stream of tokens <*t*>. Some terms *t* have been selected as content indexes and thus are in the dictionary. For each of them, the frequency in the collection is known, while in the posting list we enter the documents in which they appear with their significance degrees *F(d,t)*. The significance degree is usually defined based on statistic analysis of the document text [16].

Besides this information, in the posting list we can also find the positions of the occurrences of the index terms in the document text, $occ_k(t_i,d)$, the k-th occurrence of t_i in *d*.

The identification of the document footprint is achieved in three steps, where distinct sets of heuristic rules are evaluated. Each set of rules acts as a filter on the input terms, so that only those terms whose global satisfaction degree of the set of rules is above a threshold are selected as input to the second step.

The *GeoIndexing* module implements this geoindexing model. It operates in two subsequent phases: first, it performs a G*eoParsing* that applies the first two sets of rules to detect the candidate geographic names (gw). Then, the *GeoCoding* sub-module identifies the document footprint *Foot(d)* and stores it in a file with the document unique identifier *d*.

The first set of rules consists of Name Entity Recognition (NER) rules, aimed at reducing the set of terms among which to successively select the candidate geographic names e.g.:

```
if Language(d)="English"∧
   FirstChar(t)=Capital then return(t).
```

The second set of rules receives, in input, a stream of previously selected terms, hereafter indicated by gw, and filters the candidate geographic names, a subset of the input terms. It applies bipolar [4] context dependent rules (r_i) exploiting a *gazetteer* [10] and computing independently a satisfaction degree *s(gw)* and a dissatisfaction degree *d(gw)* that denote to what extent gw is a geo-name and is not a geo-name, respectively.

The aggregation of the positive (negative) rules is done based on a Generalized Conjunction Disjunction function (GCD) [12], that, for distinct values of the parameter *p*, can model aggregations from completely compensative (or), where each rule can replace any other, to completely not compensative (and), where all rule must be satisfied simultaneously [6]:

$$s(gw)=\left(\sum_{i=1}^{m}\lambda_i*\left(r_i(gw)\right)^{ps}\right)^{1/ps}$$
$$d(gw)=\left(\sum_{j=1}^{n}\overline{\lambda}_j*\left(\overline{r}_j(gw)\right)^{pd}\right)^{1/pd} \quad (1)$$

The rules r_i and $\overline{r}_j$ assume a value in [0,1]. *ps* and *pd* are set so as to define (partially) compensative aggregations. In our experiment we used *ps=pd*=20, i.e., towards or like aggregation [6].

$\lambda_i,\overline{\lambda}_j \in [0,1]$ with $\sum_{i=1}^{m}\lambda_i=1, \sum_{i=1}^{n}\overline{\lambda}_j=1$ are the weights of the rules, i.e. their importance degrees in the aggregation, and are determined based on statistical analysis on a sample set of documents of a collection, and are set in a configuration file that is read by the *GeoParsing* sub-module during index generation. This way, the geoindexing can be suited to the characteristics of a collection.

The satisfaction of a rule r_i with *i=1,m* is interpreted as a hint of evidence that gw is a geographic name; thus the first *m* rules have a positive influence on the recognition of gw as a candidate geoname, like, e.g., the following two rules:

$$\text{If } gw \in gazetteer \wedge \exists gw_k \in d$$
$$\wedge\ gw_k = administrative_distr(gw) \Rightarrow r_2(gw)=0.5$$
$$\wedge\ |occ_i(gw,d)-occ_j(gw_k,d)|<\Delta \Rightarrow r_2(gw)=r_2(gw)+0.5$$

E.g. if gw="*San Francisco*" and gw_k="*California*", its administrative district, occurs in the same document *d* at a maximum distance $\Delta=3$ words, then $r_2(gw)=1$, otherwise if the occurrences are at a greater distance than Δ the rule is only partially satisfied, i.e., $r_2(gw)=0.5$.

Another rule is the following:

$$\text{If } gw \in gazetteer \wedge gw_{-1} \in prefix \Rightarrow r_3(gw)=1$$

e.g. if gw="*Blanc*" is preceded by gw_{-1}="*Mount*" that belongs to the set prefix of prefixes of geographic names such as *Mount*, *lake*, *city*, *river*, then $r_3(gw)=1$.

Conversely, the satisfaction of a rule $\overline{r}_j$ with *j=1,n* is interpreted as a hint of evidence that gw is not a geographic name. Thus, these rules have a negative influence, like the following one:

$$\text{If } gw \in gazetteer \wedge gw \in Stopwords \Rightarrow \overline{r}_4(gw)=1$$

ISBN: 978-989-95079-6-8

e.g. gw ="Nice" is a stop-word too.

For each input gw, a geo-score *GeoScore*(gw)∈[0,1] is computed based on the values of *s*(gw) and *d*(gw) as follows:

$$GeoScore(\mathtt{gw}) = \begin{cases} s(\mathtt{gw}) - d(\mathtt{gw}) & if\ s(\mathtt{gw}) > d(\mathtt{gw}) \\ 0 & if\ otherwise \end{cases}$$

We select the gw with *GeoScore*(gw)$>\tau>0$ as reliable geographic names. This threshold allows restricting the footprint of a document to reduce the possibility of identifying false positives. τ must be set based on experimentations, and is also specified in the configuration file. The greater it is, the smaller is the possibility of selecting false geo-names. Generally, it is better to lose some true geo-names than to select false ones.

The third set of rules, used in the third step by the *GeoCoding* sub-module, is aimed at identifying, from the reliable geographic names, the geographic locations that belong to the document footprint (i.e., the fuzzy set *Foot*(*d*) of pairs of geographic coordinates).

For each of the selected geo-names, gw, a (set) of pair(s) of geographic coordinates $GC_{gw}=\{gc_1,\ldots,gc_n\}$, with gc=(lat,lon), is retrieved from the gazetteer. Each pair gc is a geocode of gw, uniquely identifying a geographic place on the geographic domain. Notice that homonymous geo-names have the same name but distinct gc pairs.

For each geocode associated with a selected gw, a *geo-reference score (GeoRef*(gc)) ∈[0,1] is computed, that expresses the strength by which gw, located in gc, belongs to the document footprint (i.e. it is marginal or central in defining the geographic focus of the document).

A true geographic name can be correctly identified in a document, but it can be meaningless in defining the geographic focus of the document itself. Let us consider, for example, the geographic names that are often present at the very end of web pages or in their footnotes: they generally have nothing to do with the document content, but are related with the affiliation of the web master, and thus must not define the document footprint.

Also these rules have a satisfaction degree $r'_1(gc)\in$ [0,1] that is dependent on some variable. For example, the frequency F(*d*,gw) of a geoname gw in a document *d*, increases the strength of its geocodes in the footprint proportionally to the degree by which the geoname was recognized as a candidate geo-name *s*(gw), and inversely proportional to the degree by which it was recognized as not being a geo-name *d*(gw). This is represented by the following rule:

$$r'_1(gc) = \mu_{significant}(F(d,gw) * GeoScore(gw))$$

with $gc\in GC_{gw}$ and $\mu_{significant}$ is a monotonic non-decreasing membership function.

Another rule evaluates the presence of another geoname gw_k in *d*, whose geo-code is geographically near to that of gw (at a maximum geographic distance *dist* equal top *Δ*). When this occurs, it increases *GeoRef*(gc). That is:

$$r'_2(gc) = max(0, \Delta - dist(gc, gc_k))$$

with $gc\in GC_{gw}$, $gc_k\in GC_{gwk}$ and $gw_k \in d$.

The population attribute of an administrative name gw, that has multiple geocodes, indicated by *population*(gc), is used to favour the influence of big cities w.r.t. small ones having the same geoname in the footprint:

$$r'_3(gc) = (population(gc))/max_k(population(gc_k))$$

$gw \in gazetteer,\ gc, gc_k \in GC_{gw}$

This rule allows us to resolve ambiguities, generated by homonymous geographic names, that is, to favour the geographic coordinates (geo-codes) that most likely belong to the document footprint.

These rules are aggregated based on a GCD aggregation, in which each rule has a weight $\lambda'_i\in[0,1]$, with $\sum_{i=1}^{m}\lambda'_i = 1$, determined based on statistical analysis and the parameter *p*=1, neutral aggregation (all configurable):

$$GeoRef(gc) = \left(\sum_{i=1}^{m}\lambda'_i\left(r'_i(gc)\right)^p\right)^{1/p} \quad (2)$$

A minimum threshold *Φ*>0 on *GeoRef*(gc) (configurable) restricts the footprint of a document to reduce its extent. Therefore, at the end of this model, for each geoname we have two different scores: *GeoScore*(gw)$>\tau>0$ and *GeoRef*(gc)$>\Phi>0$. Figure 2 shows the geonames extracted from a sample paper. The map was generated by the *Mapping module*: this module creates distinct graphic metafiles for each document to map the footprint in distinct graphic environments, such as Gmaps for Google maps (see figure 2), KML for Google Earth, and GPX for GPS data format environments .



Score	GeoLocation	Name	Latitude	Longitude	Country code	Population
0,54	Milano	Milan	45.461216	9.187832	IT	1306661
1,00	Italy	Italian Republic	42.833332	12.833333	IT	58103033
0,87	Abano	Abano Terme	45.360992	11.791407	IT	18206
0,76	Ferrara	Ferrara	44.82678	11.620713	IT	130992
0,78	Vicenza	Vicenza	45.557293	11.540898	IT	107223
0,92	Tuscany	Tuscany	43.416668	11.0	IT	3573454
0,63	Apulia	Andria	41.23	16.29	IT	4021957

Figure 2. Google map of a document footprint, generated by *Geo-Finder* mapping module. The hue of the red pins represents the degree of the *geo-Reference* scores of the localities in the document footprint. A table with the geographic coordinates in the footprint and their country code and population attribute is also reported.

ISBN: 978-989-95079-6-8

4.2 The GeoRetrieval model

The GeoRetrieval model takes the footprint (*Foot(q)*) of the spatial query condition *q*, that consists of a single or a set of geographic names. The footprint *Foot(q),* if not directly found in the geographic index, is identified by applying the *GeoParsing* and *GeoCoding* rules described in the previous subsection.

For each document d that has been retrieved by the *Content-based Matching* module, we match its footprint, *Foot(d)* w.r.t. the query Footprint *Foot(q)* by applying formula (3). This matching function defines the semantics of the soft spatial constraint "*close*" that computes a degree of satisfaction $GRS(d) \in [0,1]$ (*GeoRelevanceScore*) as follows:

$$GRS(d) = \mu_{close}\left(Foot(d), Foot(q)\right) \tag{3}$$

$$\mu_{close}(Foot(d), Foot(q)) =$$

$$= \max_{i,j}\left[\mu_{Foot(d)}[i] * \mu_{Foot(q)}[j] * \left(1 - \frac{qscope(dist(i,j))}{Maxdist(Foot(d) \cup Food(q))}\right)\right]$$

with

$$\mu_{Foot(e)}[i] = GeoRef(gc_i) > 0$$

and

$$qscope(x) = \begin{cases} x & if \quad x \le \delta + k * Maxdist(Foot(q)) \\ & with \qquad \delta \ge 0 \quad k > 0 \\ 0 & otherwise \end{cases}$$

and *Maxdist(X)* is the maximum geographic distance (*dist*) of any two elements in the footprint *X*.

i and *j* represent the i-th and j-th pairs of geographic coordinates *latitude* and *longitude* in the footprints of the document *d,* with membership degrees *GeoRef*(*i*), and of the query *q*, with membership degrees *GeoRef*(*j*) that is assumed equal to 1.

The *qscope* function models the "user perception" distance measure, where *Maxdist(Foot(q))* is the query footprint maximum dispersion. δ and k are constant values that define the query scope and are set without user explicit input.

We consider four main query scopes, a *full* scope considering the whole globe, a *large* scope, considering an area covered by a continent or a big country like Russia, a *meso* scope considering an area of a nation or a big region, and a *small* scope considering a city and its surroundings. Each scope has specific values for the parameters (e.g. *k=3, δ=10km* is associated with a small scope, *k=30*, *δ=100km* with a *meso* scope).

δ is the query *range*, and is useful in the case of a query footprint consisting of a single geographic coordinate pair in order to retrieve also documents with footprint in the surrounding places. Distinct δ can adapt the evaluation of the spatial constraint "*close*" to the user perception; thus, modelling strict or relaxed interpretations of the *close* surrounding of a point. *k* allows to model a tolerance on the geographic distance between a document footprint and the query footprint that is equal to *k* times the query maximum dispersion, i.e., *Maxdist(Foot(q))*. This allows enlarging or reducing the query scope. This parameter can be related to the scale of the map needed to represent the minimum bounding box of the Minkowski sum of *Foot(q)* and a circle of radius *k* [7].

For example, if one specifies the two geonames *Bergamo*, *Como* (*Como* being at about 40km from Bergamo*)* as spatial condition, and the query scope is *small* (i.e. *k=3* and *δ=10km*) documents with footprints at a maximum distance of 130 km from the query footprint are retrieved (e.g. both documents in *Milano* and *Lugano* are retrieved while a document with a footprint in *Rome* is not).

On the other side, a query with footprint in *Bergamo, Dalmine* (10 *km* from *Bergamo* in *Milano* direction) will retrieve documents at a maximum distance from the query footprint of 40 km, (e.g. it will retrieve just the document in "*Milano*" and not the one in "*Lugano*").

Figure 3 depicts the *GeoSearch* user interface of the GIR system. At the top, there are two text forms for submitting the content (left) and spatial (right) query condition.

Bottom, on the right panel the bi-dimensional relevance domain is depicted, in with each point corresponds to a retrieved document. The origin identifies the query. The document's X coordinate (Y coordinate) is its relevance degree w.r.t. spatial (general content) query constraint. The closest the document is to the origin, the most relevant it is with respect to at least one query constraint.

On the left panel, the ordered list of retrieved documents is reported corresponding with a merging of the two relevance degrees giving equal importance to the two conditions. By moving the sliding bar at the top of this panel, it is possible to modify the preference between the two conditions and thus to re-rank the documents accordingly. This is achieved by a linear combination of the relevance rankings.

5 Conclusions

The system has undergone a first evaluation based on a collection of 1100 documents in *Italian* and *English* with an average length of 800 words. The collection comprehends research papers on geological studies carried out at IDPA CNR, Reuters news of the RCV1 collection, and web pages of the Open Directory Project.

This first evaluation was aimed at estimating the ability of the *GeoParser* to identify the correct geonames in documents text. For each document, we classified its geo-names and we compared the classification with respect to the footprints of the documents. We achieved a recall of 91% and a precision of 93%. Nevertheless, these results are preliminary, and further tests are needed to complete the evaluation.

The contributions of this proposal with respect to current practice are several: first of all, the computation of a fuzzy footprint to represent the geo-reference focus of a textual document, based on a bipolar criteria decision process; second, the use of a user's perception based distance measure at the basis of the computation of the degree of satisfaction of the spatial query constraint; third, the evaluation of a bi-dimensional relevance score, and the possibility to flexibly merge the two rankings into a single one by specifying a relative importance weight of the two query constraints.

ISBN: 978-989-95079-6-8

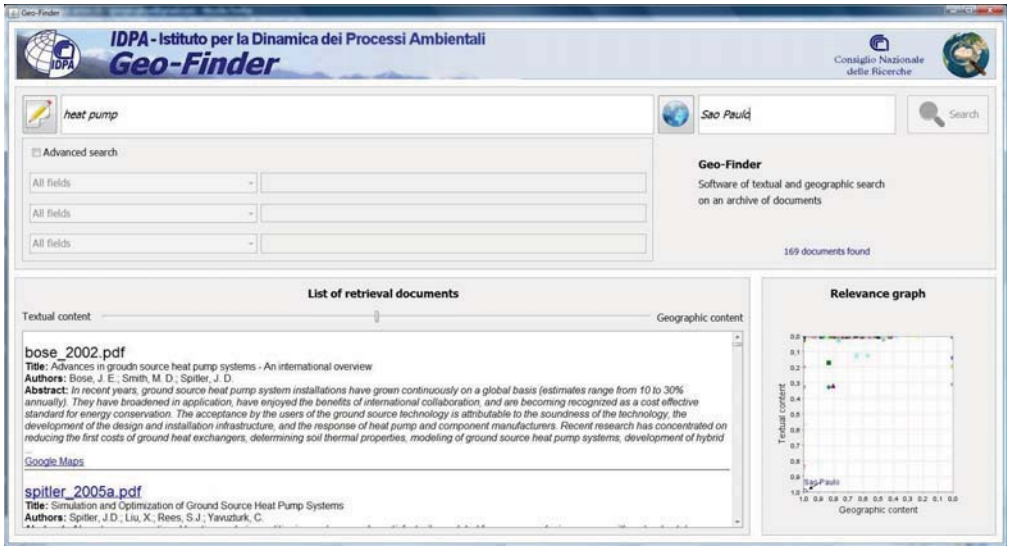


Figure 3. *GeoSearch* user interface of *Geo-Finder*: at the top the two query fields for specifying the content based condition="*heat pump*" and spatial condition="*Sao Paulo*"); below the two results panels. On the right side the bi-dimensional relevance graph, on the left the ranked list (merging with equal priority the two relevance scores).

References

[1] Amitay E., Har N., et al Web-a-where: Geotagging Web Content. SIGIR 2004.

[2] C. Wang, X. Xie et al., Detecting Geographical Locations from Web Resources. GIR'05, Bremen, Germany, 2005.

[3] Cai, G., "GeoVSM: An integrated Retrieval model for geographic information" M.J. Egenhofer and D.M. Marks Eds: GIScience 2002, LNCS 2478, Springer Verlag, 65-79, 2002.

[4] D. Dubois, H. Prade, An Introduction to Bipolar Representations of Information and Preference, Int. J. of Intelligent Systems, 23, 866–877, 2008.

[5] Egenhofer M., J., Mark D., M., Naïve Geography, eds Frank A.U., Kuhn W., Spatial Information Theory: a theoretical foundation for GIS . Springer Verlag LNCS 998, 1-15. 1995,

[6] G.Bordogna, G. Psaila, Modeling Soft Conditions with Unequal Importance in Fuzzy Databases based on the Vector p-norm, in Proc. of the IPMU Malaga, 2008.

[7] G.Bordogna, M. Pagani, G.Pasi, G. Psaila, Modeling uncertainty in Location-based spatial queries evaluation, Fuzzy Sets and Systems, accepted, september 2008.

[8] Hill L., "Introduction to Georeferencing in Digital Libraries Tutorial Document", University of California, Santa Barbara Aug. 2003.

[9] Himmelstein M.. Local search: The internet is the yellow pages. Computer, 38(2):26–34, 2005.

[10] http://www.library.uu.nl/kaartenzaal/Softwaresite/Beschrijvingen/geoname.html.

[11] http://www.metacarta.com/

[12] J.J. Dujmovic´, H. Larsen, Properties and modelingof partial conjunction/disjunction, in: B. De Baets (Ed.), Proceedings of Eurofuse Workshop on Data and Knowledge Engineering. Published as Current Issues in Data and Knowledge Engineering, Exit Publishers, Warsaw, 215–224, 2004.

[13] Jones et al, . "Spatial Information Retrieval and Geographical Ontologies: An Overview of the SPIRIT project" SIGIR 2002: Proc. of the 25th ACM SIGIR Conf.,Tampere, Finland', ACM Press, 387 – 388, 2002.

[14] Li H., Srihari R.K., Niu C., Li W., Infoxtract location normalization: a hybrid approach to geographic references in information extraction. In Proc. of the HLT-NAACL 2003 Workshop on Analysis of geographic references, 39-44.2003.

[15] Li Z., Wang C., Xie X., Wang X., Ma W.Y., Indexing implicit locations for geographical information retrieval, in Proc. of GIR'06, August 10–11, 2006, Seattle, USA, 2006.

[16] Lucene library http://lucene.apache.org/java/docs/

[17] Matthé T., De Tré G., (2009) Bipolar Query Satisfaction Using Satisfaction and Dissatisfaction Degrees, Proc. of the ACM SAC 2009, IAR track, 1699-1703.

[18] Mountain D. and MacFarlane A.. Geographic information retrieval in a mobile environment: evaluating the needs of mobile individuals. Journal of Information Science, 33(5):515–530, 2007.

[19] Reid J., A., "geoXwalk – A Gazetteer Server and Service for UK Academia" Koch/Sølvberg (Eds) Research and Advanced Technology for Digital Libraries: in Proc. of the 7th European Conf. ECDL 2003 Trondheim, Norway, August 17-22, 2003.

[20] Sanderson M. and Kohler J.. Analyzing geographic queries. In Workshop on Geographic Information Retrieval SIGIR, 2004.

[21] Smith D.A., Crane G., Disambiguating geographic names in historical digital library. In P. Costantoupolos and I. Solvberg eds, ECDL, LNCS 2163, Springer Verlag, 127-136, 2001.

[22] Vestavik Ø., Geographic Information Retrieval: an Overview, 2008.www.idi.ntnu.no/~oyvindve/article.pdf,

[23] Woodruff A.G., Plaunt C., GIPSY: Georeferenced Information Processing System, Journal of the American Society for Information Science, 1-21, 1994.

[24] Yasuda N., Toda H., Hirano T., Geographic Information Retrieval for just your surroundings, in Proc of the WWW08, April 21-25, Beijing China, 2008

[25] Zadrozny S., Kacprzyk J., (2007) Bipolar Queries Using Various Interpretations of Logical Connectives. IFSA (1), 181-190.

ISBN: 978-989-95079-6-8

A Semi-infinite Programming Approach to Possibilistic Optimization under Necessity Measure Constraints

Masahiro Inuiguchi[1]

1.Graduate School of Engineering Science, Osaka University
Toyonaka, Osaka, Japan
Email: inuiguti@sys.es.osaka-u.ac.jp

Abstract— *In this paper, possibilistic linear programming problems are investigated. After reviewing relations among conjunction and implication functions, necessity fractile optimization models with various implication functions are applied to the possibilistic linear problems. We show that the necessity fractile optimization models are reduced to semi-infinite linear programming problems. A simple numerical example is given to demonstrate the correctness of the result. The paper is concluded with some remarks for further developments.*

Keywords— Possibilistic linear programming, semi-infinite linear programming, necessity measure, implication function, conjunction function

1 Introduction

Fuzzy and possibilistic programming approaches are proposed to mathematical programming problems with ambiguity and vagueness [1, 2, 3]. By those approaches, we obtain reasonable solutions under conflicting soft constraints and goals, robust solutions under hard and soft constraints, hopeful solutions of attaining high-level goals, and so on. In possibilistic programming approaches, possibility and necessity measures are used to reduce the problems to the conventional programming problems. Many results demonstrate that possibilistic linear programming problems preserve the linearity in the reduced problems when possibility and necessity measures are defined by minimum operation and Dienes implication function. However, cases with the other conjunction and implication functions have not yet considerably investigated while several alternative approaches [4, 5] have been proposed in calculation of linear functions with fuzzy coefficients.

In this paper, we treat necessity fractile optimization model of possibilistic linear programming problems with soft constraints. We note that soft constraints include hard constraints as a special case. Then the problems are more general than problems with hard constraints. To construct necessity measures, we allow arbitrary implication satisfying weak requirements. Moreover, implication functions can be different among constraints. By the properties of implication functions, we show that necessity fractile optimization problems are reduced to a semi-infinite linear programming problems. Therefore the problems can be approximated by linear programming problems or the solutions can be obtained by a relaxation procedure together with the simplex method.

In next section, we briefly review the closure of generation procedures among conjunction and implication functions. Moreover, we introduce possibility and necessity measures. Possibilistic linear programming problems with soft constraints are given and reduced to the conventional programming problems through necessity fractile optimization models in Section 3. The main results are shown together with a conceivable solution procedure. in Section 4. In Section 5, a simple numerical example is given to demonstrate the correctness of the main result.

2 Conjunction and Implication Functions

2.1 Definitions

In this paper, a conjunction function is defined as a two-place function $T : [0,1] \times [0,1] \to [0,1]$ satisfying

(T0) T is lower semi-continuous, (semi-continuity)
(T1) $T(0,0) = T(0,1) = T(1,0) = 0$ and $T(1,1) = 1$, (boundary condition)
(T2) $T(a,b) \le T(c,d)$ if $0 \le a \le c \le 1$ and $0 \le b \le d \le 1$. (monotonicity)

A conjunction function T satisfies the following properties (t1) $T(a,1) = T(1,a) = a$ for any $a \in [0,1]$, (T3) $T(a,b) = T(b,a)$ for any $a, b \in [0,1]$ (commutativity) and (T4) $T(a,T(b,c)) = T(T(a,b),c)$ for any $a, b, c \in [0,1]$ (associativity) is known as a triangular norm.

In this paper, an implication function is a two-place function $I : [0,1] \times [0,1] \to [0,1]$ satisfying

(I0) I is upper semi-continuous, (semi-continuity)
(I1) $I(0,0) = I(0,1) = I(1,1) = 1$ and $I(1,0) = 0$, (boundary condition)
(I2) $I(a,b) \le I(c,d)$ if $0 \le c \le a \le 1$ and $0 \le b \le d \le 1$. (monotonicity)

Given a conjunction function and/or a strong negation n, an implication function can be generated through a transformation. The following three transformations are frequently adopted in the literature for implication generators:

$$I^{\mathrm{R}}[T](a,b) = \sup\{s \in [0,1] \mid T(a,s) \le b\}, \quad (1)$$

$$I^{\mathrm{S}}[T](a,b) = n(T(a,n(b))), \quad (2)$$

$$I^{\mathrm{r-R}}[T](a,b) = \sup\{s \in [0,1] \mid T(n(b),s) \le n(a)\}. \quad (3)$$

The first one, I^{R}, is encountered in the maximum solution of a fuzzy relation equation [6] and adopted in view of modus ponens. The second one, I^{S}, is introduced in analogy to Boolean logic. The last one, $I^{\mathrm{r-R}}$ is reciprocal to the first one which is obtained by taking a contraposition of the first one. When T is a t-norm, $I^{\mathrm{R}}[T]$, $I^{\mathrm{S}}[T]$ and $I^{\mathrm{r-R}}[T]$ are called R-implication (residual implication), S-implication and reciprocal R-implication, respectively. Whereas I^{S} produces an implication function from an arbitrary conjunction function T,

ISBN: 978-989-95079-6-8

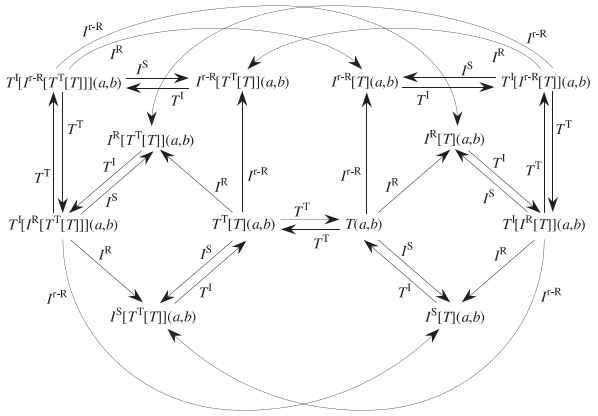


Figure 1: Conjunction and implication generation

I^{R} and $I^{\mathrm{r-R}}$ produce an implication function from a conjunction function which satisfies

$$T(1,a) > 0 \text{ for any } a > 0. \tag{4}$$

On the other hand, a conjunction function can be generated from an implication function through a transformation,

$$T^{\mathrm{I}}[I](a,b) = n(I(a,n(b))). \tag{5}$$

This transformation is symmetrical to I^{S}. From an implication function, a conjunction function is produced through T^{I}.

A conjunction function in this paper is not commutative. Thus, a new conjunction function may be generated from a conjunction function through

$$T^{\mathrm{T}}[T](a,b) = T(b,a). \tag{6}$$

2.2 Closure of generation processes

For the transformations (1)–(3), (5) and (6), the following equalities can be easily shown:

$$T^{\mathrm{I}} \circ I^{\mathrm{S}} = \mathrm{id.},\ I^{\mathrm{S}} \circ T^{\mathrm{I}} = \mathrm{id.},\ T^{\mathrm{T}} \circ T^{\mathrm{T}} = \mathrm{id.}, \tag{7}$$

$$I^{\mathrm{S}} \circ T^{\mathrm{T}} \circ T^{\mathrm{I}} \circ I^{\mathrm{R}} = I^{\mathrm{r-R}}, \tag{8}$$

where '$\circ$' denotes a composition, for example, $T^{\mathrm{I}} \circ I^{\mathrm{R}}$ is a composite transformation of I^{R} and T^{I}, i.e.,

$$T^{\mathrm{I}} \circ I^{\mathrm{R}}[T](a,b) = T^{\mathrm{I}}[I^{\mathrm{R}}[T]](a,b).$$

The notation 'id.' stands for the identical transformation.

From (T0), we have the following equalities [7]:

$$I^{\mathrm{R}} \circ T^{\mathrm{I}} \circ I^{\mathrm{R}}[T] = I^{\mathrm{S}}[T], \tag{9}$$

$$I^{\mathrm{r-R}} \circ T^{\mathrm{I}} \circ I^{\mathrm{R}} \circ T^{\mathrm{T}}[T] = I^{\mathrm{S}}[T], \tag{10}$$

$$I^{\mathrm{R}} \circ T^{\mathrm{I}} \circ I^{\mathrm{r-R}} \circ T^{\mathrm{T}}[T] = I^{\mathrm{r-R}}[T], \tag{11}$$

$$I^{\mathrm{r-R}} \circ T^{\mathrm{I}} \circ I^{\mathrm{r-R}} \circ T^{\mathrm{T}}[T] = I^{R}[T], \tag{12}$$

Equations (7)–(12) are summed up by Figure 1. As shown in Figure 1, the generation processes from a lower semi-continuous conjunction function as well as from an upper semi-continuous implication function are closed. This result is given by Inuiguchi and Sakawa [7] as the generalization of the result by Dubois and Prade [8]. Note that the semi-continuity is preserved through the generation processes [7].

As shown in Figure 1, we have six conjunction functions and six implication functions. The mappings among those twelve functions are given in Table 1. The inverse mappings are also indicated in Table 1. Some of those functions often appear in literatures on fuzzy relation equations, fuzzy logic, approximate reasoning and so on.

2.3 Possibility and necessity measures

Given a possible range V, the possibility and necessity measures $\Pi_V(S)$ and $N_V(S)$ of a fuzzy event S are defined by

$$\Pi_V(S) = \sup_{u \in U} T(\mu_V(u), \mu_S(u)), \tag{13}$$

$$N_V(S) = \inf_{u \in U} I(\mu_V(u), \mu_S(u)), \tag{14}$$

where μ_V and μ_S are membership functions of fuzzy sets V and S, respectively. U is the universal set. $\Pi_V(S)$ indicates the degree to what extent S is possible under possible range V, while $N_V(S)$ indicates the degree to what extent S is certain under possible range V.

Neither possibility measure nor necessity measure is unique. Therefore, we need to select possibility and necessity measures suitable for the given problem. By their definitions, selections of possibility and necessity measures are equivalent to selections of conjunction and implication functions. In the literature, minimum operation, i.e., $T(a,b) = \min(a,b)$ and Dienes implication, i.e., $I(a,b) = \max(1-a,b)$ are often used because they are simple and used in the original definitions of possibility and necessity measures [9, 10]. However, there is no guarantee that those possibility and necessity measures represent the decision maker's preferences well. Considering the variety of decision maker's preferences, the original possibility and necessity measures are not sufficiently flexible. Then the selections of possibility and necessity measures are important issues.

A conceivable approach to their selections are proposed by Inuiguchi and Tanino [11] and by Inuiguchi et al. [12]. The approach is based on the images of relations between two fuzzy sets represented by two modifier functions. Because of space limitations, we do not introduce this approach but we treat various kinds of necessity measures in the framework of possibilistic linear programming problems.

3 Linear Program with Necessity Measures

We consider a possibilistic linear programming problem,

$$\begin{array}{ll} \text{maximize} & \mathbf{c}^{\mathrm{T}}\mathbf{x}, \\ \text{subject to} & \mathbf{a}_i^{\mathrm{T}}\mathbf{x} \lesssim_i b_i,\ i = 1,2,\ldots,m, \\ & \mathbf{x} \geq \mathbf{0}, \end{array} \tag{15}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$ is a decision vector. b_i, $i = 1,2,\ldots,m$ are constants. Components c_j and a_{ij} of $\mathbf{c}$ and $\mathbf{a}_i$ are not known exactly but the possible ranges of those values are known as fuzzy numbers C_j and A_{ij}, respectively. A fuzzy number is a normal, convex and bounded fuzzy set on the real line whose membership function is upper semi-continuous. The notation $\lesssim_i$ is a fuzzified inequality so that $\lesssim_i\ b_i$ corresponds to a fuzzy set B_i with verbal expression 'a set of real numbers which are roughly smaller than b_i'. We assume that the membership function μ_{B_i} of B_i is non-increasing and upper semi-continuous and satisfies $\mu_{B_i}(b_i) = 1$.

ISBN: 978-989-95079-6-8

Table 1: Mappings among twelve conjunction and implication functions

conjunction → implication		
Φ	$\Phi(a,b)$	inverse
$I^{\mathrm{R}}[T]$	$\sup\{s\in[0,1] \mid T(a,s)\le b\}$	$T^{\mathrm{I}}[I^{\mathrm{R}}[T^{\mathrm{I}}[I]]]$
$I^{\mathrm{S}}[T]$	$n(T(a,n(b)))$	$T^{\mathrm{I}}[I]$
$I^{\mathrm{r-R}}[T]$	$\sup\{s\in[0,1] \mid T(n(b),s)\le n(a)\}$	$T^{\mathrm{I}}[I^{\mathrm{R}}[T^{\mathrm{T}}[T^{\mathrm{I}}[I]]]]$
$I^{\mathrm{R}}[T^{\mathrm{T}}[T]]$	$\sup\{s\in[0,1] \mid T(s,a)\le b\}$	$T^{\mathrm{I}}[I^{\mathrm{r-R}}[T^{\mathrm{I}}[I]]]$
$I^{\mathrm{S}}[T^{\mathrm{T}}[T]]$	$n(T(n(b),a))$	$T^{\mathrm{T}}[T^{\mathrm{I}}[I]]$
$I^{\mathrm{r-R}}[T^{\mathrm{T}}[T]]$	$\sup\{s\in[0,1] \mid T(s,n(b))\le n(a)\}$	$T^{\mathrm{I}}[I^{\mathrm{r-R}}[T^{\mathrm{T}}[T^{\mathrm{I}}[I]]]]$
implication → conjunction		
Φ	$\Phi(a,b)$	inverse
$T^{\mathrm{I}}[I]$	$n(I(a,n(b)))$	$I^{\mathrm{S}}[T]$
$T^{\mathrm{T}}[T^{\mathrm{I}}[I]]$	$n(I(b,n(a)))$	$I^{\mathrm{S}}[T^{\mathrm{T}}[T]]$
$T^{\mathrm{I}}[I^{\mathrm{R}}[T^{\mathrm{I}}[I]]]$	$\inf\{s\in[0,1] \mid I(a,s)\ge b\}$	$I^{\mathrm{R}}[T]$
$T^{\mathrm{I}}[I^{\mathrm{r-R}}[T^{\mathrm{I}}[I]]]$	$\inf\{s\in[0,1] \mid I(b,s)\ge a\}$	$I^{\mathrm{R}}[T^{\mathrm{T}}[T]]$
$T^{\mathrm{I}}[I^{\mathrm{R}}[T^{\mathrm{T}}[T^{\mathrm{I}}[I]]]]$	$\inf\{s\in[0,1] \mid I(n(s),n(a))\ge b\}$	$I^{\mathrm{r-R}}[T]$
$T^{\mathrm{I}}[I^{\mathrm{r-R}}[T^{\mathrm{T}}[T^{\mathrm{I}}[I]]]]$	$\inf\{s\in[0,1] \mid I(n(s),n(b))\ge a\}$	$I^{\mathrm{r-R}}[T^{\mathrm{T}}[T]]$
conjunction → conjunction		
Φ	$\Phi(a,b)$	inverse
$T^{\mathrm{T}}[T]$	$T(b,a)$	$T^{\mathrm{T}}[T]$
$T^{\mathrm{I}}[I^{\mathrm{R}}[T]]$	$n(\sup\{s\in[0,1] \mid T(a,s)\le n(b)\})$	$T^{\mathrm{I}}[I^{\mathrm{R}}[T]]$
$T^{\mathrm{I}}[I^{\mathrm{r-R}}[T]]$	$n(\sup\{s\in[0,1] \mid T(b,s)\le n(a)\})$	$T^{\mathrm{I}}[I^{\mathrm{R}}[T^{\mathrm{T}}[T]]]$
$T^{\mathrm{I}}[I^{\mathrm{R}}[T^{\mathrm{T}}[T]]]$	$n(\sup\{s\in[0,1] \mid T(s,a)\le n(b)\})$	$T^{\mathrm{I}}[I^{\mathrm{r-R}}[T]]$
$T^{\mathrm{I}}[I^{\mathrm{r-R}}[T^{\mathrm{T}}[T]]]$	$n(\sup\{s\in[0,1] \mid T(s,b)\le n(a)\})$	$T^{\mathrm{I}}[I^{\mathrm{r-R}}[T^{\mathrm{T}}[T]]]$
implication → implication		
Φ	$\Phi(a,b)$	inverse
$I^{\mathrm{R}}[T^{\mathrm{I}}[I]]$	$\sup\{s\in[0,1] \mid I(a,n(s))\ge n(b)\}$	$I^{\mathrm{R}}[T^{\mathrm{I}}[I]]$
$I^{\mathrm{r-R}}[T^{\mathrm{I}}[I]]$	$\sup\{s\in[0,1] \mid I(n(b),n(s))\ge a\}$	$I^{\mathrm{R}}[T^{\mathrm{T}}[T^{\mathrm{I}}[I]]]$
$I^{\mathrm{R}}[T^{\mathrm{T}}[T^{\mathrm{I}}[I]]]$	$\sup\{s\in[0,1] \mid I(s,n(a))\ge n(b)\}$	$I^{\mathrm{r-R}}[T^{\mathrm{I}}[I]]$
$I^{\mathrm{S}}[T^{\mathrm{T}}[T^{\mathrm{I}}[I]]]$	$I(n(b),n(a))$	$I^{\mathrm{S}}[T^{\mathrm{T}}[T^{\mathrm{I}}[I]]]$
$I^{\mathrm{r-R}}[T^{\mathrm{T}}[T^{\mathrm{I}}[I]]]$	$\sup\{s\in[0,1] \mid I(s,b)\ge a\}$	$I^{\mathrm{r-R}}[T^{\mathrm{T}}[T^{\mathrm{I}}[I]]]$

By the extension principle, the possible ranges of $\mathbf{c}^{\mathrm{T}}\mathbf{x}$ and $\mathbf{a}_i^{\mathrm{T}}\mathbf{x}$ are obtained as fuzzy sets $\mathbf{C}^{\mathrm{T}}\mathbf{x}$ and $\mathbf{A}_i^{\mathrm{T}}\mathbf{x}$, respectively, defined by the following membership functions:

$$\mu_{\boldsymbol{C}^{\mathrm{T}}\boldsymbol{x}}(y)=\sup_{\substack{r_1,\ldots,r_n\\ \boldsymbol{r}^{\mathrm{T}}\boldsymbol{x}=y}}\min\left(\mu_{C_1}(r_1),\ldots,\mu_{C_n}(r_n)\right), \quad (16)$$

$$\mu_{\boldsymbol{A}_i^{\mathrm{T}}\boldsymbol{x}}(y)=\sup_{\substack{r_1,\ldots,r_n\\ \boldsymbol{r}^{\mathrm{T}}\boldsymbol{x}=y}}\min\left(\mu_{A_{i1}}(r_1),\ldots,\mu_{A_{in}}(r_n)\right), \quad (17)$$

where $\mathbf{r}=(r_1,\ldots,r_n)^{\mathrm{T}}$.

Using a necessity measure N^i defined by an upper semi-continuous implication function I^i, in this paper, we formulate Problem (15) as a necessity fractile optimization model (see Inuiguchi and Ramík [3]):

$$\begin{array}{ll}\text{maximize} & q,\\ \text{subject to} & N^0_{\boldsymbol{C}^{\mathrm{T}}\boldsymbol{x}}([q,+\infty))\ge h^0,\\ & N^i_{\boldsymbol{A}_i^{\mathrm{T}}\boldsymbol{x}}(B_i)\ge h^i,\ i=1,2,\ldots,m,\\ & \mathbf{x}\ge\mathbf{0},\end{array} \quad (18)$$

where q is an auxiliary variable. $h^0\in(0,1]$ and $h^i\in(0,1]$, $i=1,2,\ldots,m$ are certainty levels of goal achievement and constraint satisfactions specified by the decision maker. Note that we obtain

$$N^0_{\boldsymbol{C}^{\mathrm{T}}\boldsymbol{x}}([q,+\infty))=\inf_{r<q}I^0(\mu_{\boldsymbol{C}^{\mathrm{T}}\boldsymbol{x}}(r),0), \quad (19)$$

$$N^i_{\boldsymbol{A}_i^{\mathrm{T}}\boldsymbol{x}}(B_i)=\inf_{r}I^i(\mu_{\boldsymbol{A}_i^{\mathrm{T}}\boldsymbol{x}}(r),\mu_{B_i}(r)). \quad (20)$$

The selections of necessity measures and certainty levels depend on the required robustness of goal achievement/constraint satisfactions, the meanings of total goal achievement/constraint satisfactions, the estimation of fuzzy coefficients and so on. The method proposed by Inuiguchi and Tanino [11] and by Inuiguchi et al. [12] would be useful for selecting suitable necessity measures.

Let $[S]_h$ be a h-level set of a fuzzy set S, i.e., $[S]_h=\{u\in U \mid \mu_S(u)\ge h\}$. Then, because fuzzy numbers C_i and A_{ij} are bounded and have upper semi-continuous mem-

ISBN: 978-989-95079-6-8

bership functions, we have (see Dubois and Prade [13])

$$[\mathbf{C}^{\mathrm{T}}\mathbf{x}]_h = \sum_{j=1}^{n} [C_j]_h x_j, \tag{21}$$

$$[\mathbf{A}_i^{\mathrm{T}}\mathbf{x}]_h = \sum_{j=1}^{n} [A_{ij}]_h x_j. \tag{22}$$

Let $c_j^{\mathrm{L}}(h) = \inf[C_j]_h$, $c_j^{\mathrm{R}}(h) = \sup[C_j]_h$, $a_{ij}^{\mathrm{L}}(h) = \inf[A_{ij}]_h$ and $a_{ij}^{\mathrm{R}}(h) = \sup[A_{ij}]_h$. Then, considering the non-negativity of $\mathbf{x}$, we have

$$[\mathbf{C}^{\mathrm{T}}\mathbf{x}]_h = \left[\sum_{j=1}^{n} c_j^{\mathrm{L}}(h)x_j, \sum_{j=1}^{n} c_j^{\mathrm{R}}(h)x_j \right], \tag{23}$$

$$[\mathbf{A}_i^{\mathrm{T}}\mathbf{x}]_h = \left[\sum_{j=1}^{n} a_{ij}^{\mathrm{L}}(h)x_j, \sum_{j=1}^{n} a_{ij}^{\mathrm{R}}(h)x_j \right]. \tag{24}$$

To reduce Problem (18), the following theorem is useful.

Theorem 1. *Let N^i be a necessity measure defined by an implication function I^i. Then for any fuzzy sets V and S of a universal set U, we have*

$$\begin{aligned} & N_V^i(S) \geq h \\ & \Leftrightarrow \forall u \in U,\ \forall k \in [0,1]; \\ & \quad \mu_V(u) \geq k \text{ implies } \mu_S(u) \geq T^{\mathrm{I}}[I^{\mathrm{R}}[T^{\mathrm{I}}[I^i]]](k,h), \\ & \Leftrightarrow \forall k \in [0,1];\ [V]_k \subseteq [S]_{f^i(k,h)}, \end{aligned} \tag{25}$$

where $f^i(k,h) = T^{\mathrm{I}}[I^{\mathrm{R}}[T^{\mathrm{I}}[I^i]]](k,h)$.

(**Proof**) Because of the upper semi-continuity of I^i, Figure 1 and Table 1, we have

$$\begin{aligned} & N_V^i(S) = \inf_{u\in U} I^i(\mu_V(u), \mu_S(u)) \geq h \\ & \Leftrightarrow \forall u \in U;\ I^i(\mu_V(u), \mu_S(u)) \geq h \\ & \Leftrightarrow \forall u \in U;\ \mu_V(u) \leq \sup\{s \in [0,1] \mid I^i(s, \mu_S(u)) \geq h\} \\ & \Leftrightarrow \forall u \in U;\ \mu_V(u) \leq I^{\mathrm{r-R}}[T^{\mathrm{T}}[T^{\mathrm{I}}[I^i]]](h, \mu_S(u)) \\ & \Leftrightarrow \forall u \in U,\ \forall k \in [0,1]; \\ & \quad \mu_V(u) \geq k \text{ implies } I^{\mathrm{r-R}}[T^{\mathrm{T}}[T^{\mathrm{I}}[I^i]]](h, \mu_S(u)) \geq k \\ & \Leftrightarrow \forall u \in U,\ \forall k \in [0,1];\ \mu_V(u) \geq k \text{ implies} \\ & \quad \mu_S(u) \geq \inf\{s \in [0,1] \mid I^{\mathrm{r-R}}[T^{\mathrm{T}}[T^{\mathrm{I}}[I^i]]](h,s) \geq k\} \\ & \Leftrightarrow \forall u \in U,\ \forall k \in [0,1];\ \mu_V(u) \geq k \text{ implies} \\ & \quad \mu_S(u) \geq T^{\mathrm{I}}[I^{\mathrm{r-R}}[T^{\mathrm{I}}[I^{\mathrm{r-R}}[T^{\mathrm{T}}[T^{\mathrm{I}}[I^i]]]]]](k,h) \\ & \Leftrightarrow \forall u \in U,\ \forall k \in [0,1];\ \mu_V(u) \geq k \text{ implies} \\ & \quad \mu_S(u) \geq T^{\mathrm{I}}[I^{\mathrm{R}}[T^{\mathrm{I}}[I^i]]](k,h). \end{aligned}$$

(Q.E.D.)

From the assumptions of B_i, we have $[B_i]_k = (-\infty, \bar{b}_i(k)]$, where $\bar{b}_i : [0,1] \to [0,+\infty)$ is defined by $\bar{b}_i(k) = \sup\{r \mid \mu_{B_i}(r) \geq k\}$. From (23), (24) and Theorem 1, Problem (18) is reduced to the following linear semi-infinite programming problem:

$$\begin{aligned} \text{maximize} \quad & q, \\ \text{subject to} \quad & \sum_{j=1}^{n} c_j^{\mathrm{L}}(k)x_j \geq \bar{q}(f^0(k,h^0)),\ \forall k \in [0,1], \\ & \sum_{j=1}^{n} a_{ij}^{\mathrm{R}}(k)x_j \leq b_i(f^i(k,h^i)),\ \forall k \in [0,1], \\ & \qquad\qquad i = 1,2,\ldots,m, \\ & \mathbf{x} \geq \mathbf{0}, \end{aligned} \tag{26}$$

where $\bar{q} : [0,1] \to \{-\infty, q\}$ is defined by

$$\bar{q}(k) = \begin{cases} q, & \text{if } k > 0, \\ -\infty, & \text{if } k = 0. \end{cases} \tag{27}$$

We showed that fractile optimization models based on necessity measures defined by any implication functions of possibilistic linear programming problems are reduced to linear semi-infinite programming problems. Thus, we can solve the problems by using linear semi-infinite programming techniques [14]. For example, let $\varepsilon > 0$ be a sufficiently small number and let $k_v = (v-1)(1-\varepsilon)/(l-1)+\varepsilon$, $v = 1,2,\ldots,l$. Then Problem (26) can be approximated by the following linear programming problem:

$$\begin{aligned} \text{maximize} \quad & q, \\ \text{subject to} \quad & \sum_{j=1}^{n} c_j^{\mathrm{L}}(k_v)x_j \geq \bar{q}(f^0(k_v,h^0)),\ v = 1,\ldots,l, \\ & \sum_{j=1}^{n} a_{ij}^{\mathrm{R}}(k_v)x_j \leq b_i(f^i(k_v,h^i)), \\ & \qquad\qquad v = 1,\ldots,l,\ i = 1,\ldots,m, \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned} \tag{28}$$

Let $M = \{1,2,\ldots,m\}$. As another approach to computation of an approximate solution to Problem (26), the following algorithm based on a relaxation procedure is conceivable:

Algorithm

S1. Let $z = 0$ and $\varepsilon > 0$ be a sufficiently small positive number. Solve the following linear programming problem:

$$\begin{aligned} & \text{maximize } q, \\ & \text{subject to} \\ & \sum_{j=1}^{n} c_j^{\mathrm{L}}(k_0)x_j \geq \bar{q}(f^0(k_0,h^0)), k_0 = \varepsilon, 1, \\ & \sum_{j=1}^{n} a_{ij}^{\mathrm{R}}(k_i)x_j \leq b_i(f^i(k_i,h^i)),\ k_i = \varepsilon, 1,\ i \in M, \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned}$$

Let $\mathbf{x}^z = (x_1^z,\ldots,x_n^z)^{\mathrm{T}}$ be the obtained optimal solution and q^z be the optimal value.

S2. Calculate $d_0 = \inf_{r\in[0,1]} \sum_{j=1}^{n} c_j^{\mathrm{L}}(r)x_j^z - \bar{q}(f^0(r,h^0))$. If $d_0 < 0$, we define $k_0^z = \arg\inf_{r\in[0,1]} \sum_{j=1}^{n} c_j^{\mathrm{L}}(r)x_j^z - \bar{q}(f^0(r,h^0))$ and otherwise, $k_0^z = 0$.

S3. For $i = 1,2,\ldots,m$, the following processes (a) and (b) are applied. (a) calculate $d_i = \sup_{r\in[0,1]} \sum_{j=1}^{n} a_{ij}^{\mathrm{R}}(r)x_j^z - b_i(f^i(r,h^i))$. (b) If $d_0 > 0$, we define $k_0^z = \arg\sup_{r\in[0,1]} \sum_{j=1}^{n} a_{ij}^{\mathrm{R}}(r)x_j^z - b_i(f^i(r,h^i))$ and otherwise, $k_i^z = 0$.

ISBN: 978-989-95079-6-8

Table 2: Fuzzy numbers and constraints

$C_1 = \langle 5, 1\rangle$	$C_2 = \langle 7, 0.7\rangle$	
$A_{11} = \langle 2, 0.7\rangle$	$A_{12} = \langle 3, 0.5\rangle$	$B_1 = (230, 60\rangle$
$A_{21} = \langle 4, 1.5\rangle$	$A_{22} = \langle 2, 0.3\rangle$	$B_2 = (370, 100\rangle$
$A_{31} = \langle 1, 0.5\rangle$	$A_{32} = \langle 3, 0.3\rangle$	$B_3 = (140, 90\rangle$

S4. If $\sum_{i=0}^{m} k_i^z = 0$, terminate the algorithm. An approximate solution is obtained as $\mathbf{x}^z$ and q^z.

S5. Let $z = z + 1$. Solve the following linear programming problem:

$$
\begin{aligned}
&\text{maximize } q,\\
&\text{subject to}\\
&\sum_{j=1}^{n} c_j^{\mathrm{L}}(k_0)x_j \geq \bar{q}(f^0(k_0, h^0)), k_0 = \varepsilon, 1,\\
&\sum_{j=1}^{n} c_j^{\mathrm{L}}(k_0^w)x_j \geq \bar{q}(f^0(k_0^w, h^0)), w = 1, \ldots, z\\
&\sum_{j=1}^{n} a_{ij}^{\mathrm{R}}(k_i)x_j \leq b_i(f^i(k_i, h^i)),\ k_i = \varepsilon, 1,\ i \in M,\\
&\sum_{j=1}^{n} a_{ij}^{\mathrm{R}}(k_i^w)x_j \leq b_i(f^i(k_i^w, h^i)),\ w = 1, \ldots, z,\ i \in M,\\
&\mathbf{x} \geq \mathbf{0}.
\end{aligned}
$$

Let $\mathbf{x}^z = (x_1^z, \ldots, x_n^z)^{\mathrm{T}}$ be the obtained optimal solution and q^z be the optimal value. Go to S2.

4 A Numerical Example

In order to demonstrate that Problem (26) is equivalent to Problem (18), in other words, an optimal solution to Problem (28) is an approximate solution to Problem (18), we consider the possibilistic linear programming problem (15) with $m = 3$. Fuzzy numbers C_j and A_{ij} are given by symmetric triangular fuzzy numbers whose center values and spreads are shown in Table 2, where a symmetric triangular fuzzy number Q=$\langle q^{\mathrm{C}}, q^{\mathrm{S}}\rangle$ ($q^{\mathrm{S}} > 0$) is characterized by a membership function,

$$
\mu_Q(r) = \begin{cases} 1 - \dfrac{|r - q^{\mathrm{C}}|}{q^{\mathrm{S}}} & \text{if } r \in [q^{\mathrm{C}} - q^{\mathrm{S}}, q^{\mathrm{C}} + q^{\mathrm{S}}],\\ 0 & \text{otherwise.}\end{cases} \tag{29}
$$

As shown in Table 2, fuzzy constraints B_i are defined by fuzzy sets $B_i = (b_i^{\mathrm{C}}, b_i^{\mathrm{S}}\rangle$ with a linear membership function,

$$
\mu_{B_i}(r) = \begin{cases} 1 & \text{if } r < b_i^{\mathrm{C}},\\ 1 - \dfrac{r - b_i^{\mathrm{C}}|}{b_i^{\mathrm{S}}} & \text{if } r \in [b_i^{\mathrm{C}}, b_i^{\mathrm{C}} + b_i^{\mathrm{S}}],\\ 0 & \text{otherwise.}\end{cases} \tag{30}
$$

We apply Problem (18) with reciprocal Goguen implication $I^0(a, b) = \min(1, (1-a)/(1-b))$ with definition $(1-a)/0 = +\infty$ for any $a \leq 1$, Lukasiewicz implication $I^1(a, b) = \min(1, 1-a+b)$, Reichenbach implication $I^2(a, b) = 1-a+ab$ and Goguen implication $I^3(a, b) = \min(1, b/a)$ with definition $b/0 = +\infty$ for any $b \geq 0$. Then, for $h > 0$ and $k > 0$, we obtain $f_0(k, h) = \max(0, 1 - (1 - k)/h)$, $f_1(k, h) = \max(0, k + h - 1)$, $f_2(k, h) = \max(0, (k + h - 1)/k)$ and $f_3(k, h) = kh$. Moreover, we set $h^0 = 0.5$, $h^1 = 0.5$, $h^2 = 0.62$ and $h^3 = 0.62$.

Let $\varepsilon = 0.0000001$ and $l = 20$, we solve Problem (28), i.e.,

$$
\begin{aligned}
&\text{maximize } q,\\
&\text{subject to}\\
&4.500x_1 + 6.650x_2 \geq q, && 4.550x_1 + 6.685x_2 \geq q,\\
&4.600x_1 + 6.720x_2 \geq q, && 4.650x_1 + 6.755x_2 \geq q,\\
&4.700x_1 + 6.790x_2 \geq q, && 4.750x_1 + 6.825x_2 \geq q,\\
&4.800x_1 + 6.860x_2 \geq q, && 4.850x_1 + 6.895x_2 \geq q,\\
&4.900x_1 + 6.930x_2 \geq q, && 4.950x_1 + 6.965x_2 \geq q,\\
&5x_1 + 7x_2 \geq q, && 2.350x_1 + 3.250x_2 \leq 290.00,\\
&2.315x_1 + 3.225x_2 \leq 287.00, && 2.280x_1 + 3.200x_2 \leq 284.00,\\
&2.245x_1 + 3.175x_2 \leq 281.00, && 2.210x_1 + 3.150x_2 \leq 278.00,\\
&2.175x_1 + 3.125x_2 \leq 275.00, && 2.140x_1 + 3.100x_2 \leq 272.00,\\
&2.105x_1 + 3.075x_2 \leq 269.00, && 2.070x_1 + 3.050x_2 \leq 266.00,\\
&2.035x_1 + 3.025x_2 \leq 263.00, && 2x_1 + 3x_2 \leq 260,\\
&4.900x_1 + 2.180x_2 \leq 465.00, && 4.825x_1 + 2.165x_2 \leq 454.44,\\
&4.750x_1 + 2.150x_2 \leq 446.00, && 4.675x_1 + 2.135x_2 \leq 439.09,\\
&4.600x_1 + 2.120x_2 \leq 433.33, && 4.525x_1 + 2.105x_2 \leq 428.46,\\
&4.450x_1 + 2.090x_2 \leq 424.29, && 4.375x_1 + 2.075x_2 \leq 420.67,\\
&4.300x_1 + 2.060x_2 \leq 417.50, && 4.225x_1 + 2.045x_2 \leq 414.71,\\
&4.150x_1 + 2.030x_2 \leq 412.22, && 4.075x_1 + 2.015x_2 \leq 410.00,\\
&4x_1 + 2x_2 \leq 408, && 1.500x_1 + 3.300x_2 \leq 230.00,\\
&1.475x_1 + 3.285x_2 \leq 227.21, && 1.450x_1 + 3.270x_2 \leq 224.42,\\
&1.425x_1 + 3.255x_2 \leq 221.63, && 1.400x_1 + 3.240x_2 \leq 218.84,\\
&1.375x_1 + 3.225x_2 \leq 216.05, && 1.350x_1 + 3.210x_2 \leq 213.26,\\
&1.325x_1 + 3.195x_2 \leq 210.47, && 1.300x_1 + 3.180x_2 \leq 207.68,\\
&1.275x_1 + 3.165x_2 \leq 204.89, && 1.250x_1 + 3.150x_2 \leq 202.10,\\
&1.225x_1 + 3.135x_2 \leq 199.31, && 1.200x_1 + 3.120x_2 \leq 196.52,\\
&1.175x_1 + 3.105x_2 \leq 193.73, && 1.150x_1 + 3.090x_2 \leq 190.94,\\
&1.125x_1 + 3.075x_2 \leq 188.15, && 1.100x_1 + 3.060x_2 \leq 185.36,\\
&1.075x_1 + 3.045x_2 \leq 182.57, && 1.050x_1 + 3.030x_2 \leq 179.78,\\
&1.025x_1 + 3.015x_2 \leq 176.99, && 1x_1 + 3x_2 \leq 174.2,\\
&x_1 \geq 0,\ x_2 \geq 0.
\end{aligned}
\tag{31}
$$

Solving this problem, we obtain an optimal solution $\mathbf{x}^* = (x_1^*, x_2^*) = (79.508, 31.564)$ with the optimal value $q^* = 567.69$. As shown in Figures 2∼5, we can observe that the obtained solution approximately satisfies constraints, $N^0_{\boldsymbol{C}^{\mathrm{T}}\boldsymbol{x}}([q^*, +\infty)) \geq 0,5$ and $N^i_{\boldsymbol{A}_i^{\mathrm{T}}\mathbf{x}}(B_i) \geq h^i$, $i = 1, 2, 3$.

5 Concluding Remarks

Utilizing the closure of generation processes of conjunction and implication functions, we have shown that necessity fractile optimization models of possibilistic linear programming problems are reduced to semi-infinite linear programming problems. Therefore, we can solve the problems approximately by linear programming techniques. As a solution procedure for the reduced semi-infinite linear programming problem, we have described a relaxation procedure. In order to demonstrate the main result, we have given a simple numerical example. In the example, the reduced semi-infinite linear programming problem is approximated by a linear programming problem having coarsely-sampled constraints. It has been confirmed that the obtained solution approximately satisfies the required constraints on necessity measures by figures.

The investigation on possibilistic linear programming problems with necessity measures is important to have flexible

ISBN: 978-989-95079-6-8

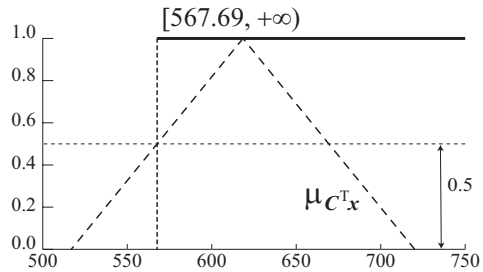


Figure 2: The relation between $\mathbf{C}^{\mathrm{T}}\mathbf{x}^*$ and the optimal value

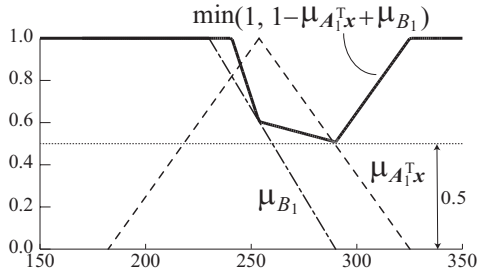


Figure 3: The satisfaction of $N^1_{\boldsymbol{A}_1^{\mathrm{T}}\mathbf{x}}(B_1) \geq 0.5$

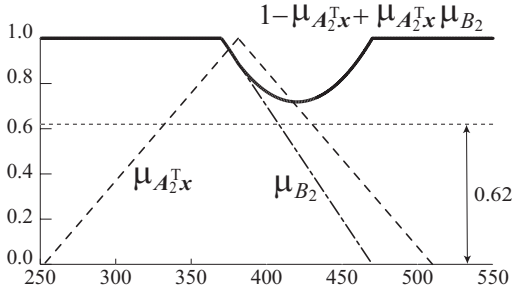


Figure 4: The satisfaction of $N^2_{\boldsymbol{A}_2^{\mathrm{T}}\mathbf{x}}(B_2) \geq 0.62$

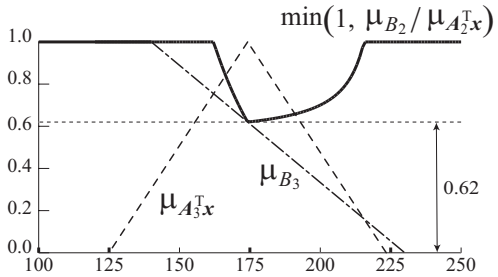


Figure 5: The satisfaction of $N^3_{\boldsymbol{A}_3^{\mathrm{T}}\mathbf{x}}(B_3) \geq 0.62$

models of robust optimization reflecting decision makers' various attitudes toward the uncertainty. Fortunately, models with necessity measures are often easier than models with possibility measures. Analyzing some special implication functions defining necessity measures and special shapes of fuzzy coefficients, the reduced semi-infinite constraints can further reduced to finite constraints. Moreover, we may solve necessity measure optimization models of possibilistic linear programming problems by the simplex method together with the bisection method. These would be future topics in our studies.

Acknowledgment

The author acknowledges that this work has been partially supported by the Grant-in-Aid for Scientific Research (B) No. 17310098, and thanks Yuji Matsumoto (Osaka University) for his help in solving a linear programming problem.

References

[1] M. Inuiguchi, H. Ichihashi and H.Tanaka. Fuzzy programming: A survey of recent developments. in: R. Słowiński and J. Teghem (eds.). *Stochastic versus Fuzzy Approaches to Multiobjective Mathematical Programming under Uncertainty*. Kluwer Academic Publishers, Dordrecht, 45–68, 1990.

[2] H. Rommelfanger and R. Słowiński. Fuzzy linear programming with single or multiple objective functions. in: R. Słowiński (ed.). *Fuzzy Sets in Decision Analysis, Operations Research and Statistics*. Kluwer Academic Publishers, Boston, 179–213, 1998.

[3] M. Inuiguchi and J. Ramík. Possibilistic linear programming: A brief review of fuzzy mathematical programming and a comparison with stochastic programming in portfolio selection problem. *Fuzzy Sets and Systems*, 111: 29–45, 2000.

[4] H. Rommelfanger and T. Keresztfalvi. Multicriteria fuzzy optimization based on Yager's parameterized t-norm. *Foundations of Computing and Decision Sciences*, 16: 99–110, 1991.

[5] M. Inuiguchi. Linear programming with interactive fuzzy numbers. *Proceedings of 40th Annual Convention of ORSI*, 27–37, 2007.

[6] E. Sanchez. Resolution of composite fuzzy relation equations. *Information and Control*, 30: 38–48, 1976.

[7] M. Inuiguchi and M. Sakawa. On the closure of generation processes of implication functions from a conjunction function. in: T. Yamakawa and G. Matsumoto (eds.). *Methodologies for the Conception, Design, and Application of Intelligent Systems*. World Scientific, Singapore, 1: 327–330, 1996.

[8] D. Dubois and H. Prade. A theorem on implication functions defined from triangular norms. *Stochastica*, 8: 267–279, 1984.

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

[10] D. Dubois and H. Prade. *Fuzzy Sets and Systems: Theory and Applications*. Academic Press, New York, 1980.

[11] M. Inuiguchi and T. Tanino. Necessity measures and parametric inclusion relations of fuzzy sets. *Fundamenta Informaticae*. 42: 279–302, 2000.

[12] M. Inuiguchi, S. Greco, R. Słowiński and T. Tanino. Possibility and necessity measure specification using modifiers for decision making under fuzziness. *Fuzzy Sets and Systems*. 137: 151–175, 2003.

[13] D. Dubois and H. Prade. Fuzzy numbers: An overview. in: J.C. Bezdek (ed.). *Analysis of Fuzzy Information, Vol.I: Mathematics and Logic*. CRC Press, Boca Raton, FL, pp.3–39, 1987.

[14] M.A. Goberna and M.A. López. *Linear Semi-infinite Optimization*. John Wiley & Sons, Chichester, 1998.

ISBN: 978-989-95079-6-8

On Aristotle's NC and EM Principles in Three-valued Logics*

E. Trillas, C. Moraga, E. Renedo†

European Centre for Soft Computing
Mieres (Asturias), Spain
Email: {enric.trillas, claudio.moraga, eloy.renedo }@softcomputing.es.

Abstract— *By interpreting* 'p *is* impossible' *by* 'p *is* self-contradictory'*, and* 'p *is* always' *by* 'not p *is* self-contradictory'*, this paper studies which of the three-valued systems of Łucasiewicz, Gödel, Kleene, Bochvar, and Post, do verify the Aristotle's principles of Non-Contradiction (NC), and Excluded-Middle (EM).*

Keywords— Systems of three-valued logic, Non-Contradiction, Excluded-Middle.

1 Introduction

As it is well known, the systems of multiple-valued logic do not verify the *principles of non-contradiction* and *excluded-middle*, once presented in the forms (see [5])

$$a \cdot a' = 0,\ a + a' = 1.$$

Nevertheless, in [7] a different way of looking at these two principles, and based on the concept of self-contradiction, respectively $a \cdot a' \leq (a \cdot a')'$, and $(a + a')' \leq ((a + a')')'$, for all a, b, was introduced. These expressions, that are even in more agreement with the original formulation given by Aristotle (see [1]), allow the verification of the two principles for a wide class of structures that include all algebras of fuzzy sets, as well as De Morgan algebras (see [5]). All that meant some progress for what concerns the verification of the two principles.

This paper tries to study the verification of the two Aristotle's principles once interpreted as in [7] , by some of the most well known systems of three-valued logic, namely, by those of the Łukasiewicz, Gödel, Kleene, Bochvar and Post. It is first shown that in Łukasiewicz, Gödel and Kleene cases the two principles are verified, that in the case of Bochvar only one of them is verified, and that in the case of Post no one is verified. Everything is done in agreement with the implication's semantics, as given at each case by its truth-table. At the end, although far from the semantical interpretation at each case, it is shown how the five systems could verify the two principles in the sense introduced in [7].

2 Systems of MV logic

2.1

Let Ω be a set consisting of propositions $p, q, r, \ldots$ and such that,

- If $p \in \Omega$, then '*not p*'$\in \Omega$
- If $p, q \in \Omega$, then '*p and q*', '*p or q*', and '*If p, then q*', are in Ω

2.2

Let L be a set with at least three elements, and endowed with a unary operation $' : L \to L$, and three binary operations $\cdot, +, \to: L \times L \to L$. Suppose there exists 0, 1 in L such that $0 \neq 1$, and $0' = 1$.

2.3

(Ω, L, t) is a *system of multiple-valued logic*, if $t : \Omega \to L$ (truth-function) verifies:

- $t(\text{not } p) = t(p)'$
- $t(p \text{ and } q) = t(p) \cdot t(q)$
- $t(p \text{ or } q) = t(p) + t(q)$
- $t(\text{if } p, \text{ then } q) = t(p) \to t(q)$

for all p, q in Ω. If L is with n elements, the system is an *n-valued* one.

2.4

Two systems of multiple-valued logic, (Ω_1, L_1, t_1) and (Ω_2, L_2, t_2), are *isomorphic* if there exist a pair of bijections (Φ, φ) such that

- $\Phi : \Omega_1 \to \Omega_2, \quad \varphi : L_1 \to L_2$
- $\varphi(a') = \varphi(a)', \varphi(a \cdot b) = \varphi(a) \cdot \varphi(b), \varphi(a + b) = \varphi(a) + \varphi(b), \varphi(a \to b) = \varphi(a) \to \varphi(b)$ with the corresponding operations in L_1, L_2
- $t_1 = \varphi^{-1} \circ t_2 \circ \Phi$, or $t_2(\Phi(p)) = \varphi(t_1(p))$ for all $p \in \Omega_1$.

Example.
The three-valued system of Łukasiewicz, Ł$_3$, (see [5]) is obtained with $L = \{T, F, I\}$ and the four operations given by the following tables:

	$'$
T	F
I	I
F	T

$\bullet$	T	I	F
T	T	I	F
I	I	I	F
F	F	F	F

$+$	T	I	F
T	T	T	T
I	T	I	I
F	T	I	F

$\to$	T	I	F
T	T	I	F
I	T	T	I
F	T	T	T

*This work has been partially supported by the Foundation for the Advancement of Soft Computing (Asturias, Spain), and CICYT (Spain) under project TIN2008-06890-C02-01

†Corresponding author.

ISBN: 978-989-95079-6-8

where $T, I, F \in L, T \neq F$, and $F' = T$.

With $L_0 = \{1, \frac{1}{2}, 0\}$, and $\varphi : L \to L_0$ given by $\varphi(T) = 1, \varphi(I) = \frac{1}{2}, \varphi(F) = 0$, the tables are translated into,

$'$	
1	0
$\frac{1}{2}$	$\frac{1}{2}$
0	1

$\bullet$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	0
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0
0	0	0	0

$+$	1	$\frac{1}{2}$	0
1	1	1	1
1	1	$\frac{1}{2}$	$\frac{1}{2}$
0	1	$\frac{1}{2}$	0

$\to$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	0
$\frac{1}{2}$	1	1	$\frac{1}{2}$
0	1	1	1

Then, *any system* $Ł_3(\Omega, L, t)$ *is isomorphic to* (Ω, L_0, t^*), with $t^*(p) = \varphi(t(p))$, for all $p \in \Omega$.

Notice that it is,
$a' = 1 - a, a \cdot b = \min(a, b), a + b = \max(a, b)$,
$a \to b = \min(1, 1 - a + b)$, for all $a, b \in L_0$.
Hence,
$t(\text{not } p) = 1 - t(p), t(p \text{ and } q) = \min(t(p), t(q))$,
$t(p \text{ or } q) = \max(t(p), t(q))$,
$t(\text{If } p, \text{ then } q) = \min(1, 1 - t(p) + t(q))$,
for all $p, q \in \Omega$.
Notice that $(a')' = a'' = a$.

3 The other L_0 three-valued systems

Analogously, we can obtain the corresponding systems of Gödel, Bochvar, Kleene and Post (see [5], [2]), by their isomorphic images with $\varphi : L \to L_0$ given by $\varphi(T) = 1, \varphi(I) = \frac{1}{2}, \varphi(F) = 0$.

3.1

Gödel (G_3), is translated into L_0 by

$'$	
1	0
$\frac{1}{2}$	0
0	1

$\bullet$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	0
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0
0	0	0	0

$+$	1	$\frac{1}{2}$	0
1	1	1	1
$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$
0	1	$\frac{1}{2}$	0

$\to$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	0
$\frac{1}{2}$	1	1	0
0	1	1	1

or, for all $a \in L_0$,

$$a' = \begin{cases} 1, & \text{if } a = 0 \\ 0, & \text{if } a \neq 0 \end{cases}$$

$a \cdot b = \min(a, b)$
$a + b = \max(a, b)$

$$a \to b = \begin{cases} 1, & \text{if } a \leq b \\ b, & \text{if } a > b \end{cases}$$

Notice that, the negation $'$ is not strong, since $(\frac{1}{2}')' = 0' = 1$.

3.2

Kleene (K_3), is translated into L_0 by

$'$	
1	0
$\frac{1}{2}$	$\frac{1}{2}$
0	1

$\bullet$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	0
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0
0	0	0	0

$+$	1	$\frac{1}{2}$	0
1	1	1	1
$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$
0	1	$\frac{1}{2}$	0

$\to$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	0
$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$
0	1	1	1

or, for all $a \in L_0$,
$a' = 1 - a$,
$a \cdot b = \min(a, b)$,
$a + b = \max(a, b)$,
$a \to b = \max(1 - a, b)$.

Notice that $(a')' = a'' = a$, for all $a \in L_0$ (the negation is strong).

3.3

Bochvar (B_3) is translated into L_0 by

$'$	
1	0
$\frac{1}{2}$	$\frac{1}{2}$
0	1

$\bullet$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	0
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
0	0	$\frac{1}{2}$	0

$+$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	1
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
0	1	$\frac{1}{2}$	0

$\to$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	0
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
0	1	$\frac{1}{2}$	1

or,
$a' = 1 - a$,

$$a \cdot b = \begin{cases} \frac{1}{2}, & \text{if } (a, b) \in \{(\frac{1}{2}, 0), (0, \frac{1}{2})\} \\ \min(a, b), & \text{otherwise} \end{cases},$$

$$a + b = \begin{cases} \frac{1}{2}, & \text{if } (a, b) \in \{(1, \frac{1}{2}), (\frac{1}{2}, 1)\} \\ \max(a, b), & \text{otherwise} \end{cases},$$

$a \to b = a' + b$.

Notice that $(a')' = a'' = a$, for all $a \in L_0$ (the negation is strong).

3.4

Post (P_3), is translated into L_0 by

$'$	
1	$\frac{1}{2}$
$\frac{1}{2}$	0
0	1

$\bullet$	1	$\frac{1}{2}$	0
1	0	0	$\frac{1}{2}$
$\frac{1}{2}$	0	1	$\frac{1}{2}$
0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$

$+$	1	$\frac{1}{2}$	0
1	1	1	1
$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$
0	1	$\frac{1}{2}$	0

$\to$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	$\frac{1}{2}$
$\frac{1}{2}$	1	$\frac{1}{2}$	0
0	1	1	1

or,

$$a' = \begin{cases} 1, & \text{if } a = 0 \\ |\frac{1}{2} - a|, & \text{if } a \neq 0 \end{cases},$$

$a \cdot b = (a' + b')'$,
$a + b = \max(a, b)$,
$a \to b = a' + b$.

Notice that, for all $a \in L_0$, is $(a')' = a'' \neq a$ (the negation is not strong).

Remark. 3.1. In the systems $Ł_3, K_3, G_3, B_3$ but not in P_3, holds the following:

- $0 \cdot 1 = 1 \cdot 0 = 0 \cdot 0 = 0, 1 \cdot 1 = 1$
- $0 + 1 = 1 + 0 = 1 + 1 = 1, 0 + 0 = 0$
- $0' = 1, 1' = 0$

ISBN: 978-989-95079-6-8

- $1 \to 1 = 0 \to 1 = 0 \to 0 = 1, 1 \to 0 = 0$

Then, these four three-valued systems do contain the classical system of two-valued logic based on the values $\{0, 1\}$ that, notwithstanding, is not contained in P_3, a system in which only the operation + preserves such classical case.

4 The two principles

4.1

Triplets $(L, \vDash, ')$ consisting of a non-empty set L, a transitive relation $\vDash \subset L \times L$, and a mapping $' : L \to L$ reversing $\vDash$: if $a \vDash b$, then $b' \vDash a'$, are called *transitive reversing systems* (TRS for short) (see [7])

4.2

An operation $\cdot : L \times L \to L$, is a *type-1 operation* in $(L, \vDash, ')$, provided $a \cdot a' \vDash a$ and $a \cdot a' \vDash a'$, for all $a \in L$.

An operation $+ : L \times L \to L$, is a *type-2 operation* in $(L, \vDash, ')$, provided $a \vDash a + a'$ and $a' \vDash a + a'$, for all $a \in L$.

Theorem 4.1. *(ANC, for Aristotle's Non-contradiction). If $(L, \vDash, ')$ is a TRS, and $\cdot$ is a type-1 operation, it holds*

$$a \cdot a' \vDash (a \cdot a')' \tag{1}$$

for all $a \in L$ (see [7])

Theorem 4.2. *(AEM, for Aristotle's Excluded-middle). If $(L, \vDash, ')$ is a TRS, and $+$ is a type-2 operation, it holds*

$$(a + a')' \vDash ((a + a')')' \tag{2}$$

for all $a \in L$ (see [7])

Remark. 4.3. By defining 'a is selfcontradictory' as $a \vDash a'$, and by taking $'$ representing 'not', $\cdot$ representing 'and', it is clear that (1) can be read

'a and not a, is self-contradictory'.

Hence, interpreting 'impossible' by 'self-contradictory' (1) represents 'a and not a, is impossible'. That is, (1) can be read as an algebraic translation of the Aristotle's Non-contradiction principle.

Remark. 4.4. Like in the previous remark, and taking + representing 'or', (2) can be read as 'not (a or not a) is self - contradictory', or 'not (a or not a) is impossible', an algebraic translation of the Aristotle's Excluded-middle principle usually stated as 'a or not a is always'.

4.3

Theorem 4.5. *Given a triplet $(L, \cdot, ')$, with $L \neq \emptyset$, $\cdot : L \times L \to L$, and $' : L \to L$, the relation $\vDash_{NC} \subset L \times L$, given by the set of pairs*

$$\vDash_{NC} = \{(a \cdot a', (a \cdot a')'); a \in L\},$$

assures the verification of the principle ANC.

Proof. Obviously, $a \cdot a' \vDash_{NC} (a \cdot a')'$, for all $a \in L$. □

Theorem 4.6. *Given a triplet $(L, +, ')$, with $L \neq \emptyset$, $+ : L \times L \to L$, and $' : L \to L$, the relation $\vDash_{EM} \subset L \times L$, given by the set of pairs*

$$\vDash_{EM} = \{((a + a')', ((a + a')')'); a \in L\},$$

assures the verification of the principle AEM

Proof. Obviously, $(a + a')' \vDash_{NC} (a + a')')'$, for all $a \in L$. □

Theorem 4.7. *Given a quadruplet $(L, \cdot, +, ')$, with $L \neq \emptyset$,$\cdot : L \times L \to L$, $+ : L \times L \to L$, and $' : L \to L$, the relation*

$$\vDash = \vDash_{NC} \cup \vDash_{EM} \subset L \times L$$

assures the verification of the two principles ANC and AEM.

Proof. Obviously, for all $a \in L$, is

- $a \cdot a' \vDash_{NC} (a \cdot a')'$ and $a \cdot a' \vDash (a \cdot a')'$
- $(a + a')' \vDash_{EM} ((a + a')')'$, and $(a + a')' \vDash ((a + a')')'$. □

It is to be noticed that *when $'$ verifies $(a')' = a'' = a$, ($'$ is strong), the triplets $(L, \vDash_{NC}, ')$* and $(L, \vDash_{EM}, ')$ are TRS, since:

- $a \cdot a' \vDash_{NC} (a \cdot a')'$, and $(a \cdot a')'' = a \cdot a'$
- $(a+a')' \vDash_{EM} ((a+a')')'$ is equivalent to $(a+a')' \vDash_{EM} a + a'$

Of course, *this does not mean that those triplets are TRS only when $'$ is strong*.

Remark. 4.8. It is obvious that the set of relations $\vDash$ for which $(L, \cdot, +, ')$ verifies, respectively, the principle ANC or the principle AEM, is not empty. If $\vDash$ allows to verify ANC, it is $\vDash_{NC} \subset \vDash$, if $\vDash$ allows to verify AEM is $\vDash_{EM} \subset \vDash$, and if $\vDash$ allows both principles is $\vDash_{NC} \cup \vDash_{EM} \subset \vDash$

Remark. 4.9. It should be pointed out that the last theorems ignore what follows. *The unary operation $'$ should represent 'not', and the relation $a \vDash b$ should represent* '*If a, then b*'. Otherwise, the interpretation of 'If a, then not a' by means of $a \vDash a'$ (a is self-contradictory) has no sense at all. Hence, *and provided $a \cdot b$ does represent* 'a and b', *those representations are crucial for asserting that $a \cdot a' \vDash (a \cdot a')'$ is an interpretation of* "'a and not a' is a self-contradictory statement". And analogously with $(a + a')' \vDash ((a + a')')'$.

4.4

Let us call *Modern Non-contradiction Principle* (MNC, for short), the statement '$a \cdot a' = 0$ for all $a \in L$', and *Modern Excluded-middle Principle* (MEM, for short), '$a + a' = 1$ for all $a \in L$'

Theorem 4.10. *Given a triplet $(L, \vDash, ')$, an operation $\cdot : L \times L \to L$, and $0 \in L$, such that*

- $0 \vDash 0'$
- $a \cdot a' = 0$ *(MNC)*

then, it holds ANC.

ISBN: 978-989-95079-6-8

Proof. $a \cdot a' = 0 \vDash 0' = (a \cdot a')'$. $\square$

Theorem 4.11. *Given a triplet* $(L, \vDash, ')$*, an operation* $+ : L \times L \to L$*, and* $1 \in L$*, such that*

- $1' \vDash (1')'$
- $a + a' = 1$ *(MEM)*

then, it holds AEM.

Proof. $(a + a')' = 1' \vDash (1')' = ((a + a')')'$. $\square$

Remark. 4.12. If $1' = 0$, then $1' \vDash (1')'$ is equivalent to $0 \vDash 0'$

Remark. 4.13. In the sense of theorems 4.10 and 4.11, MNC and MEM are particular cases of ANC and AEM, respectively.

Remark. 4.14. Apart of the three operations $(\cdot, +, ')$, the principles ANC and AEM only involve the relation $\vDash$, but the principles MNC and MEM do involve the relation =, and the 'singular' elements 0,1.

4.5

In what follows this paper is devoted to study if the three-valued systems $Ł_3, G_3, K_3, B_3$, and P_3, do verify the principles ANC or AEM once expressed in the above algebraic terms.

5 Posing the problem for the five three-valued systems

Given a MVL system (Ω, L_0, t), consider its 'values' part $(L_0, ', \cdot, +, \to)$, and the binary relations $\leq$ and $\vDash$ given, respectively, by

$a \vDash b \Leftrightarrow a \to b = 1$, and $\leq$ is the linear order in $L_0 = \{1, \frac{1}{2}, 0\}$, inherited from $\mathbb{R}$.

- *Case* $Ł_3$.

  It is $a \to b = 1 \Leftrightarrow \min(1, 1 - a + b) = 1 \Leftrightarrow a \leq b$. That is, $\vDash_{Ł} = \leq$.

  Since $a' = 1 - a$ implies "$a \leq b \Leftrightarrow b' \leq a'$", $(L_0, \leq, ')$ is a TRS in which, obviously, the operation $\cdot = \min$ is one of type-1, and $+ = \max$ is one of type-2. Hence, it holds

  $$a \cdot a' \leq (a \cdot a')', \ (a + a')' \leq ((a + a')')',$$

  for all $a \in L_0$.

- *Case* G_3.

  It is $a \to b = 1 \Leftrightarrow a \leq b$. That is, $\vDash_G = \leq$, like in the case $Ł_3$. Since it is also $\cdot = \min$, $+ = \max$, it also holds

  $$a \cdot a' \leq (a \cdot a')', \ (a + a')' \leq ((a + a')')',$$

  for all $a \in L_0$.

- *Case* K_3.

  Since, $\vDash_K = \{(1,1), (\frac{1}{2}, 1), (0,1), (0, \frac{1}{2}), (0,0)\} \subset \leq$, and $\vDash'_K = \{(0,0), (0, \frac{1}{2}), (0,1), (\frac{1}{2}, 1), (1,1)\} = \vDash_K$, it means that $(L_0, \vDash_K, ')$ is a TRS. Since, $a' = 1 - a$, it is also $(L_0, \leq, ')$ a TRS.

  From $\frac{1}{2} \cdot \frac{1}{2}' = \frac{1}{2}$, $(\frac{1}{2} \cdot \frac{1}{2}')' = \frac{1}{2}$, and $(\frac{1}{2}, \frac{1}{2}) \notin \vDash_K$, follows that it is not $a \cdot a' \vDash (a \cdot a')'$, for all $a \in L_0$.

  Analogously, $(\frac{1}{2} + \frac{1}{2}')' = \frac{1}{2}$, $(\frac{1}{2} + \frac{1}{2})'' = \frac{1}{2}$, shows that it is not $(a + a')' \vDash ((a \cdot a')')$, for all $a \in L_0$.

  Nevertheless, since $\cdot$ denoting $\min$ is a type-1 operation in $(L_0, \leq, ')$, and $+$ denoting $\max$ is a type-2 operation in $(L_0, \leq, ')$, it follows

  $$a \cdot a' \leq (a \cdot a')', \ (a + a')' \leq ((a + a')')',$$

  for all $a \in L_0$.

- *Case* B_3.

  It is $\vDash_B = \{(1,1), (0,1), (0,0)\} \subset \leq$, and $\vDash'_B = \{(b', a'); (a, b) \in \vDash_B\} = \{(0,0), (0,1), (1,1)\} = \vDash_B$. That is, $(L_0, \vDash_B, ')$ and $(L_0, \leq, ')$ are TRSs.

  Since, $\frac{1}{2} \cdot \frac{1}{2}' = \frac{1}{2}$ and $(\frac{1}{2} \cdot \frac{1}{2}')' = \frac{1}{2}$, but $(\frac{1}{2}, \frac{1}{2}) \notin \vDash_B$, it is not $a \cdot a' \vDash_B (a \cdot a')'$, for all $a \in L_0$.

  Since, $(\frac{1}{2} + \frac{1}{2}')' = \frac{1}{2}' = \frac{1}{2}$ and $((\frac{1}{2} + \frac{1}{2}')')' = \frac{1}{2}$, it is not $(a + a')' \vDash_B ((a + a')')'$, for all $a \in L_0$.

  Notwithstanding,

  $$\left.\begin{array}{lcl} 1 \cdot 1' = 0, (1 \cdot 1')' = 0' = 1 & \Rightarrow & 0 \leq 1 \\ \frac{1}{2} \cdot \frac{1}{2}' = \frac{1}{2}, (\frac{1}{2} \cdot \frac{1}{2}')' = \frac{1}{2} & \Rightarrow & \frac{1}{2} \leq \frac{1}{2} \\ 0 \cdot 0' = 0, (0 \cdot 0')' = 0' = 1 & \Rightarrow & 0 \leq 1 \end{array}\right\} \Rightarrow$$

  $a \cdot a' \leq (a \cdot a')'$, for all $a \in L_0$.

  Analogously,

  $$\left.\begin{array}{lcl} (1 + 1')' = 1, ((1 + 1')')' = 1' + 1 = 1 & \Rightarrow & 1 \leq 1 \\ (\frac{1}{2} + \frac{1}{2}')' = \frac{1}{2}, ((\frac{1}{2} + \frac{1}{2}')')' = \frac{1}{2} & \Rightarrow & \frac{1}{2} \leq \frac{1}{2} \\ (0 + 0')' = 0, ((0 + 0')')' = 0 + 0' = 1 & \Rightarrow & 0 \leq 1 \end{array}\right\} \Rightarrow$$

  $(a + a')' \leq ((a + a')')'$, for all $a \in L_0$.

- *Case* P_3.

  It is $\vDash_P = \{(1,1), (\frac{1}{2}, 1), (0,1), (0, \frac{1}{2}), (0,0)\} \subset \leq$, but since $(\frac{1}{2}, \frac{1}{2}) \in \vDash'_P$ it is not $\vDash'_P \subset \vDash_P$, and $(L_0, \vDash_P, ')$ is not a TRS. In addition, also $(L_0, \leq_P, ')$ it is not a TRS, since $\frac{1}{2} \leq 1$, but $1' = \frac{1}{2} \not\leq 0 = \frac{1}{2}'$.

  From, $0 \cdot 0' = 0 \cdot 1 = \frac{1}{2}, (0 \cdot 0')' = \frac{1}{2}' = 0$, and neither $(\frac{1}{2}, 0) \vDash 0$, nor $\frac{1}{2} \leq 0$, it follows that it is not $a \cdot a' \vDash_P (a \cdot a')'$, nor $a \cdot a' \leq (a \cdot a')'$, for all $a \in L_0$.

  From, $(0 + 0')' = \frac{1}{2}, ((0 + 0')')' = 0$, it follows that it is neither $(a + a')' \vDash_P ((a + a')')'$, nor $(a + a')' \leq ((a + a')')'$, for all $a \in L_0$.

6 The solutions to the problem

6.1

Given a multiple-valued system (Ω, L_0, t), consider the two relations:

$$p \vDash q \Leftrightarrow t(p) \vDash t(q), \text{ and } p \leq q \Leftrightarrow t(p) \leq t(q),$$

and define:

ISBN: 978-989-95079-6-8

- The MV-system (Ω, L_0, t) verifies the principle ANC respect to $\vDash, \leq$, respectively, whenever

  p *and not* $p \vDash$ *not* (p *and not* p),
  or, p *and not* $p \leq$ *not* (p *and not* p)

- The MV-system (Ω, L_0, t) verifies the principle AEM respect to $\vDash, \leq$, respectively, whenever

  not (p *or not* p) $\vDash$ *not*(*not* (p *or not* p)),
  or, *not*(p *or not* p) $\leq$ *not*(*not* (p *or not* p)),

  for all $p \in \Omega$.

6.2

Hence, from the results in section 5, follows:

1. Three-valued *Łukasiewicz* systems *do verify* ANC *and* AEM, *with respect to* $\vDash_{\mathbf{Ł}} = \leq$
2. Three-valued *Gödel* systems *do verify* ANC *and* AEM, *with respect to* $\vDash_G = \leq$
3. Three-valued *Kleene* systems *verify neither* ANC *,nor* AEM, *with respect to* $\vDash_K \subset \leq$
4. Three-valued *Kleene* systems *do verify* ANC *and* AEM, *with respect to* $\leq$
5. Three-valued *Bochvar* systems *verify neither* ANC, *nor* AEM, *with respect to* $\vDash_B \subset \vDash_K \subset \leq$
6. Three-valued *Bochvar* systems *do verify* ANC *and* AEM, *with respect to* $\leq$
7. Three-valued *Post* systems *verify neither* ANC, *nor* AEM, *with respect to* $\vDash_P = \vDash_K \subset \leq$
8. Three-valued *Post* systems *verify neither* ANC, *nor* AEM, *with respect to* $\leq$.

Anyway, and in the line of theorems in section 4.2, it is possible to go further for what concerns points 3, 5, 7 and 8.

6.3

Obviously, *the most surprising case is* P_3. By just adding $(\frac{1}{2}, 0)$ to $\vDash_P$, that is, with the relation

$$\vDash^*_P = \vDash \cup \{(\tfrac{1}{2}, 0)\},$$

it is easy to prove that *both* ANC *and* AEM *hold with respect to* $\vDash^*_P$. But the triplet $(L_=, \vDash^*_P, ')$ is not a TRS, since $(1,1) \in \vDash^*_P$ but $(1', 1') = (\frac{1}{2}, \frac{1}{2}) \notin \vDash^*_P$.

Taking $\vDash^{**}_P = \vDash^*_P \cup \{(\frac{1}{2}, \frac{1}{2})\} = \{(\frac{1}{2}, 0), (0, \frac{1}{2}), (0,0), (1,1), (\frac{1}{2}, 1), (0,1), (\frac{1}{2}, \frac{1}{2})\}$, it is obtained the TRS $(L_0, \vDash^{**}_P, ')$, and *it also holds* ANC *and* AEN *with respect to* $\vDash^{**}_P$. In addition, and as it is easy to check, *the operation* $\cdot$ *in* P_3 *is a type-1 operation in* $(L_0, \vDash^{**}_P, ')$, and *the operation* + in P_3 *is a type-2 operation in* $(L_0, \vDash^{**}_P, ')$.

It should be pointed out that neither $\cdot$ is a type-1 operation in $(L_0, \vDash^*_P, ')$, nor + is a type-2 operation in $(L_0, \vDash^*_P, ')$. The reasons are, respectively, that $\frac{1}{2} \cdot \frac{1}{2}' = \frac{1}{2}$, and $\frac{1}{2} + \frac{1}{2}' = \frac{1}{2}$, but $(\frac{1}{2} \cdot \frac{1}{2}) \notin \vDash^*_P$.

6.4

In the case K_3, taking $\vDash^*_K = \vDash_K \cup \{(\frac{1}{2}, \frac{1}{2})\} = \leq$, the triplet $(L_0, \vDash^*_K, ')$ is a TRS in which the operation $\cdot$ denoting min is a type-1 operation, and the operation + is a type-2 operation. Hence, *it holds ANC with respect to* $\vDash^*_K$ *, and also AEM with respect to the same* $\vDash^*_K$.

6.5

Concerning the case B_3, with $\vDash^*_B = \vDash_B \cup \{(\frac{1}{2}, \frac{1}{2})\}$, the TRS $(L_0, \vDash^*_B, ')$ is reached in which *it holds neither ANC, nor AEM, since*

$$(1+1')' = 1, ((1+1')')' = 1' = 0, \text{ but } (1,0) \notin \vDash^*_B .$$

Anyway, with $\vDash^{**}_B = \vDash^*_B \cup \{(1,0)\} = \{(1,1), (0,1), (0,0), (\frac{1}{2}, \frac{1}{2}), (1,0)\}$, it is obtained the TRS $(L_0, \vDash^{**}_B, ')$, *in which the two principles ANC and AEM obviously do hold.*

6.6

At the end, in the five cases there exist a TRS $(L_0, \vDash, ')$ for which the two principles ANC and AEM do hold. In the cases Ł$_3$ and G_3, it is $\vDash = \leq$ with $\leq$ the lineal order of the real line. In the case K_3, there is a solution with $\leq$, but also with $\vDash^*_K$ that, nevertheless, is again a part of the order $\leq$.

What is different are the cases of Bochvar and Post, where $\vDash^{**}_B$ is not a part of $\leq$, and $\vDash^{**}_P$ is not a part of $\leq$, but on the contrary, it is $\leq \subset \vDash^{**}_P$.

This fact, of not being a part of $\leq$, seems to be the red line between the three system Ł$_3, G_3, K_3$ and the two systems B_3, P_3. The trick to reach the principles ANC and AEM seems to be based on the identification of some pairs $(a, b) \in L_0 \times L_0$ that, added to $\vDash$ in the line of section 4.2, give a new relation for which (without altering the original negation) the two principles do hold.

6.7

Form Remark 4.8, it follows that the relation considered from 6.2 to 6.5 are larger than $\vDash_{NC}$, and $\vDash_{EM}$. As it is easy to check,

- In Ł$_3$, $\vDash_{NC} = \vDash_{EM} = \{(0,1), (\frac{1}{2}, \frac{1}{2})\} \subset \vDash_{\mathbf{Ł}} = \leq$
- In G_3, $\vDash_{NC} = \vDash_{EM} = \{(0,1)\} \subset \vDash_G = \leq$
- In K_3, $\vDash_{NC} = \vDash_{EM} = \{(0,1), (\frac{1}{2}, \frac{1}{2})\} \subset \leq$, but $\vDash_{NC} \nsubseteq \vDash_K$, and $\vDash_{NC} \nsubseteq \vDash^*_K$
- In B_3, $\vDash_{NC} = \vDash_{EM} = \{(0,1), (\frac{1}{2}, \frac{1}{2})\} \subset \leq$, but $\vDash_{NC} \nsubseteq \vDash_B$, and $\vDash_{NC} \nsubseteq \vDash^*_B$
- In P_3, $\vDash_{NC} = \vDash_{EM} = \{(0,1), (\frac{1}{2}, 0)\} \subset \vDash^*_P$, but $\vDash_{NC} \nsubseteq \leq$, and $\vDash_{NC} \nsubseteq \vDash_P$.

Remark. 6.1. An element $a \in L$ is $\vDash$-self-contradictory, if $a \vDash a'$.

In Ł$_3$, since $\vDash_{\mathbf{Ł}_3} = \leq$ and $a' = 1 - a$, it is $a \leq a'$ whenever $a \leq \frac{1}{2}$, that is, the self-contradictory elements are 0 and $\frac{1}{2}$. In G_3, with $\vDash_G = \leq$, the only element that is self-contradictory is 0.

In K_3, with $\vDash_K \subset \leq$, the only self-contradictory is 0, but with $\vDash^*_K$ also $\frac{1}{2}$ is self-contradictory.

ISBN: 978-989-95079-6-8

In B_3, with $\vDash_B \subset \leq$, the only element that is self-contradictory is 0, but with $\vDash_B^*$ and $\vDash_B^{**}$ also $\frac{1}{2}$ is self-contradictory.

In P_3, with $\vDash_P$, the only element that is self-contradictory is 0, but with $\vDash_P^*$ and $\vDash_P^{**}$ also $\frac{1}{2}$ is self-contradictory.

7 Conclusion

7.1

When the principles NC and EM are interpreted in the usual 'modern' forms,

$$a \cdot a' = 0,\ a + a' = 1$$

for all $a \in L_0$, neither Ł$_3$ nor G_3, K_3, B_3, P_3 do verify one of them. For example

- in Ł$_3, G_3, K_3, B_3$: $\frac{1}{2} \cdot \frac{1}{2}' = \frac{1}{2} \neq 0, \frac{1}{2} + \frac{1}{2}' = \frac{1}{2} \neq 1$
- in $P_3 : 0 \cdot 0' = \frac{1}{2} \neq 0, \frac{1}{2} + \frac{1}{2}' = \frac{1}{2} \neq 1$.

Hence, the results presented in this paper mean some progress in what concerns, jointly with what is in [7] and [6], the general verification of the non-contradiction and excluded-middle principles. If they can fail at the just mentioned 'modern' level, it seems to exist at least another level (that of self-contradiction) at which they do not fail.

7.2

Nevertheless, right now the authors are not yet able to completely answer the question of *why the systems Ł$_3, G_3, K_3$ do have solution in $\leq$, but the systems B_3, P_3 do have to exit from $\leq$ to reach a solution, and the system K_3 also admits a second solution in $\leq$*. In all these cases, the pairs that are to be added to the initial relation given by $a \to b = 1$, verify $a \to b \neq 1$. For example,

- In K_3, the added pair $(\frac{1}{2}, \frac{1}{2})$, verifies $\frac{1}{2} \to \frac{1}{2} = \frac{1}{2} \neq 1$
- In B_3, the added pairs $(\frac{1}{2}, \frac{1}{2})$ and $(1, 0)$, verify $\frac{1}{2} \to \frac{1}{2} = \frac{1}{2} \neq 1, 1 \to 0 = 0 \neq 1$
- In P_3, the added pairs $(\frac{1}{2}, 0)$ and $(\frac{1}{2}, \frac{1}{2})$, verify $\frac{1}{2} \to 0 = 0 \neq 1, \frac{1}{2} \to \frac{1}{2} = \frac{1}{2} \neq 1$

Notice that in the cases Ł$_3$ and G_3 it is $a \to b = 1 \Leftrightarrow a \leq b$. Instead, in K_3 it is only $a \to b = \max(1-a, b) = 1 \Rightarrow a = 0$, or $b = 1 \Rightarrow a \leq b$, but not reciprocally. It seems that the relation 'If a, then b' given by $a \to b = 1$, needs to be enlarged to reach both principles but without an agreement with the truth table of $\to$. The case K_3, with the change of $\frac{1}{2} \to \frac{1}{2} = \frac{1}{2}$ by $\frac{1}{2} \to \frac{1}{2} = 1$, results to be coincidental with Ł$_3$. This is a subject that, related to the semantics of the system, deserves more thinking.

7.3

Another consideration about the operators $\to$ drives us to see into its conditional behaviour related to the $\leq$ or $\vDash$ relations. That is [4], whether or not the *Modus Ponens* rule of inference, $a \cdot (a \to b) \leq b$, or $a \cdot (a \to b) \vDash b$, do hold.

Only the Gödel system, G_3, verifies them, as is easy to prove. For the rest of the cases the following counter-examples answer the question.

1. Ł$_3$, with respect to $\vDash_{Ł} = \leq$. It is $\frac{1}{2} \cdot (\frac{1}{2} \to 0) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2} \not\leq 0$, that is $(\frac{1}{2}, 0) \notin \leq$.
2. K_3 , with respect to $\vDash_K$, and $\vDash_K^* = \leq$. It is $\frac{1}{2} \cdot (\frac{1}{2} \to 0) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$, and $(\frac{1}{2}, 0) \notin \leq = \vDash_K^*$ and $(\frac{1}{2}, 0) \notin \vDash_K$ too.
3. B_3 , with respect to $\vDash_B, \vDash_B^*, \vDash_B^{**}$, and $\leq$. It is $\frac{1}{2} \cdot (\frac{1}{2} \to 0) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$, and $(\frac{1}{2}, 0) \notin \leq$, $(\frac{1}{2}, 0) \notin \vDash_B$, $(\frac{1}{2}, 0) \notin \vDash_B^*$ and $(\frac{1}{2}, 0) \notin \vDash_B^{**}$ too.
4. P_3 , with respect to both $\vDash_P$ and $\leq$. It is $\frac{1}{2} \cdot (\frac{1}{2} \to 0) = \frac{1}{2} \cdot 0 = \frac{1}{2}$, and $(\frac{1}{2}, 0) \notin \leq$, $(\frac{1}{2}, 0) \notin \vDash_P$. With respect to both $\vDash_P^*$ and $\vDash_P^{**}$. It is $\frac{1}{2} \cdot (\frac{1}{2} \to \frac{1}{2}) = \frac{1}{2} \cdot \frac{1}{2} = 1$, and $(1, \frac{1}{2}) \notin \vDash_P^*$, $(1, \frac{1}{2}) \notin \vDash_P^{**}$.

Hence, it is difficult to see the interest of such systems from the point of view of inference.

7.4

Last, but not least significant, is to notice how the different operators $\to$ verify the five properties characterizing implication functions [3], namely, i) $0 \to y = 1$, ii) $1 \to y = y$, iii) $x \to (y \to z) = y \to (x \to z)$, iv) $\to$ is decreasing in the first variable, and v) $\to$ is increasing in the second variable.

A simple calculation shows that in Ł$_3$, G_3, and K_3 the operation $\to$ verifies the five properties, therefore in these systems $\to$ is an implication function. On the contrary, B_3 does verify only properties ii and iii, and P_3 only i, iii, and iv. Hence, neither in B_3, nor in P_3, $\to$ is an implication function.

In the same vein, since in P_3 it is $1' = \frac{1}{2}$, but not $1' = 0$, operation $'$ is not, properly speaking, what is usually designed as a negation.

Acknowledgment

The authors express their thanks to the three anonymous reviewers for their hints and comments.

References

[1] Aristotle. *Metaphysics*. Book IV,4. Translated by W. D. Ross, eBooks@Adelaide, 2007.

[2] L. Bolc and P. Borowik. *Many-valued Logics*. Springer-Verlag, Berlin, 1992.

[3] D. Dubois and H. Prade. *Fundamentals of Fuzzy Sets*. Kluwer, Upper Saddle River, NJ, 2000.

[4] G. M. Hardegree. Material implication in orthomodular (and boolean) lattices. *Notre Dame Journal of Formal Logic*, 22(2):163–182, 1981.

[5] N. Rescher. *Many-valued Logic*. McGraw-Hill, New York, 1969.

[6] E. Trillas. An algebraic view on aristotle's nc and em principles. Submitted to *Theoria*.

[7] E. Trillas, C. Alsina, and A. Pradera. Searching for the roots of non-contradiction and excluded-middle. *International Journal of General Systems*, 31(5):499–513, 2002.

ISBN: 978-989-95079-6-8

Parametric Fuzzy Modelling Framework for Complex Data–Inherent Structures

Arne–Jens Hempel Steffen F. Bocklisch
Department of Electrical Engineering, Chair for Systems Theory, Chemnitz University of Technology
Chemnitz, Germany
Email: arne-jens.hempel@etit.tu-chemnitz.de, steffen.bocklisch@etit.tu-chemnitz.de

Abstract— *The present article dedicates itself to fuzzy modelling of data–inherent structures. In particular two main points are dealt with: the introduction of a fuzzy modelling framework and the elaboration of an automated, data–driven design strategy to model complex data–inherent structures within this framework.*

The innovation concerning the modelling framework lies in the fact that it is consistently built around a single, generic type of parametrical and convex membership function. In the first part of the article this essential building block will be defined and its assets and shortcomings will be discussed.

The novelty regarding the automated, data–driven design strategy consist in the conservation of the modelling framework when modelling complex (nonconvex) data–inherent structures. Instead of applying current clustering methods the design strategy uses the inverse of the data structure in order to created a fuzzy model solely based on convex membership functions.

Throughout the article the whole model design process is illustrated, section by section, with the help of an academic example.

Keywords— pattern recognition, fuzzy classification, fuzzy modelling of data structures, data–driven fuzzy classifier design

1 Introduction

Nowadays the world suffocates in vast amounts of data. To give those data an interpretable and thus economical meaning it has to be analysed. The goal of such an analysis is the creation of a model or the classification of the considered phenomenon, e. g. modelling of the traffic flow in cities, medical or machine diagnosis [1, 2, 3].

Basically there are two main philosophies to deduce such a model, theoretical and experimental modelling. In experimental modelling it is assumed that measurement data (objects) reflect the complexity of the phenomenon under consideration through data–inherent structures. Unfortunately, the same data might also exhibit imprecision (e. g. measuring inaccuracies) or depict interesting phenomena characteristics just vaguely (because of missing information). With the help of fuzzy set theory these occurring inaccuracies can be taken into account as a supplementary model feature [4].

In this work the whole modelling problem is understood as a fuzzy classification task, where specific fuzzy sets form a model equivalent. As it is pointed out in [5], there are a lot of sophisticated solutions for such a task, which in general apply nonparametric fuzzy sets or a composition of different fuzzy sets. Contrary to those approaches, the main philosophy behind this work is the exclusive usage of one specific parametrical fuzzy set to model complex data–inherent structures as well as the data itself. Another aspect of the here pursued type of structure modelling is that it works in the original feature space without any transformation like fuzzy support vector classifiers or any assumption of fuzzy functions [6, 7].

2 Fuzzy Pattern Classes

In order to become acquainted with the modelling philosophy it is necessary to understand its core component, the so called *fuzzy pattern class* (FPC). The subsequent sections provide a basic survey about the definition, composition, capabilities and utilisation of *fuzzy pattern class models*.

From a fuzzy theoretical perspective fuzzy pattern classes correspond to a side–specific parametrical multivariate membership function. FPCs are referred to as classes since they emerge from an agglomeration of class supporting objects (see section 2.2), consequently they represent an superordinate entity.

2.1 Definition of a Fuzzy Pattern Class

Although the usual FPC–membership function is multidimensional it derives from one–dimensional basis functions. Hence, it is reasonable to study this basis functions being equivalent to one–dimensional fuzzy pattern classes first. Generally a one–dimensional fuzzy pattern class A is defined over its class space U based on a side–specific parametrical function concept, see (1).

$$\mu^A(a,\vec{p}) = \begin{cases} \dfrac{a}{1+\left(\dfrac{1}{b_l}-1\right)\left|\dfrac{u}{c_l}\right|^{d_l}}, & u<0 \\ \dfrac{a}{1+\left(\dfrac{1}{b_r}-1\right)\left|\dfrac{u}{c_r}\right|^{d_r}}, & u\geq 0 \end{cases} \tag{1}$$

The function concept comprises a set of seven parameters a and $\vec{p}=(b_l,b_r,c_l,c_r,d_l,d_r)$. The further specification of these parameters results from the fact that the parameter a characterises an entire fuzzy pattern class, whereas the parameters combined in $\vec{p}$ are related to a dimension of the class space [8]. Beyond their mere mathematical functionality all parameters possess the following semantical meaning:

- The parameter a represents the maximum membership value of the FPC μ^A. Regarding a structure of classes the parameter a expresses the weight of a specific class. Considering a dynamic classification process a embodies the topicality or authenticity of the information represented by that class [9, 10].
- In the normalised case $a=1$, the parameters b_l, b_r of $\vec{p}$ assign left– and right–sided membership values at the class borders $u=-c_l$ and $u=c_r$.
- c_l, c_r mark the support of a class in a crisp sense. Both parameters characterise the left– and right–sided expansions of a fuzzy pattern class.
- The continuous descent of the membership function is specified by the parameters d_l, d_r. From a graphical point of view d_l, d_r determine the shape of the membership function, or in

ISBN: 978-989-95079-6-8

other words, the fuzziness of a class. Fig. 1 illustrates the introduced concept of the membership function considering the general unidimensional case.

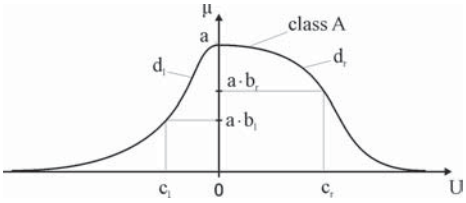


Figure 1: Membership function and parameters.

To obtain the common multidimensional fuzzy pattern class A the basis functions of each class dimension are accumulated using the N-fold compensatory Hamacher intersection operator (2), where n denotes the index of the basis functions and N the total number of dimensions [11].

$$k_{Ham\cap}\mu_N^A = \frac{1}{\frac{1}{N}\sum_{n=1}^{N}\frac{1}{\mu_n^A}} \tag{2}$$

Regarding the main philosophy behind this paper, the most important feature of this intersection operation is the conservation of the parametrical function concept for the multidimensional case [11].

Considering the multidimensional FPC form the class describing set of parameters is supplemented by a class specific position $\vec{u}_0$ in the original feature space and a class specific orientation $\vec{\varphi}$. Fig. 2 depicts the influence of the additional parameters for a two dimensional three class structure. The different location of each class results from $\vec{u}_{c10} = (0.8, 0.2)^T$ $\vec{u}_{c20} = (0.5, 0.5)^T$ and $\vec{u}_{c30} = (0.2, 0.8)^T$, whereas an additional class orientation φ of 60° has been applied to the middle class.

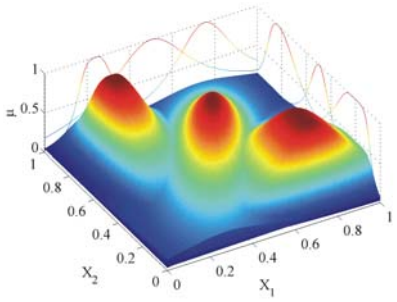


Figure 2: Two dimensional three class structure

2.2 *Data Driven Design of Fuzzy Pattern Classes*

Being familiar with the definition of fuzzy pattern classes a more intriguing question is how fuzzy pattern class models can be deduced. As a matter of principle such models can be obtained via two different approaches [8].

First they can be defined by expertise. That is an expert determines all class parameters based upon task and domain specific knowledge. This approach is not pursued here.

The second approach is a data–driven method, strongly advocating the here featured goal to model data–inherent structures. Based upon a class labelled set of learning data x_l (e.g. measurement objects) all class parameters are assigned automatically by a two step aggregation procedure [9, 10]. The class labels might result from a preliminary conducted cluster analysis.

For the sake of clarity only the basic aggregation principle will be outlined in the following. A full description can be found in [9, 12]. In order to perform the aggregation on sound mathematical foundations, the crisp learning dataset is extended to a set of fuzzy objects, using the introduced function concept (1). This extension is justified by the fact that every observation (measurement) inheres a so called "elementary fuzziness" (e.g impression of a sensor) [9]. In the first aggregation step the class position $\vec{u}_0$, alignment φ and extensions c_l, c_r are calculated in a dimension–wise manner. As exemplified in Fig. 3 the position $\vec{u}_0$ and alignment $\vec{\varphi}$ of the class space is obtained via a principal component analysis (PCA), where $\vec{u}_0$ is defined by the mean over all learning objects and $\vec{\varphi} = (\varphi_1, \varphi_2, \ldots, \varphi_{N-1})^T$ by sequential rotation of the class space U into the principal axes.

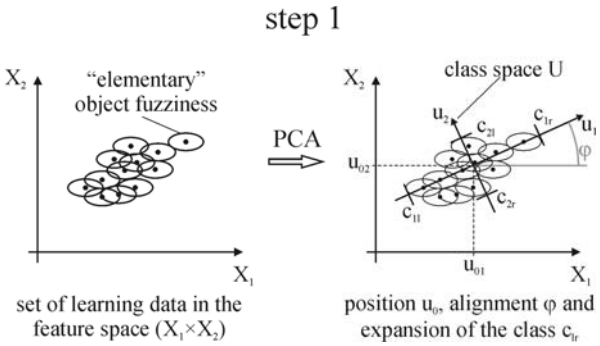


Figure 3: Step 1: aggregation procedure

The extensions c_l, c_r of the class are determined by the outermost objects in each class space dimension. In the subsequent second aggregation step the class shape d_l, d_r and the border memberships b_l, b_r are determined also dimension–wisely.

After the transformation of the objects x_l into their corresponding class space U the shape of a fuzzy patter class (d_l, d_r) is assigned based on their agglomeration properties. The more the data resembles an agglomeration according to a geometric series the smoother the class shape. The rate of resemblance is determined by the mean distance between two adjacent objects. The smoothest class shape is obtained for $d_{lr} = 2$ where the objects are cumulating in the centre of the class conform to a geometric series. The crisp case results for ($d_{lr} \to \infty$) where the objects are equally distributed over the class space, however for calculation purposes $d_{lr} = 20$ has proven to be a sufficient value to represent the crisp case.

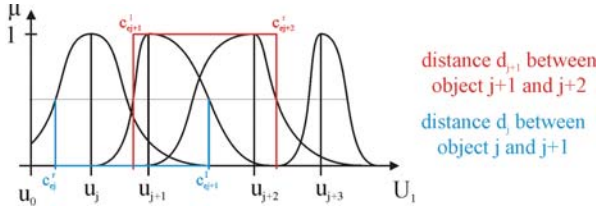


Figure 4: Step 2a: determination of the class form

The values for border memberships b_l, b_r are derived by the

ISBN: 978-989-95079-6-8

conservation of the object cardinality taking into account the results of c_{lr} and d_{lr}. Since the integral over the class membership function (1) cannot be solved analytically the border memberships b_l, b_r are estimated by a binary search over the unity interval $b_{lr} \in [0,1]$ [9].

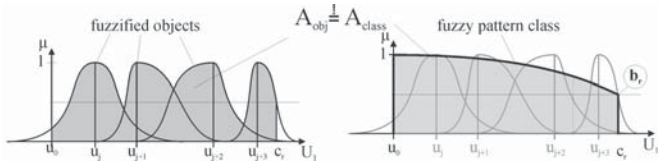


Figure 5: Step 2b: claim for cardinality

The class "weight" a is calculated by a logistic function given the number of the class supporting objects, see [9].

2.3 Application of FPC and Mode of Operation

For applicative purposes all task relevant fuzzy pattern classes are grouped together in a so called fuzzy pattern classifier. In operating mode the fuzzy pattern classifier assigns unknown objects to this class structure. The objects to be classified are each denoted by a vector $\vec{x}$ of their features:

$$\vec{x} = (x_1, x_2, ..., x_N)^T , \tag{3}$$

where N represents the number of feature dimensions. The results of the classification process are stored into a so called vector of sympathy $\vec{s}$. The components of $\vec{s}$ express the membership of a classified object to the corresponding class:

$$\vec{s} = (s_1, s_2, ..., s_K)^T , \tag{4}$$

where K is the total number of classes. The gradual membership of an object to a given class is calculated using (1).

$$s_k = \mu^k(\vec{x}) \quad \text{for } k = 1, 2, ..., K \tag{5}$$

Figure 6 illustrates the process of classification with the help of a one-dimensional three class structure. The object to be classified is situated in the centre of class two, the right outskirts of the first class and in the left centre of the third class. Alongside with the classification task the classification results are listed.

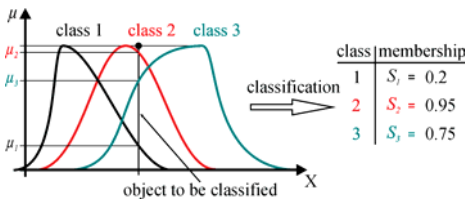


Figure 6: Object classification

According to Fig.6 the vector of sympathy describes a unique assignment of the object to the class structure with respect to its location in the feature space, since there are three classes it contains three values of membership.

2.4 Properties of Fuzzy Pattern Classes

In order to round off the comprehension about the afore introduced concept of fuzzy pattern class models its major properties (advantages and drawbacks) will be outlined subsequently. The main features of the fuzzy pattern class model lie in its versatility, its uniformity and its closed modelling framework. All these features can be attributed to the unimodal, side–specific and parametric class membership function. In the most general case the class membership function offers multivariate FPC models with various and asymmetric shapes, ranging from peak- over bell- to crisp shaped fuzzy sets. In connection with the introduced data–driven design procedure the class membership function allows to map class internal object distributions onto its shape and to model correlative relations without loosing its fuzzy logic basis. Besides this flexibility it has to be stressed that the parameters are semantically motivated or have at least a semantical meaning. It is therefore that the fuzzy pattern classes are considered to be well interpretable and transparent. Another feature resulting from the utilisation of the parametric membership concept is its good trade off between data compression, computational cost and generality. Due to the choice of the membership function (1) each fuzzy pattern class is defined on a set of eight parameters per dimension, providing a sufficient data compression, especially for high dimensional models. Due to the choice of the conjunction operator (2) the intersection operation is exclusively performed on *parameter level* saving computational cost. Both advantages are traded off for generality in so far as class membership functions are convex models, specifying a convex area of the feature space. Consequently FPC are best suited to model convex data–inherent structures. But their convex nature causes fuzzy pattern class models to be afflicted with significant errors when it comes to model nonconvex data–inherent structures.

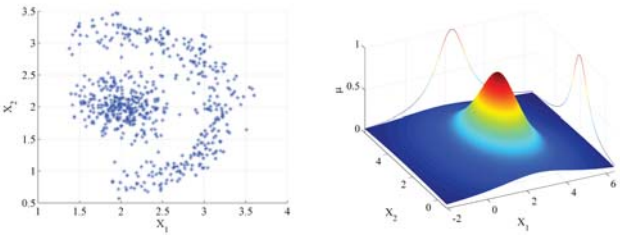


Figure 7: Nonconvex data structure and according FPC model

An example for such an error is depicted in Fig.7 where a central object accumulation enclosed by a half circle shaped data structure was aggregated to a fuzzy pattern class. Obviously the region between the object accumulations does not belong to the given data structure but the associated fuzzy pattern model μ^C will assign high grades of memberships for this region.

In order to circumvent this major drawback two possibilities can be thought of:
The first way, is to segment the data into convex subsets, for example with the help of cluster algorithms. Aside from the fact that this approach works on every data–inherent structure it might create considerably large structures of fuzzy pattern classes at the expenses of model clarity and computational costs.

3 Fuzzy Pattern Anti–classes

The hereafter elaborated access to dissolve the convexity drawback arises from the negation of a class assertion over its unsupported class space. The idea behind this approach

ISBN: 978-989-95079-6-8

can be stated as follows: *The difference of convex sets can be a non–convex set.*
Fig.8 sketches such a negation of the FPC model for a ring shaped problem.

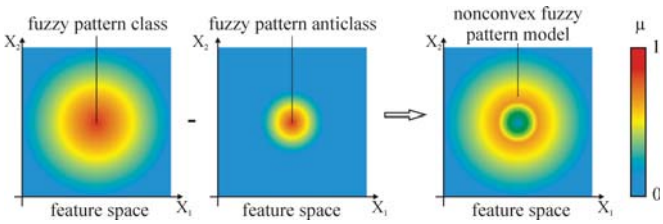


Figure 8: Nonconvex fuzzy pattern model via negation

As it is depicted the negation will be introduced in terms of fuzzy pattern anti–classes (FPAC). It works on semantical level and from this point of view FPACs can be seen as a further specification of a preceding FPC.
Mathematically fuzzy pattern anti–classes are defined upon the same membership function concept as fuzzy pattern classes, see (1). The main concern of of this definition is conservation of the fuzzy pattern modelling framework and with it the *automated model generation* and the *model properties* (such as flexibility, interpretability, computational efficiency, etc.). The enormous shape diversity of fuzzy pattern classes together with the mutual negation of such membership functions allows to model almost any form of data–inherent structures. The only task that needs to be solved can be formulated as follows:
Determine the fuzzy pattern anti–classes given a set of learning data, containing an arbitrary (nonconvex) data–inherent structure and the appendent FPC model.

3.1 Design of Fuzzy Pattern Anti–classes

Assuming that FPACs, like usual fuzzy pattern classes, can be supported by objects or, better so called "anti–objects", then it is in accordance with modelling framework that FPCAs can be designed in a data–driven manner. Aiming to elaborate an automated FPAC design method the already introduced automated databased algorithm can be exploited. However the introduced data–driven FPC design relies on class supporting objects. Correspondingly, in order to setup FPACs these "anti–objects" have to be found first. When considering the general case there are no "anti–objects" given. Consequently they have to be *generated*.
For the sake of clearness the anti–object generation procedure will be illustrated for the two–dimensional case with the help of an academic example. The general multidimensional case follows analogously.

3.1.1 Generation of Anti–objects

According to the task (section 3) the information given are the location of each class supporting object in the learning dataset and the modelling fuzzy pattern class, see Fig.7. Suppose that there is no prior information about the distribution of the learning data it is impossible to make an assumption about the shape of a data–inherent structure. That is why the central idea behind the anti–object generation is led by the assumption that the *anti–objects will agglomerate in the class space being unsupported by learning objects*, such that they will adopt a kind of inverse data–inherent structure and form FPACs.
To ascertain whether a partition of the class space is actually supported by learning objects requires a discretisation of the class space. The size of the class space to be discretised is determined based upon the class borders c_{lr} in each dimension, whereas the discretisation resolution amounts to 2% of the class space leading to a 50×50 matrix F with 2500 cells. A cell of an N–dimensional class space then posses the following the extend:

$$V_c = \frac{1}{50}\left((c_{l1}+c_{r1})\cdot(c_{l2}+c_{r2})\cdot\ldots\cdot(c_{lN}+c_{rN})\right) \qquad (6)$$

Fig.9 depicts the discretised class space for the example along with the class borders in green, the object supported cells highlighted in red and the unsupported class space cells coloured in dark blue.

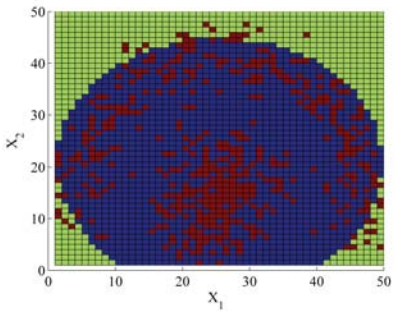


Figure 9: Discretised class space for the example

According to the central idea, the anti–objects have to agglomerate in these unsupported cells. In other words it is necessary to define an model for the agglomeration process. In particular two ways have been pioneered to model the accumulation of anti–objects, namely accumulation based on the elemental fuzziness and accumulation based on power series expansion. The most promising agglomeration model resulted from the assumption of a Fibonacci series.
Its leading thought is the expansion of Fibonacci numbers around the cell of interest until an object supported cell or a class border cell is reached. The result of the expansion around an arbitrary cell (i, j) is a 50×50 matrix $F_{d(i,j)}$ containing only the expanded Fibonacci numbers and zeros.

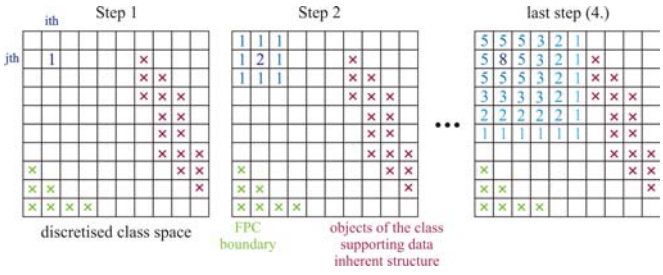


Figure 10: Fibonacci expansion for an arbitrary cell

In this context it has to be mentioned that the fuzzy pattern class borders are treated like objects. Purpose of this procedural manner is to provide an insight of the inner class anti–object distribution by preventing an anti–object accumulation close to the class borders.

ISBN: 978-989-95079-6-8

Fig.10 illustrates the implementation of the Fibonacci expansion for an arbitrary cell at position (i, j). The sum over all Fibonacci matrices $F_{d(i,j)}$ yields the result of the agglomeration process, as in (7).

$$F_d = \sum_{i=1}^{N} \sum_{j=1}^{N} F_{d(i,j)} \tag{7}$$

F_d can be interpreted as a kind of anti–object density matrix. It is broken down into an explicit number of anti–objects per cell based on the maximum object density of all cells d_{max}, see (8).

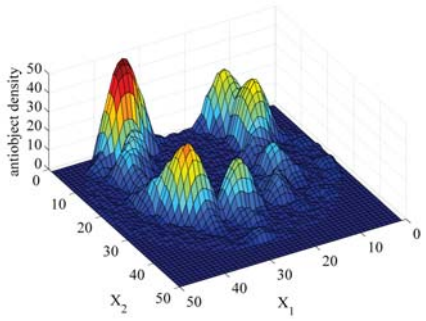


Figure 11: Anti–object density matrix F_d for the example

$$F_a = rd\left(\left(\frac{d_{max}}{max\,(F_d)}\right) F_d\right) \tag{8}$$

With the usual rounding operation rd, in (8), it is assured that each cell contains a natural number of anti–objects ($F_a \in \mathbb{N}^{50\times 50}$).

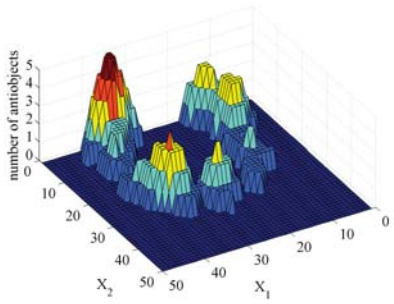


Figure 12: Anti–object matrix F_a for the example

Fig.12 presents the resulting number of anti–objects in their corresponding cells, after applying (8), for the example. The total number of anti–objects, being distributed around the centre region amounts to 887 with a maximum density of five anti–objects per cell.

The final step to complete the anti–object generation comprises the assignment of a position to each anti–object. For this purpose all anti–objects are uniformly distributed over their associated cell. In the here discussed two–dimensional case the anti–object position of an arbitrary cell with the coordinates (i, j) follows from (9), where *rand* draws an random number from the unity interval.

$$\begin{pmatrix} u_{a1} \\ u_{a2} \end{pmatrix} = \begin{pmatrix} i \cdot \frac{1}{50}\,(c_{l1} + c_{r1}) - rand \\ j \cdot \frac{1}{50}\,(c_{l2} + c_{r2}) - rand \end{pmatrix} \tag{9}$$

After their distribution the set of anti–objects forms an inverse of the data–inherent structure, see left–hand side of Fig.13.

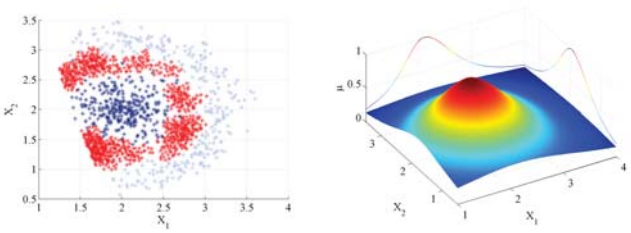


Figure 13: left: Objects, anti–objects, right: according FPAC

However, as it can be seen the anti–objects (in red) form a nonconvex data structure. Modelling this with the help of the automated FPC algorithm again creates a faulty FPAC μ^{AC}, see right–hand side of Fig.13. This might lead to the misconception that the problem was just shifted to the anti–objects. But that is by no means the case, since the whole procedure can be applied again to the created anti–class.

Like a block letter and its print anti–objects and objects complement each other to a complete convex model. Or with another emphasis anti–objects together with objects cover the *entire* class space. Because of that complementary relation, the original objects can be seen as the anti–objects of the anti–objects. Consequently, instead of running through the entire process again for the anti–class, the original objects can be recycled. The reutilisation of the original objects is justified by the facts that a part of the original objects is already complementing the set of anti–objects to a convex description and that their position is already known. Original objects being considered to negate the anti–class are required, firstly to posses a high membership to the anti–class and secondly to be located within the class borders of the anti–class. Generally this applies to original objects with a anti–class membership above $\mu = 0.5$.

For the considered example, the only objects meeting this requirement are the ones located in the centre region, highlighted in dark blue see Fig.13. After setting up the "anti–anti–class" μ^{AAC} by the selected original objects the negation procedure stops owing to the fact the average membership of possible anti–objects drops below a threshold of $\mu = 0.5$.

3.2 Combination of Fuzzy Pattern Anti–class Models

After their generation the anti–class models have to be combined together in order to from an overall model of the data–inherent structure. This combination process is characterised by the following keynote: *If an object does not belong to the anti–class it belongs to the preceding class.* A reasonable implementation of this key concept derives from the concatenation of the natural complement and minimum conjunction [13, 14].

$$\mu = min\left(\mu^{class}, \left(1 - \mu^{anti-class}\right)\right) \tag{10}$$

If the anti–class generation yields several levels of negation their sequence has to be respected for the concatenation. Or differently it has to be respected that *each level of negation creates an new level of hierarchy*. Equation (11) represents this aspect for the example.

ISBN: 978-989-95079-6-8

$$\mu = min\left(\mu^{C}, 1 - min\left(\mu^{AC}, \left(1 - \mu^{AAC}\right)\right)\right) \qquad (11)$$

Fig.14 depicts the nonconvex and fuzzy model, resulting from the class–anti–class combination, together with the supporting objects in black.

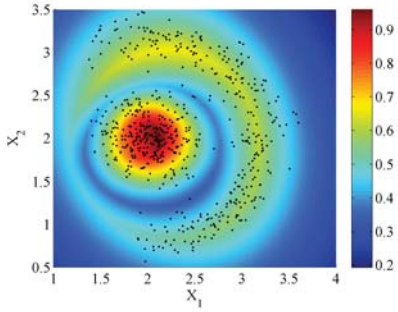


Figure 14: Resulting FPC Model

3.3 Properties of the Fuzzy Pattern Anti–class Design

The introduction of FPAC design is round off by having its assets and drawbacks summarised subsequently.
The main advantage of the introduced design approach lies in the exploitation of a single type of membership function. All the beneficial properties of fuzzy pattern classes, such as transparency, interpretability, computational efficiency etc. are imparted and conserved into the design of nonconvex FPC models. Equally important, by complementing the objects over the considered class space the FPAC design converts the major drawback of fuzzy pattern models (convexity) into an advantage. As a consequence the anti–objects have to be constructed only once. Furthermore the complete object anti–object covering of the class space guarantees convergence of the FPAC design method, since there is no further anti–object generation. Besides that, each anti–class is smaller or equal in size compared to its preceding class such that the area of interest is also converging to zero. Another aspect worth to mention is that the FPAC design works independent from clustering algorithms but features structure capturing.

The major drawback of the design strategy arises from its numerical character, the class space has to be discretised in order to distribute the anti–objects. As a consequence their creation is computationally costly.

4 Conclusions

This paper presents a data driven approach toward modelling of complex data–inherent structures. Its main philosophy is the exclusive usage of a general fuzzy modelling framework. Instead of applying cluster analysis techniques the design strategy aims to complete the convex fuzzy pattern model with the help of so called anti–objects. These anti–objects are not available prior to the design. They have been generated and distributed over the class space. For this purpose a Fibonacci expansion model was elaborated and demonstrated.
With the help of the introduced automated FPC design the anti–objects have been agglomerated to a negating fuzzy pattern classes (FPAC). By preserving the membership function concept the FPAC are on one hand afflicted again with the drawbacks of fuzzy pattern classes. But on the other hand the same design approach can be applied again, eventually leading to a pure convex fuzzy pattern model. A possible combination of the setup FPACs and the original fuzzy pattern model to a hierarchical nonconvex overall FPC model has been demonstrated with the help of an example.

Finally it has to be stressed that the introduced design strategy is a universal approach in so far as it can be applied to any multivariate unimodal convex parametrical membership function.

Further points of research are:
- reducing the numerical character of the method
- combination with cluster algorithms
- creation of fuzzy pattern class and anti–class networks

References

[1] Gernot Herbst and Steffen F. Bocklisch. Classification of keystroke dynamics - A case study of fuzzified discrete event handling. In Bengt Lennartson, Martin Fabian, Knut Akesson, Alessandro Giua, and Ratnesh Kumar, editors, *9th International Workshop on Discrete Event Systems*, 2008.

[2] Arne-Jens Hempel and Steffen F. Bocklisc. Hierarchical modelling of data inherent structures using networks of fuzzy classifiers. In *Tenth International Conference on Computer Modeling and Simulation, 2008. UKSIM 2008.*, pages 230–235. IEEE, 2008. ISBN: 0-7695-3114-8.

[3] C. Weihs & W. Gaul, editor. *Classification - the Ubiquitous Challenge: Proceedings of the 28th Annual Conference of the Gesellschaft Fur Klassifikation E.V., University of Dortmund*, volume 28. Springer, June 2005.

[4] Lotfi Asker Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.

[5] James C. Bezdek, James Keller, Raghu Krisnapuram, and Nikhil R. Pal. *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing (The Handbooks of Fuzzy Sets)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

[6] B. Schölkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. The MIT Press, 2006.

[7] Xuehua Li and Lan Shu. Fuzzy theory based support vector machine classifier. *Fuzzy Systems and Knowledge Discovery, Fourth International Conference on*, 1:600–604, 2008.

[8] Steffen F. Bocklisch. *Prozeßanalyse mit unscharfen Verfahren*. VEB Verlag Technik, Berlin, 1987.

[9] Arne-Jens Hempel. Aggregation und Identifikation von Fuzzy−Objekten mit unterschiedlichen elementaren Unschärfen. Master's thesis, TU Chemnitz, 2005.

[10] Michael Päßler. Mehrdimensionale Zeitreihenmodellierung und Prognose mittels Fuzzy Pattern Modellen. Master's thesis, TU Chemnitz, 1998.

[11] Ulrich Scheunert. *Neue Verfahren der Fuzzy−Mengen−Verknüpfung und Fuzz−Arithmetik und ihrer Anwendung bei der Sensor−Daten−Fusion*. PhD thesis, TU Chemnitz, 2001.

[12] Steffen F. Bocklisch and Norman Bitterlich. *Fuzzy systems in computer science*, chapter Fuzzy pattern classification – methodology and application. Vieweg, Wiesbaden, 1994.

[13] Leszek Rutkowski. *Flexible Neuro-fuzzy Systems: Structures, Learning and Performance Evaluation (Kluwer International Series in Engineering and Computer Science)*. Kluwer Academic Publishers, Norwell, MA, USA, 2004.

[14] H. J. Zimmermann. *Fuzzy Set Theory and its Applications*. Springer, 4 edition, October 2001.

ISBN: 978-989-95079-6-8

On Triangular Norm Based Fuzzy Description Logics

Àngel García-Cerdaña Eva Armengol Francesc Esteva

IIIA, Artificial Intelligence Research Institute
CSIC, Spanish Council for Scientific research
Campus UAB, 08193 Bellaterra, Spain
Email: {angel, eva, esteva}@iiia.csic.es

Abstract— *Description Logics (DLs) are knowledge representation languages useful to represent concepts and roles. Fuzzy Description Logics (FDLs) incorporate both vague concepts and vague roles modeling them as fuzzy sets and fuzzy relations respectively. In the present paper, following ideas from Hájek, we propose the use of t-norm based (fuzzy) logics with truth constants in the language as logics underlying the fuzzy description language. We introduce the languages $\mathcal{ALC}_{L^*(\mathbf{S})}$ and $\mathcal{ALC}_{L^*_\sim(\mathbf{S})}$ as an adequate syntactical counterpart of some semantic calculi given in different works dealing with FDLs.*

Keywords— Description Logics, Fuzzy Description Logics, t-Norm Based Fuzzy Logics, Truth-constants, Involutive negation.

1 Introduction

Description Logics (DLs) are knowledge representation languages (particularly suited to specify formal ontologies), which have been studied extensively over the last two decades. A full reference manual of the field is [1]. The vocabulary of DLs consists of *concepts*, which denote sets of individuals, and *roles*, which denote binary relations among individuals. From atomic concepts and roles DL systems allow, by means of *constructors*, to build complex descriptions of both concepts and roles, which are used to describe a domain through a knowledge base (KB) containing the definitions of relevant domain concepts or some hierarchical relationships among them (TBox) and a specification of properties of the domain instances (ABox). One of the main issues of DLs is the fact that the semantics is given in a Tarski-style presentation and the statements in both TBox and ABox can be identified with formulae in first-order logic or a slight extension of it, and hence we can use reasoning to obtain implicit knowledge from the explicit knowledge in the KB.

A natural generalization to cope with vague concepts and relations consists in interpret DL concepts and roles as fuzzy sets and fuzzy relations, respectively. From this point of view, it is at the end of the last decade (from 1998) when several proposals of Fuzzy Description Logics (FDLs) were introduced (e.g., the first ones by Yen [19], Tresp and Molitor [18] and Straccia [14]). However, the logic framework behind these initial works is very limited. With the aim of enriching the expressive possibilities Hájek [9] proposes to take t-norm based fuzzy logics as logics underlying FDLs. This change of view gives a wide number of choices on which a DL can be based: for every particular problem we can consider the t-norm based (fuzzy) logic that seems to be more adequate. As an example, Hájek studies the FDL associated with the description language $\mathcal{ALC}$. After this work, several researchers on FDLs have developed approaches based on the spirit of Hájek's paper, even though their work is more related to expressiveness and algorithms than in its logical base (see for instance [16, 13, 17]).

The main motivation of the present work is based on the following consideration: since the axioms of the bases of knowledge in FDLs include truth degrees (see for instance [14]), a natural choice is to include symbols for these degrees in both, the description language and, as truth constants, in the t-norm based logic where that language is interpreted.

To this goal in the present paper we propose two new families of description languages, denoted by $\mathcal{ALC}_{L^*(\mathbf{S})}$ and $\mathcal{ALC}_{L^*_\sim(\mathbf{S})}$ that are extensions of the language $\mathcal{ALC}$ considered by Hájek in [9]. After some introductory notions we define their semantics and describe the corresponding knowledge base (TBox and ABox) from a syntactic and semantic perspective and, taking advantage of having truth constants in the logic, we define graded notions of validity, satisfiability and subsumption. We also give some representative example and some new results for the case of $\mathcal{ALC}$ language over Gödel logic with truth constants and an involutive negation.

2 Fuzzy Logic: basics

In the last decade a family of fuzzy logics as t-norm based fuzzy logics has defined and studied (see, for instance, the monograph [8] for the main notions used in this section). They are multi-valued systems with additive conjunction and disjunction, multiplicative conjunction, implication, negation and the constant $\bar{0}$ which are interpreted in $[0, 1]$ as $\min$, $\max$, a continuous t-norm $*$, its residuum $\rightarrow_*$, the negation function $n(x) = x \rightarrow_* 0$ and 0, respectively. This interpretation is mainly defined by a continuous t-norm and its residuum which justify the name t-norm based logics. Remember that the main continuous t-norms are the minimum, the product and the Lukasiewicz since all other continuous t-norm are ordinal sum of these three basic ones.

2.1 From BL to the logic of a continuous t-norm

The Basic fuzzy Logic (BL) (defined in [8]) has the following basic connectives: *multiplicative conjunction* ($\&$) *implication* ($\rightarrow$) (both binary) and *falsity* ($\bar{0}$) (nullary). It is defined by the inference rule of *Modus Ponens* and the following schemata (taking $\rightarrow$ as the least binding connective):

ISBN: 978-989-95079-6-8

(BL1) $(\varphi \to \psi) \to ((\psi \to \chi) \to (\varphi \to \chi))$
(BL2) $\varphi\&\psi \to \varphi$
(BL3) $\varphi\&\psi \to \psi\&\varphi$
(BL4) $\varphi\&(\varphi \to \psi) \to \psi\&(\psi \to \varphi)$
(BL5a) $(\varphi \to (\psi \to \chi)) \to (\varphi\&\psi \to \chi)$
(BL5b) $(\varphi\&\psi \to \chi) \to (\varphi \to (\psi \to \chi))$
(BL6) $((\varphi \to \psi) \to \chi) \to (((\psi \to \varphi) \to \chi) \to \chi)$
(BL7) $\bar{0} \to \varphi$

The usual defined connectives are introduced as follows:

$\varphi \vee \psi := ((\varphi \to \psi) \to \psi) \wedge ((\psi \to \varphi) \to \varphi)$,
$\varphi \wedge \psi := \varphi\&(\varphi \to \psi)$, $\varphi \leftrightarrow \psi := (\varphi \to \psi)\&(\psi \to \varphi)$,
$\neg\varphi := \varphi \to \bar{0}$, $\bar{1} := \neg\bar{0}$.

Łukasiewicz, Product and Gödel Logics can be obtained as axiomatic extensions of BL with the following axioms: $\neg\neg\varphi \to \varphi$ for Łukasiewicz; $(\varphi \wedge \neg\varphi) \to \bar{0}$ and $\neg\neg\chi \to (((\varphi\&\chi) \to (\psi\&\chi)) \to (\varphi \to \psi))$ for Product; and $\varphi \to \varphi\&\varphi$ for Gödel.

An *evaluation of propositional variables* is a mapping e assigning to each variable p a *truth value* $e(p) \in [0,1]$. Given a continuous t-norm $*$, the evaluation e is extended inductively to a mapping of all formulas into the so-called *standard algebra* $[0,1]_* = \langle [0,1], *, \to_*, \max, \min, 0, 1\rangle$ defined on [0,1] by the t-norm and its residuum in the following way: $e(\varphi\&\psi) = e(\varphi) * e(\psi)$; $e(\varphi \to \psi) = e(\varphi) \to_* e(\psi)$; $e(\bar{0}) = 0$. In [3] it is proved that the formal system BL is sound and complete w.r.t. the standard algebras defined in $[0,1]$ by continuous t-norms and their residua, i.e., a formula φ is provable in BL if and only if it is a common tautology of all standard algebras defined by a continuous t-norm and its residuum. Consequently we say that BL is the logic of continuous t-norms and their residua. It is also well known that Łukasiewicz (Product, Gödel) Logic is the logic of the t-norm of Łukasiewicz (Product, Gödel) and its residuum in the sense that a formula is provable in each logic if and only if it is a tautology over the standard algebra defined by the corresponding t-norm and its residuum.

In [6] the logic L^* of each continuous t-norm $*$ and its residuum is proved to be finitely axiomatizable as extension of BL. Moreover it is also given an algorithm to find a finite set of axioms characterizing each logic L^*. When the negation $\neg\varphi := \varphi \to \bar{0}$ defined in L^* is not involutive, a new logic $L^*_\sim$, expanding L^* by adding an involutive negation, can be considered. This negation, denoted by $\sim$, could be introduced, as is done in the context of intuitionistic logic (see [11]) or in the context of Gödel or Product logics (cf.[5]) by adding the axioms

(~ 1) $\sim\sim\varphi \to \varphi$
(~ 2) $\sim(\varphi \vee \psi) \leftrightarrow (\sim\varphi \wedge \sim\psi)$
(~ 3) $\neg\varphi \to \sim\varphi$

Notice that having an involutive negation in the logic enriches the representational power of the logical language in a non-trivial way because a multiplicative (or strong) disjunction $\varphi \veebar \psi$ is definable now (by duality) as $\sim(\sim\varphi\& \sim\psi)$, being the associated truth function $\oplus$ defined as $x \oplus y := n(n(x) * n(y))$, where n is the truth function of $\sim$ and, on the other hand, a contrapositive implication $\varphi \hookrightarrow \psi$ is definable as $\sim\varphi \veebar \psi$, with truth function $\hookrightarrow_\oplus$ defined as $x \hookrightarrow_\oplus y = n(x) \oplus y$.

2.2 The logics $BL\forall$ and $L^*\forall$

The language of the *basic fuzzy predicate logic* $BL\forall$ consists of a set of *predicate symbols* $\mathcal{P} = \{P, Q, \dots\}$, each together with its *arity* $n \geq 1$ and a set of *constant symbols* $\mathcal{C} = \{c, d, \dots\}$. The logical symbols are *variable symbols* $x, y, \dots$, *connectives* $\&, \to, \bar{0}$ and *quantifiers* $\forall, \exists$. Other connectives $(\vee, \wedge, \neg, \leftrightarrow, \bar{1})$ are defined as in BL. *Terms* are constant symbols and variable symbols. An *atomic formula* is an expression of the form $P(t_1, \dots, t_n)$, where P is a predicate letter of arity n and $t_1, \dots, t_n$ are terms. The set of predicate formulas is built from atomic formulas in the usual way.

A *fuzzy interpretation* for our language is a tuple $\mathbf{M} = \langle M, (r_P)_{P\in\mathcal{P}}, (m_c)_{c\in\mathcal{C}}\rangle$, where M is a set, for each n-ary predicate symbol P, r_P is a fuzzy n-ary relation $M^n \to [0,1]$; and for each constant symbol c, m_c is an element of M.

Given a continuous t-norm $*$, an $\mathbf{M}$-*evaluation* of the variables assigns to each variable x an element $v(x)$ of M. From M and v we define the *truth value of a term* t in the following way: $\|t\|_{\mathbf{M},v} = v(t)$ when t is a variable, and $\|t\|_{\mathbf{M},v} = m_c$ when t is a constant c. The *truth value of a formula* φ *for an evaluation* v, denoted by $\|\varphi\|^*_{\mathbf{M},v}$, is a value in $[0,1]$ defined inductively as follows:

$$\begin{array}{ll} \|\varphi\|_{\mathbf{M},v} & \text{if } \varphi \text{ is an atomic formula,} \\ 0 & \text{if } \varphi = \bar{0}, \\ \|\alpha\|^*_{\mathbf{M},v} * \|\beta\|^*_{\mathbf{M},v} & \text{if } \varphi = \alpha\&\beta, \\ \|\alpha\|^*_{\mathbf{M},v} \to_* \|\beta\|^*_{\mathbf{M},v} & \text{if } \varphi = \alpha \to \beta, \\ \inf\{\|\alpha\|^*_{\mathbf{M},v} : v \equiv_x v'\} & \text{if } \varphi = (\forall x)\alpha, \\ \sup\{\|\alpha\|^*_{\mathbf{M},v} : v \equiv_x v'\} & \text{if } \varphi = (\exists x)\alpha, \end{array}$$

where $v \equiv_x v'$ means that for all variables $y \neq x$, $v(y) = v'(y)$.

The *truth value of a formula* φ is defined by $\|\varphi\|^*_{\mathbf{M}} := \inf\{\|\varphi\|^*_{\mathbf{M},v} : v \text{ is an } \mathbf{M}\text{-evaluation}\}$. A formula φ is a $*$-*tautology* if $\|\varphi\|^*_{\mathbf{M}} = 1$ for every fuzzy interpretation $\mathbf{M}$; φ is a *standard tautology* if it is a $*$-*tautology* for each continuous t-norm $*$. The following standard tautologies are taken as axioms of the basic fuzzy predicate logic $BL\forall$ (see [8]): *a)* the axioms of BL; *b)* the following axioms on quantifiers:

$(\forall 1)$ $(\forall x)\varphi(x) \to \varphi(t)$ (t substitutable for x in $\varphi(x)$),
$(\exists 1)$ $\varphi(t) \to (\exists x)\varphi(x)$ (t substitutable for x in $\varphi(x)$),
$(\forall 2)$ $(\forall x)(\varphi \to \psi) \to (\varphi \to (\forall x)\psi)$ (x not free in φ),
$(\exists 2)$ $(\forall x)(\varphi \to \psi) \to ((\exists x)\varphi \to \psi)$ (x not free in ψ),
$(\forall 3)$ $(\forall x)(\varphi \vee \psi) \to (\forall x)\varphi \vee \psi$ (x not free in ψ).

Deduction rules are (as in classical logic) *Modus Ponens* and *Generalization*. Notions of *proof*, *provability*, *theory*, etc., are defined in the usual way. Let $\mathcal{C}$ be an axiomatic extension of BL. $\mathcal{C}\forall$ is obtained by taking the axioms and rules of $BL\forall$ plus the axioms characterizing $\mathcal{C}$. Thus, given a continuous t-norm $*$, the predicate logic $L^*\forall$ is the logic obtained from L^* by adding to its axiomatization the schemas for quantifiers and the rule of generalization.

2.3 Adding truth constants to the language

T-norm based logics are infinite-valued logics. However, the advantage of being a many-valued logic is not used in current approaches since the semantic deduction of formulas do not take into account the *intermediate or partial truth degrees*. That is to say, current approaches use a truth-preserving con-

ISBN: 978-989-95079-6-8

sequence relation in the same way as in the classical logic, i.e. deduce true formulas (having value 1) from sets of true formulas. An elegant way to take advantage from being many-valued is to introduce truth constants into the language, as it is done by Pavelka in [12] and more recently in [8, 6, 4]. The approach considered in this paper is based in these ideas.

Given a continuous t-norm $*$, its residuum $\rightarrow_*$ and its corresponding logic L^*, let $\mathbf{S} = \langle S, *, \rightarrow_*, \max, \min, 0, 1\rangle$ be a *countable* (i.e., finite or enumerable) *subalgebra* of the corresponding standard algebra $[0,1]_*$. The expansion of L^* adding into the language a truth constant $\bar{r}$ for each $r \in S$, denoted by $L^*(\mathbf{S})$, is defined as follows:

i) the language of $L^*(\mathbf{S})$ is the one of L^* plus a truth constant $\bar{r}$ for each $r \in S$,

ii) the axioms and rules of $L^*(\mathbf{S})$ are those of L^* plus the *book-keeping* axioms: for each $r, s \in S \setminus \{0,1\}$, $\bar{r}\&\bar{s} \leftrightarrow \overline{r * s}$ and $(\bar{r} \rightarrow \bar{s}) \leftrightarrow \overline{r \rightarrow_* s}$.

Completeness results for propositional logic $L^*(\mathbf{S})$ when $*$ is a continuous t-norm has been fully studied in [4].

2.4 *Defining the logics $L^*(\mathbf{S})\forall$ and $L^*_\sim(\mathbf{S})\forall$*

When the negation associated to the continuous t-norm $*$ is not involutive, the logic $L^*_\sim(\mathbf{S})$ can be defined in a similar way although in this case $\mathbf{S}$ has to be a countable subalgebra of the algebra obtained by adding the truth function of the involutive negation $n(x) := 1 - x$ to the operations of $[0,1]_*$. Moreover we need to add the book-keeping axioms for the involutive negation: $\sim \bar{r} \leftrightarrow \overline{n(r)}$. The corresponding predicate logics $L^*(\mathbf{S})\forall$ and $L^*_\sim(\mathbf{S})\forall$ are respectively obtained from $L^*\forall$ and $L^*_\sim\forall$ by expanding the language with a truth constant $\bar{r}$ for every $r \in S$ and by adding the book keeping axioms. The truth value of the formula $\bar{r}$ is given by $\|\bar{r}\|^*_{\mathbf{M}} = r$. The logics $L^*(\mathbf{S})\forall$ and $L^*_\sim(\mathbf{S})\forall$ will be the basis of our proposal for the description languages presented in the next section.

3 The description languages $\mathcal{ALC}_{L^*(\mathbf{S})}$ and $\mathcal{ALC}_{L^*_\sim(\mathbf{S})}$

Similarly as first-order logic is the counterpart for interpreting the classic description language $\mathcal{ALC}$, the logics $L^*(\mathbf{S})\forall$ and $L^*_\sim(\mathbf{S})\forall$ will be the counterpart for interpreting the description languages $\mathcal{ALC}_{L^*(\mathbf{S})}$ and $\mathcal{ALC}_{L^*_\sim(\mathbf{S})}$ we will define in this section. In these languages, take an special role the so-called evaluated formulas. An *evaluated formula* is a formula of one of the types $\bar{r} \rightarrow \varphi$, $\varphi \rightarrow \bar{r}$ where φ is a formula without new truth constants (i.e., different from $\bar{0}$ and $\bar{1}$). In this setting $\bar{r} \leftrightarrow \varphi$ is definable as $(\bar{r} \rightarrow \varphi)\&(\varphi \rightarrow \bar{r})$. The name of evaluated formulas cames from the fact that $e(\bar{r} \rightarrow \varphi) = 1$ (resp. $e(\varphi \rightarrow \bar{r}) = 1$) if and only if $e(\varphi) \geq r$ (resp. $e(\varphi) \leq r$). Thus, evaluated formulas correspond to the type of formulas (under the notation $\langle \varphi, \succcurlyeq r\rangle$ and $\langle \varphi, \preccurlyeq r\rangle$) used for the knowledge bases in papers on FDLs (cf.[14, 15, 17]). Next we define the description languages $\mathcal{ALC}_{L^*(\mathbf{S})}$ and $\mathcal{ALC}_{L^*_\sim(\mathbf{S})}$ from a syntactic and semantic perspective. Then we introduce the notions of *TBox* and *ABox* for that languages.

Syntax. In the languages of description we start from *atomic concepts* and *atomic roles*. Complex descriptions are built inductively with constructors of concepts. We will use the letters A for atomic concepts, R for atomic roles and both C and D for descriptions of concepts. Using the connectives $\bar{0}, \&, \rightarrow$ (falsity, conjunction, implication), the quantifiers $\forall, \exists$ and the point . as an auxiliary symbol, the description of concepts in classic $\mathcal{ALC}$ can be built using the following syntactical rules

$$C, D \rightsquigarrow A \mid \bar{0} \mid C\&D \mid C \rightarrow D \mid \forall R.C \mid \exists R.C$$

Given a continuous t-norm $*$ and a countable subalgebra $\mathbf{S}$ of the corresponding standard algebra $[0,1]_*$, let us consider the logic $L^*(\mathbf{S})$. We define $\mathcal{ALC}_{L^*(\mathbf{S})}$ by adding to $\mathcal{ALC}$, for every $r \in S$, a nullary connective $\bar{r}$ and the rule $C \rightsquigarrow \bar{r}$.

The language $\mathcal{ALC}_{L^*_\sim(\mathbf{S})}$ is defined by adding to $\mathcal{ALC}_{L^*(\mathbf{S})}$ the connective $\sim$ and the syntactic rule $C \rightsquigarrow \sim C$.

Following [8], the notions of *instance of a concept* and *instance of a role* allow us to read the formulas of both languages as formulas of the corresponding predicate fuzzy logic. For each term t (variable or constant). The *instance* $D(t)$ of a concept D is defined as follows:

$$\begin{array}{ll} A(t), & \text{if } D \text{ is an atomic concept } A,\\ \bar{0}, & \text{if } D = \bar{0},\\ \sim C(t), & \text{if } D = \sim C,\\ C_1(t) \circ C_2(t), & \text{if } D = C_1 \circ C_2, \text{ where } \circ \in \{\&, \rightarrow\},\end{array}$$

and, if y is a variable not occurring in $C(t)$,

$$\begin{array}{ll} (\forall y)(R(t,y) \rightarrow C(y)), & \text{if } D = \forall R.C,\\ (\exists y)(R(t,y)\&C(y)), & \text{if } D = \exists R.C,\end{array}$$

where, given two terms t_1 and t_2, $R(t_1, t_2)$ is an atomic formula corresponding to the atomic role R. We will refer to the expressions of the form $R(t_1, t_2)$ as *instances of the atomic role R*.

Semantics. According to semantics for $L^*(\mathbf{S})\forall$ and $L^*_\sim(\mathbf{S})\forall$ a fuzzy interpretation $\mathbf{M}$ associates a fuzzy set $A^{\mathbf{M}}$ to each atomic concept A and a fuzzy binary relation $R^{\mathbf{M}}$ to each atomic role R, and the truth value for complex descriptions is given as follows:

$$\begin{array}{rcl} \bar{0}^{\mathbf{M}}(a) & = & 0\\ \bar{r}^{\mathbf{M}}(a) & = & r, \text{ for every } r \in S\\ (C\&D)^{\mathbf{M}}(a) & = & C^{\mathbf{M}}(a) * D^{\mathbf{M}}(a)\\ (C \rightarrow D)^{\mathbf{M}}(a) & = & C^{\mathbf{M}}(a) \rightarrow_* D^{\mathbf{M}}(a)\\ (\forall R.C)^{\mathbf{M}}(a) & = & \inf\{R^{\mathbf{M}}(a,b) \rightarrow_* C^{\mathbf{M}}(b) : b \in M\}\\ (\exists R.C)^{\mathbf{M}}(a) & = & \sup\{R^{\mathbf{M}}(a,b) * C^{\mathbf{M}}(b) : b \in M\}\end{array}$$

In the case of complex descriptions in $\mathcal{ALC}_{L^*_\sim(\mathbf{S})}$ we must to add the following:

$$(\sim C)^{\mathbf{M}}(a) \quad = \quad 1 - C^{\mathbf{M}}(a)$$

Fuzzy TBox and Fuzzy ABox. Now we define the notions of $TBox$ and $ABox$ for $\mathcal{ALC}_{L^*(\mathbf{S})}$ and $\mathcal{ALC}_{L^*_\sim(\mathbf{S})}$. In these definitions we use the following graded notion of inclusion between fuzzy sets: $\textit{degree}(C \subseteq D) = \inf_x(C(x) \rightarrow_* D(x))$. Of course this degree is 1 if and only if $C(x) \leq D(x)$ for all x and 0 if the support[1] of the two fuzzy sets are disjoint. Having the truth constants in the language allows us to associate sentences like "*degree* $(C \subseteq D) \leq r$" with formulas such as $(\forall x)(C(x) \rightarrow D(x)) \rightarrow \bar{r}$.

A *fuzzy concept inclusion axiom* is a sentence of one of the following forms:

- $\bar{r} \rightarrow (\forall x)(C(x) \rightarrow D(x))$

[1] The *support* of a fuzzy set is the cardinal of the set of elements which membership degree is greater than 0.

ISBN: 978-989-95079-6-8

Table 1: The graded notation for fuzzy KB.

Evaluated Formula	FDL Graded Notation
$\bar{r} \rightarrow (\forall x)(C(x) \rightarrow D(x))$	$\langle C \sqsubseteq D, \succcurlyeq \bar{r} \rangle$
$(\forall x)(C(x) \rightarrow D(x)) \rightarrow \bar{r}$	$\langle C \sqsubseteq D, \preccurlyeq \bar{r} \rangle$
$\bar{r} \leftrightarrow (\forall x)(C(x) \rightarrow D(x))$	$\langle C \sqsubseteq D, \approx \bar{r} \rangle$
$\bar{r} \rightarrow C(a)$	$\langle a : C, \succcurlyeq \bar{r} \rangle$
$C(a) \rightarrow \bar{r}$	$\langle a : C, \preccurlyeq \bar{r} \rangle$
$\bar{r} \leftrightarrow C(a)$	$\langle a : C, \approx \bar{r} \rangle$
$\bar{r} \rightarrow R(a,b)$	$\langle (a,b) : R, \succcurlyeq \bar{r} \rangle$
$R(a,b) \rightarrow \bar{r}$	$\langle (a,b) : R, \preccurlyeq \bar{r} \rangle$
$\bar{r} \leftrightarrow R(a,b)$	$\langle (a,b) : R, \approx \bar{r} \rangle$

- $(\forall x)(C(x) \rightarrow D(x)) \rightarrow \bar{r}$
- $\bar{r} \leftrightarrow (\forall x)(C(x) \rightarrow D(x))$

A *fuzzy assertion axiom* is a sentence of one of the following forms:

- $\bar{r} \rightarrow C(a)$ or $\bar{r} \rightarrow R(a,b)$
- $C(a) \rightarrow \bar{r}$ or $R(a,b) \rightarrow \bar{r}$
- $C(a) \leftrightarrow \bar{r}$ or $R(a,b) \leftrightarrow \bar{r}$

Now a *fuzzy TBox* is defined as a finite set of *fuzzy concept inclusion axioms* while a *fuzzy ABox* is defined as a finite set of *fuzzy assertion axioms*. A *fuzzy KB* is a pair $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$, where the first component is a *fuzzy TBox* and the second one is a *fuzzy ABox*.

Notice that all the axioms of the *fuzzy KB* are evaluated formulas. Thus, the syntactic notion of *fuzzy KB* according to our approach, both the $TBox$ and the $ABox$ can be seen as theories of the logic $L^*(\mathbf{S})\forall$ (or $L^*_{\sim}(\mathbf{S})\forall$).

In Tab. 3 we present an alternative graded notation for fuzzy inclusion and fuzzy assertions that we will use in the example. This notation is similar to the one used in some papers of FDLs (see for instance [17]). Moreover this notation is according to the semantical interpretation in the sense that, for instance,

$$\langle \sigma, \succcurlyeq \bar{r} \rangle^{\mathbf{M}} = r \rightarrow_* \sigma^{\mathbf{M}}$$

and so

$$\mathbf{M} \models \langle \sigma, \succcurlyeq \bar{r} \rangle \quad \text{iff} \quad r \rightarrow_* \sigma^{\mathbf{M}} = 1 \quad \text{iff} \quad \sigma^{\mathbf{M}} \geq r$$

It is interesting to remark that in $\mathcal{ALC}_{L^*_{\sim}(\mathbf{S})\forall}$ the involutive negation allows us to define graded expressions like, for instance, $\langle a : C, \succ \bar{r} \rangle$ as $\sim \langle a : C, \preccurlyeq \bar{r} \rangle$ which corresponds to the formula $\sim (C(a) \rightarrow \bar{r})$.

An example. We will use a data set composed of nine robots (Fig. 1), each one with either the same of different shape of head and body (i.e., they are homogeneous or not homogeneous respectively), they can or cannot wear a tie, they can or cannot smile, and they hold some object. Taking into account all these characteristics, robots can have different *friendliness* degree. The domain of interpretation of the robots is the set: $M_{\mathcal{R}} = \{r_i : 1 \leq i \leq 9\} \cup \{o_i : 1 \leq i \leq 9\}$, where the r_i are the robots and each o_i is the object that the robot r_i holds (e.g., the object o_4 is the flower that r_4 holds). Atomic concepts of the language are the following: Robot, Happy, Object, FriendlyObject, Homogeneous, Balloon, Flag, Flower, Sword, Ax and HasTie. There is only one atomic role: hasObject.

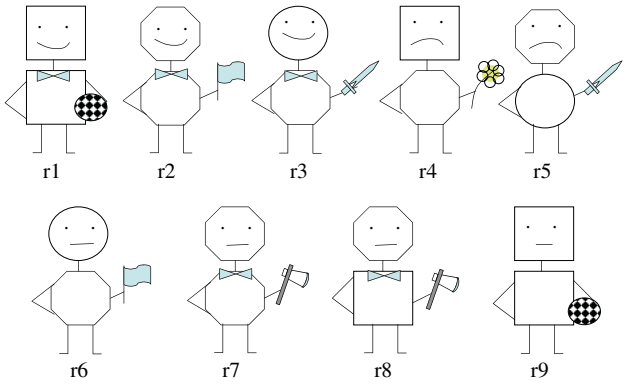


Figure 1: The 9 little robots.

The $TBox$ concerning the robots domain is the following:

Friendly $\equiv$ Robot&(∃hasObject.FriendlyObject)&(Happy $\veebar$ Homogeneous)
$\langle$Robot & Object $\sqsubseteq \bar{0}, \approx \bar{1}\rangle$
$\langle \bar{1} \sqsubseteq$ Robot $\veebar$ Object, $\approx \bar{1}\rangle$
$\langle$Flower $\sqsubseteq$ FriendlyObject, $\approx \bar{1}\rangle$
$\langle$Balloon $\sqsubseteq$ FriendlyObject, $\approx \overline{0.75}\rangle$
$\langle$Flag $\sqsubseteq$ FriendlyObject, $\approx \overline{0.50}\rangle$
$\langle$Sword $\sqsubseteq$ FriendlyObject, $\approx \overline{0.25}\rangle$
$\langle$Ax $\sqsubseteq$ FriendlyObject, $\approx \bar{0}\rangle$

where $C \veebar D$ is an abbreviation of $\sim (\sim C \& \sim D)$ and $C \equiv D$ is an abbreviation for the conjunction of the formulas $\langle C \sqsubseteq D, \approx \bar{1} \rangle$ and $\langle D \sqsubseteq C, \approx \bar{1} \rangle$.

Notice that objects have different friendliness degree. For instance, a sword is a friendly object with degree 0.25 and an ax is a friendly object with degree 0 (i.e., it should be considered unfriendly in the classical case). On the other hand, the $TBox$ also contains a definition of Friendly allowing to assess the friendliness degree of a robot.

The $ABox$ containing the descriptions of the robots is the following:

For each i, $1 \leq i \leq 9$, $\langle r_i$: Robot, $\approx \bar{1}\rangle$, $\langle (r_i, o_i)$: hasObject, $\approx \bar{1}\rangle$
$\langle r_1$: Homogeneous, $\approx \bar{1}\rangle$, $\langle o_1$: Balloon, $\approx \bar{1}\rangle$, $\langle r_1$: Happy, $\approx \bar{1}\rangle$
$\langle r_2$: Homogeneous, $\approx \bar{1}\rangle$, $\langle o_2$: Flag, $\approx \bar{1}\rangle$, $\langle r_2$: Happy, $\approx \bar{1}\rangle$
$\langle r_3$: Homogeneous, $\approx \overline{0.75}\rangle$, $\langle o_3$: Sword, $\approx \bar{1}\rangle$, $\langle r_3$: Happy, $\approx \bar{1}\rangle$
$\langle r_4$: Homogeneous, $\approx \overline{0.50}\rangle$, $\langle o_4$: Flower, $\approx \bar{1}\rangle$, $\langle r_4$: Happy, $\approx \bar{0}\rangle$
$\langle r_5$: Homogeneous, $\approx \overline{0.50}\rangle$, $\langle o_5$: Sword, $\approx \bar{1}\rangle$, $\langle r_5$: Happy, $\approx \bar{0}\rangle$
$\langle r_6$: Homogeneous, $\approx \overline{0.75}\rangle$, $\langle o_6$: Flag, $\approx \bar{1}\rangle$, $\langle r_6$: Happy, $\approx \overline{0.50}\rangle$
$\langle r_7$: Homogeneous, $\approx \bar{1}\rangle$, $\langle o_7$: Ax, $\approx \bar{1}\rangle$, $\langle r_7$: Happy, $\approx \overline{0.50}\rangle$
$\langle r_8$: Homogeneous, $\approx \overline{0.75}\rangle$, $\langle o_8$: Ax, $\approx \bar{1}\rangle$, $\langle r_8$: Happy, $\approx \overline{0.50}\rangle$
$\langle r_9$: Homogeneous, $\approx \bar{1}\rangle$, $\langle o_9$: Balloon, $\approx \bar{1}\rangle$, $\langle r_9$: Happy, $\approx \overline{0.50}\rangle$

Notice that, using truth constants, we can assess different degrees of homogeneity according to the shape of both head and body. In particular, we assess, in a subjective way, that a combination of round shapes of head and body (i.e., a circle and an octagon) give a more homogeneous aspect to the robot than combining round and square shapes. Thus, robots r_6 and r_8 are considered more homogeneous than robot r_4. Similarly, robots have different form of mouth that give them different degree of happiness (i.e., robot r_1 is assessed as more happy than robots r_8 and r_4).

Fuzzy reasoning. Reasoning in fuzzy description logics consists on the same kind of tasks than in the classical case but

ISBN: 978-989-95079-6-8

now they depend on the chosen continuous t-norm $*$. One of the advantages of introducing truth constants in the language is the possibility to define the graded versions of the notions of $*$-satisfiability, $*$-subsumption and $*$-validity defined in [9] without modifying the semantics. Thus, given a concept C, a t-norm $*$ and a truth value $r \in S$ we introduce the following graded notions with respect to a knowledge base $\mathcal{K}$:

• C is *$*$-satisfiable in a degree greater or equal than r* iff there is a model $\mathbf{M}$ of $\mathcal{K}$ such that $\|\bar{r} \to C(a)\|^*_{\mathbf{M}} = 1$ (being a a constant).

• C is *$*$-valid in a degree greater or equal than r* in a model $\mathbf{M}$ of $\mathcal{K}$ iff $\|\bar{r} \to (\forall x)C(x)\|^*_{\mathbf{M}} = 1$.

• C is *$*$-subsumed* by D *in a degree greater or equal than* r in a model $\mathbf{M}$ of $\mathcal{K}$ iff the concept $C \to D$ is *$*$-valid in a degree greater or equal than r* in the model $\mathbf{M}$, that is, iff $\|\bar{r} \to (\forall x)(C(x) \to D(x))\|^*_{\mathbf{M}} = 1$.

On the other hand, we can analogously define the notions of lower thresholds. For instance, a concept C is *$*$-satisfiable in a degree lower or equal than r* iff $\|C(a) \to \bar{r}\|^*_{\mathbf{M}} = 1$ for some model $\mathbf{M}$ of $\mathcal{K}$. Moreover, it is also possible to define an interval where a concept is either valid, satisfiable or subsumed. For instance a concept C is *$*$-satisfiable in an interval of degrees* $[r, s]$ iff $\|\bar{r} \to C(a)\|^*_{\mathbf{M}} = 1$ and $\|C(a) \to \bar{s}\|^*_{\mathbf{M}} = 1$ for some model $\mathbf{M}$. In particular, when $r = s$ the interval became a value called the *degree of satisfiability*.

Example 3.1 *The concept C below is $*$-satisfiable with degree 0.75 in the robots model using any continuous t-norm.*

$$C := \mathsf{Homogeneous} \,\&\exists \mathsf{hasObject}.\mathsf{FriendlyObject}$$

According to the semantics, C is $*$-satisfiable with degree 0.75 if there is at least one robot such that

$C(x) = \mathsf{Homogeneous}(x) * \sup_{y \in M_{\mathcal{R}}}(\mathit{hasObject}(x, y) * \mathit{FriendlyObject}(y)) = 0.75$

The equality is true since the robots r_1 and r_9 have $\mathsf{Homogeneous}(x) = 1$ and both hold a ballon that, as the $TBox$ states, it is a friendly object with degree 0.75. Thus, because 1 is the unity element of any t-norm, both robots have friendliness degree 0.75.

Example 3.2 *Let us to analyze the $*$-subsumption degree of the concept* Object & Sword & FriendlyObject *by* $\bar{0}$ *with respect to the $TBox$.*

According to the definition of $*$-subsumption, we have to analyze the following formula

$$\bar{r} \leftrightarrow (\forall x)((\mathit{Object} \,\&\, \mathit{Sword} \,\&\, \mathit{FriendlyObject})(x) \to \bar{0})$$

and thus, according to the interpretation, the $*$-subsumption has degree

$$r = \inf_{x \in M_{\mathcal{R}}} \{\mathit{Object}(x) * \mathit{Sword}(x) * \mathit{FriendlyObject}(x) \to_* 0\}$$

For all t-norms: a) when x is an element that is neither an object nor a sword, then r is $0 \to_* 0 = 1$, and b) when x is a sword, $r = \mathit{FriendlyObject}(x) \to_* 0$. In such case r depends on the t-norm. For instance, taking the Łukasiewicz's t-norm, $r = \min(1, 1 - \mathit{FriendlyObject}(x) + 0) = 0.75$. Taking either the minimum or the product t-norms $r = 0.25 \to_* 0 = 0$.

Concerning the entailment, we say that a fuzzy assertion α is *$*$-entailed* by the knowledge base $\mathcal{K}$ if every model $\mathbf{M}$ of $\mathcal{K}$ also satisfies α. For instance, the reader can easily prove that the fuzzy $ABox$ of the robots' example $*$-entails the assertion $\langle \mathsf{Friendly}(r_6), \succcurlyeq 0.50\rangle$ with respect to the fuzzy $TBox$ using either Łukasiewicz or minimum t-norms. However, using the product t-norm the above $*$-entailment is not satisfied.

4 A case study: Fuzzy description logics associated to $G_\sim\forall(S)$

The first FDL systems were related to the initial Zadeh proposal for fuzzy sets operations. The logic underlying this proposals is the logic associated to the calculus over $[0, 1]$ defined by the functions $\min$, $\max$, $n(x) = 1 - x$, the Kleene-Dienes implication $x \to y = \max(1 - x, y)$, and quantifiers interpreted as in section 2.2. This is referred as *minimalistic* apparatus of fuzzy logic by Hájek in [9]. But, as already noted by Hájek, the (non residuated) implication of this logic has some no nice behaviour. For example: (a) in this logic $\varphi \to \varphi$ is not a tautology and (b) an implication $\varphi \to \psi$ is evaluated as 1 only if either φ is evaluated as 0 or ψ is evaluated as 1. The languages $\mathcal{ALC}$ based on this logic was studied by Straccia in [14]. The notion of subsumption is defined using the notion of fuzzy subsets ($C \subseteq D$ if and only if $C(x) \leq D(x)$ for all x in the univers) with no relation to implication function due to the lack of a good relation between the implication function of this logic and the order relation in $[0, 1]$. Moreover it is not known any (Hilbert style) axiomatization of this logic.

For these reasons in this short case study we propose the interest of studying the $\mathcal{ALC}$ language associated to Gödel Logic with truth constants $G(\mathbf{S})\forall$ and specially the $\mathcal{ALC}$ language associated to its expansion with an involutive negation $G_\sim(\mathbf{S})\forall$. In [7] the authors prove the canonical completenes of the logic $G(\mathbf{S})\forall$. The canonical completeness can also be proved for $G_\sim(\mathbf{S})\forall$.

Theorem 4.1 *The logic $G_\sim(\mathbf{S})\forall$ has the Canonical Finite Strong completeness, i.e., for every finite set of formulas $\Gamma \cup \{\varphi\}$,*

$$\Gamma \vdash_{G_\sim(\mathbf{S})\forall} \varphi \quad \textit{iff} \quad \Gamma \models_{[0,1]_{G_\sim(\mathbf{S})}} \varphi$$

where
$[0, 1]_{G_\sim(\mathbf{S})} = \langle [0, 1], \min, \max, \to_G, n, \{r \mid r \in S\}, 0, 1\rangle$, *with $n(x) = 1 - x$ being the truth function associated to the involutive negation.*

The proof is a simplified version of [7, Theorem 11]. As a consequence of this theorem, we have an equivalence between $*$-entailment in $\mathcal{ALC}_{G_\sim(\mathbf{S})}$ and the semantics consequence relation on $[0, 1]_{G_\sim(\mathbf{S})}$.

Notice that in $G_\sim(\mathbf{S})\forall$ we have, in addition to the connectives of this logic, the ones being the counterpart of the minimalistic logic, since an implication having as truth function the Kleene-Dienes implication is definable as $\sim \varphi \vee \psi$. Moreover the truth constants allow us to have evaluated formulas and their negation by the involutive negation that makes possible to represent by formulas the semantical expressions saying that the interpretation of a formula is greater, greater or equal, less, and less or equal than a value r.

ISBN: 978-989-95079-6-8

Thus the description language based on $G_\sim\forall(S)$ seems to be a good choice because both it is actually very expressive and it maintains the canonical completeness. In this setting we have two implication connectives, that semantically correspond to the residuum of the minimum plus the Kleene-Dienes implication. Of course, as in the general case of $L^*_\sim(\mathbf{S})\forall$, the degree of subsumption between concepts is defined by means of the residuated implication, i.e, the Gödel implication. The Straccia's work [14] and the Bobillo's work [2] gives algorithms for proving satisfiability and subsumption in the $\mathcal{ALC}$ description language based on the minimalistic logic and in the $\mathcal{SROIQ}(D)$ description language based on Gödel logic respectively. As future work we want to find analogous algorithms for the $\mathcal{ALC}$ description languages based on $G_\sim(\mathbf{S})\forall$.

5 Conclusions

This paper is a first step on the direction proposed by Hájek concerning the relationships between some proposals of FDLs and the recent developments in mathematical fuzzy logics. The main contributions of our approach is the use of truth constants in the language of description and the introduction of an involutive negation in the required cases. This allows us to recover graded notions of satisfiability, validity and subsumption that have been used in the fuzzy logic setting. This choice is oriented to search for the syntactical counterpart of the semantic calculi proposed in some works dealing with FDLs.

Acknowledgments

This work has been supported by the Spanish projects MULOG2 TIN2007-68005-C04-01/04, "Agreement Technologies" (CONSOLIDER CSD 2007-0022, INGENIO 2010), and the MCYT-FEDER Project MID-CBR (TIN2006-15140-C03-01), the CSIC grant 200750I005, and the Generalitat de Catalunya grant 2005-SGr-00093. The authors thank Félix Bou for his assistance in the improvement of this paper.

References

[1] F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi and P. F. Patel-Schneider, editors. The Description Logic Handbook: Theory, Implementation and Applications. Cambridge University Press, 2007.

[2] F. Bobillo. Managing vagueness in Ontologies *PhD disertation, Univ. Granada* , 2008.

[3] R. Cignoli, F. Esteva, L. Godo, and A. Torrens. Basic fuzzy logic is the logic of continuous t-norms and their residua. *Soft Computing*, 4(2):106–112, 2000.

[4] F. Esteva, J. Gispert, L. Godo, and C. Noguera. Adding truth-constants to logics of continuous t-norms: Axiomatization and completeness results. *Fuzzy Sets and Systems*, 158(6):597–618, 2007.

[5] F. Esteva, L. Godo, P. Hájek, and M. Navara. Residuated fuzzy logics with an involutive negation. *Archive for Mathematical Logic*, 39(2):103–124, 2000.

[6] F. Esteva, L. Godo, and F. Montagna. Equational characterization of the subvarieties of BL generated by t-norm algebras. *Studia Logica* 76(2): 161–200, 2004.

[7] F. Esteva, L. Godo, and C. Noguera. On completeness results for predicate Łukasiewicz, product, Gödel and nilpotent minimum logics expanded with truth-constants. *Mathware & Soft Computing*, 14(3):223–246, 2007.

[8] P. Hájek. *Metamathematics of fuzzy logic*, volume 4 of *Trends in Logic—Studia Logica Library*. Kluwer Academic Publishers, Dordrecht, 1998.

[9] P. Hájek. Making fuzzy description logic more general. *Fuzzy Sets and Systems*, 154(1):1–15, 2005.

[10] P. Hájek. What does mathematical fuzzy logic offer to description logic? In Elie Sanchez, editor, *Fuzzy Logic and the Semantic Web*, Capturing Intelligence, chapter 5, pages 91–100. Elsevier, 2006.

[11] A. A. Monteiro. Sur les algèbres de Heyting symmétriques. *Port. Math.*, 39: 1–237, 1980.

[12] J. Pavelka. On fuzzy logic. I, II, III. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 25(1):45–52, 25(2):119–134, 25(5):447–464,1979.

[13] G. Stoilos, G. Stamou, J.Z. Pan, V. Tzouvaras, and I. Horrocks. Reasoning with very expressive fuzzy description logics. *Journal of Artificial Intelligence Research*, 30(8):273–320, 2007.

[14] U. Straccia. Reasoning within fuzzy description logics. *Journal of Artificial Intelligence Research*, 14:137–166, 2001.

[15] U. Straccia. Fuzzy alc with fuzzy concrete domains. In *Proceeedings of the International Workshop on Description Logics (DL-05)*, pages 96–103, Edinburgh, Scotland, 2005. CEUR.

[16] U. Straccia. A fuzzy description logic for the semantic web. In Elie Sanchez, editor, *Fuzzy Logic and the Semantic Web*, Capturing Intelligence, chapter 4, pages 73–90. Elsevier, 2006.

[17] U. Straccia and F. Bobillo. Mixed integer programming, general concept inclusions and fuzzy description logics. *Mathware & Soft Computing*, 14(3):247–259, 2007.

[18] C. B. Tresp and R. Molitor. A description logic for vague knowledge. Technical Report RWTH-LTCS Report 98-01, Aachen University of Technology, 1998.

[19] J. Yen. Generalizing term subsumption languages to fuzzy logic. In *Proc. of the 12th IJCAI*, pages 472–477, Sidney, Australia, 1991.

[20] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[21] L. A. Zadeh. Fuzzy logic and approximate reasoning. *Synthese*, 30:407–428, 1975.

ISBN: 978-989-95079-6-8

Evaluating premises, partial consequences and partial hypotheses *

I. García-Honrado, E. Trillas, S. Guadarrama and E. Renedo

European Centre for Soft Computing
Mieres (Asturias), Spain
Email: {itziar.garcia, enric.trillas, sergio.guadarrama, eloy.renedo}@softcomputing.es

Abstract— *To evaluate premises, consequences and hypotheses, on this paper relevance and support ratios are defined for each of them. This allows to distinguish consequences based on the number of premises that support them, and also to reduce the set of premises while maintaining the same consequences. Since the relation between premises and hypotheses is, in some sense, similar to the relation between consequences and premises, analogous ratios are defined for hypotheses and premises.*

Keywords— Conjectures, Consequences, Hypotheses, Relevance, Support.

1 Introduction

The aim of most problems is to make choice between possible solutions, a clear example is a medical diagnosis problem. In this paper we allocate degrees for the elements in the set of consequences, hypotheses or premises in order to choose the consequence, hypothesis or premise with the biggest degree. Papers [2] and [3] had dealt with that idea of graded consequences.

The following section will show Conjectures, Hypotheses, and Consequences (CHC) models introduced in [7], which was suggested, in part by Watanabe in [10], and takes into account the particular case of partial consequence's operator [8] [9]. Partial operators of consequences are that allow to get consequences of each premise, or subset of premises, and obtaining the final set of consequences as the union of all these partial consequences.

To define support for each consequence (section 3) we consider that consequences with bigger support, are those that are supported by more premises or subsets of premises. By the way, different degrees are allocated for consequences. So, for example, following with a medical diagnosis problem in which the premises are diseases and the consequences are symptoms, we can choose between consequences, and select as the stronger, the one with biggest support.

In section 4, we deal with a measure of relevance for premises that is useful for knowing which premises have more importance, in the sense of how many consequences can be deduced from them. Thanks to that measure, the set of premises can be reduced to a smaller set with the same relevance. This reduced set of premises gets rid of superfluous premises and yet allows to work with less premises, while getting the same set of consequences. Till now all premises seemed to have the same importance.

Finally, in section 5, we also consider partial hypotheses, that is, hypotheses of one premise, and not hypotheses of all premises. And analogous measure of support for premises, as well as of relevance for partial hypotheses, are defined. This allows to evaluate subset of partial hypotheses by counting how many premises they give as consequences.

2 Basic concepts

2.1 CHC models

Reasoning can be understood as a process allowing to get conjectures from a set of premises, P. There are three basic types of reasoning: deduction, abduction and induction. A process that allows to get consequences is a deductive reasoning, a process that allows to get hypotheses is an abductive reasoning, and finally, if the process allows to get speculations, it is an speculative reasoning.

In this paper, CHC models are defined on a preordered set $(L, \leq)$, endowed with a negation, $'$. And when it is necessary, the preordered set will be endowed with an infimum, $\cdot$, and a supremum, $+$, operations $(L, \leq, ', \cdot, +)$, a preorder with infimum and supremum operations is a partial ordered set (poset)with these operations. The infimum of L is called first element and it is denoted by 0, the supremum of L is called the last element and it is denoted by 1. The paper only deals with finite algebraic structures, that is, with a finite set L.

CHC models can be based on consequences operators [8, 9],

Definition 2.1 *If L is a set, and $\mathfrak{F} \subset \mathbb{P}(L)$, it is said that $(L, \mathfrak{F}, C)$ is a* structure of consequences*, provided that $C : \mathfrak{F} \to \mathfrak{F}$ verifies,*

1. *$P \subset C(P)$, for all $P \in \mathfrak{F}$ (C is* extensive*)*
2. *If $P \subset Q$, then $C(P) \subset C(Q)$, for all $P, Q \in \mathfrak{F}$ (C is* monotonic*)*
3. *$C(C(P)) = C(P)$, or $C^2 = C$, for all $P \in \mathfrak{F}$ (C is a* clausure*)*

i.e. C is an operator of consequences *(in the sense of Tarski) for $\mathfrak{F}$ in L.*

For each $\{q\} \in \mathfrak{F}$, let us write $C(q) = C(\{q\})$.

*This work has been partially supported by the Foundation for the Advancement of Soft Computing (Asturias, Spain), and CICYT (Spain) under project TIN2008-06890-C02-01

ISBN: 978-989-95079-6-8

Definition 2.2 *A consequences' operator C is consistent in P, if for all $q \in C(P)$, $q' \notin C(P)$.*
A structure of consequences $(L, \mathfrak{F}, C)$ is consistent if C is consistent for all $P \in \mathfrak{F}$.

Let P be the set of premises, $P \neq \emptyset$, and $C(P)$ a set of consequences for P. Conjectures of P can be defined from a consistent consequences operator C, as those elements whose negation is not in $C(P)$, $Conj_C(P) = \{q \in L; q' \notin C(P)\}$. Hypothesis can be defined by $Hyp_C(P) = \{h \in Conj_C(P) - C(P); \{h\} \in \mathfrak{F},\ P \subset C(h)\}$.

Finally, speculations are those conjectures that are neither consequences, nor hypotheses. Hence, $Sp_C(P) = Conj_C(P) - [C(P) \cup Hyp_C(P)] = \{q \in L, q \notin C(P), q' \notin C(P), q \notin Hyp_C(P)\}$.

2.2

Definition 2.3 *A consequences' operation C is a partial consequences operator if $C(P) = \bigcup_{R \subset P, R \in \mathfrak{F}} C(R)$.*

Definition 2.4 *A decomposable consequences operator is a consequences' operator such that $C(P) = \bigcup_{p \in P} C(R)$.*

This paper considers *partial hypothesis*, elements that are hypotheses of a subset of the set of premises P. This idea comes from that of *partial consequences*.

Definition 2.5 *For each set P of premises the partial hypotheses set is,*

$$Hyp^*_C(P) = \{h \in \{L - 0\} - P; \{h\} \in \mathfrak{F}, \exists R \subset P,$$
$$R \subset C(h)\}.$$

Obviously, hypotheses are partial hypotheses, since $P \subset P$ and $P \subset C(h)$, provided h is a hypothesis.

Remark 2.6 *Although hypotheses are anti-monotonic ($P_1 \subset P_2$, implies $Hyp(P_2) \subset Hyp(P_1)$), partial hypotheses are (as it is easy to prove) monotonic ($P_1 \subset P_2$, implies $Hyp^*(P_1) \subset Hyp^*(P_2)$). That is why they can not be considered classical hypotheses.*

2.3

The paper deals with a general concept of measure [6], defined in a preordered set $(L, \leq)$. A measure is a mapping $m : L \to [0, 1]$, such that:

- There exists a minimal $x_0 \in L$, such that $m(x_0) = 0$
- There exists a maximal $x_1 \in L$, such that $m(x_1) = 1$
- If $x \leq y$, then $m(x) \leq m(y)$.

3 Consequences support

This section introduces a ratio in order to distinguish which consequences are the more supported by a given set of premises. And proof in which cases is a measure

Let's recall that in this paper L is assumed to be a finite set.

Definition 3.1 *The support of $q \in L$ is the ratio of subsets of premises that allow getting q as a consequence, to all possible subsets of premises.*

$$Supp_{C,P}(q) = \frac{|\{R \in \mathbb{P}(P); q \in C(R)\}|}{2^{|P|} - 1} = \quad (1)$$
$$\frac{|\{R \in \mathbb{P}(P); q \in C(R)\}|}{|\mathbb{P}(P) - \emptyset|}.$$

Since $P \neq \emptyset$, it is $|P| > 0$ and the quotient in the definition is possible.

The bigger support a consequence has, the more subsets of premises allow deducing it.
Notice that if $q \notin C(P), Supp_{C,P}(q) = 0$, since if there were $R \in \mathbb{P}(P)$ such that $q \in C(R)$, because of the monotonicity of the consequence operator, $C(R) \subset C(P)$ would imply $q \in C(P)$.

This ratio verifies the following properties,

- If $P \subset Q$, it is $Supp_{C,P}(q) \leq Supp_{C,Q}(q)$, for all $q \in L$.
- If $P = \{p\}, \forall q \in C(P)$, it is $Supp_{C,P}(q) = 1$.
- For all $q \in C(P)$, $Supp_{C,P}(q) > 0$.
- $Supp_{C,P}(q) = 1$ means that q is a consequence for all $R \in \mathbb{P}(P)$. Particularly, q is consequence of all $p \in P$.

The support defined by (1), is not a measure in general. For example, if $C(P) = P,\ \forall P \in \mathbb{P}(L)$, let P be a set with more than one element. If $q \in L$, it is either $q \in P$, or $Supp(q) = 0$. If $q \in P$, there exists $p \in P$ such that $p \neq q$, and $q \notin C(p)$ and $Supp(q) \neq 1$. Therefore, there is no $q \in L$ such that $Supp_{C,P}(q) = 1$.

Remark 3.2 *$Supp_{C,P}$ is monotonic with respect to the preorder given by C, $q_1 \leq_C q_2$ iff $q_2 \in C(q_1)$ [1],*

Proof. Since if $q_1 \leq_C q_2$, for each R such that $q_1 \in C(R)$, it is $C(q_1) \in C(C(R)) = C(R)$, and, as $q_2 \in C(q_1)$, it is also $q_2 \in C(R)$. So, $Supp_{C,P}(q_1) \leq Supp_{C,P}(q_2)$.
□

Since, given P, the relation defined between the pairs of elements in L with the same value of $Supp_{C,P}$, is an equivalence, the classes

$$[q] = \{v \in L;\ Supp_{C,P}(v) = Supp_{C,P}(q)\}$$

give a partition on L in a number of parts that is, at most, $2^{|P|}$.

3.1 The case of the operator $C_\bullet$

Let $(L, \leq)$ be now a preordered set in which is defined an infimum operation denoted by '$\cdot$'.

The partial consequences operator $C_\bullet$ gives consequences that are consequences for a subset of the set of premises P, it is $C_\bullet(P) = \{q \in L; \exists \{p_1, p_2, ..., p_n\} \subset P : p_1 \cdot p_2 \cdot ... \cdot p_n \leq q\}$. It is a partial consequences' operator and it obviously verifies $C_\bullet(P) = \bigcup_{R \subset P, R \in \mathfrak{F}} C_\bullet(R)$.

ISBN: 978-989-95079-6-8

Notice, that as L is finite, P is also finite and $C_\bullet(P) = \{q \in L; \inf P \leq q\}$, which equal to the infimum operator of consequences $C_\wedge$.

$$Supp_{C_\bullet,P}(q) = \frac{|\{R \in \mathbb{P}(P) \in P; \inf(R) \leq q\}|}{2^{|P|}-1} \quad (2)$$

Let's see what specific properties are verified by $Supp_{C_\bullet,P}$,

- If L has last element, it implies $1 \in L$, then $1 \in C_\bullet(P)$ and $Supp_{C_\bullet,P}(1) = 1$.
- If $q_1 \leq q_2$, then $Supp_{C_\bullet,P}(q_1) \leq Supp_{C_\bullet,P}(q_2)$. That is the function $Supp_{C_\bullet,P}$ is monotonic.
- $Supp_{C_\bullet,P}(\sup\{q_1,q_2\}) \geq \max\{Supp_{C_\bullet,P}(q_1), Supp_{C_\bullet,P}(q_2)\}$, provided $\sup\{q_1,q_2\}$ exists.
- $Supp_{C_\bullet,P}(\inf\{q_1,q_2\}) \leq \min\{Supp_{C_\bullet,P}(q_1), Supp_{C_\bullet,P}(q_2)\}$

Corollary 3.3 *Let $(L, \leq, ', \cdot, +)$ be a partial ordered set with infimum and supremum operations and first and last elements. If $P \neq \{0\}$, the function $Supp_{C_\bullet,P} : L \to [0,1]$ is a measure.*

Proof. It is monotonic, and it verifies the boundary conditions, since 0 is not a consequence $Supp_{C_\bullet,P}(0) = 0$ and $Supp_{C_\bullet,P}(1) = 1$. □

3.2 The case of the operator $C_\leq$

$C_\leq$ is the partial consequences operator that gives as consequences those elements that follow from at least one premise in P, formally, it is $C_\leq(P) = \{q \in L; \exists p \in P : p \leq q\}$, see [8]. Hence, it can be considered as a decomposable consequences' operator, since allows getting consequences that are not deduced from all premises. It is straightforward that $C_\leq(P) = \bigcup_{R \subset P, R \in \mathfrak{F}} C(R) = \bigcup_{p \in P} C_\leq(p)$.

In this case, a different definition of the support's ratio seems to be more convenient, since it deals only with consequences of each $p \in P$, nor with consequences of each subset of $\mathbb{P}(P)$.

Definition 3.4 *The support of $q \in C_\leq(P)$ is the ratio of premises that allow getting q as consequence to all premises.*

$$\widehat{Supp}_{C_\leq,P}(q) = \frac{|\{p \in P; p \leq q\}|}{|P|} \quad (3)$$

Since $P \neq \emptyset$, it is $|P| > 0$ and the quotient in the definition is possible.

If $q \notin C_\leq$, then $\widehat{Supp}_{C_\leq,P}(q) = 0$.

So, the bigger Support an element has, the more premises allow to reach it.

$\widehat{Supp}_{C_\leq,P}$ verifies,

- If $P = \{p\}, \forall q \in C_\leq(p)$, it is $\widehat{Supp}_{C_\leq,P}(q) = 1$.
- If L has last element, 1, then $1 \in C_\leq(P)$ and $\widehat{Supp}_{C_\leq,P}(1) = 1$.
- For no $q \in C_\leq(P)$ is $\widehat{Supp}_P(q) = 0$. That is, for all $q \in C_\leq(P)$, $\widehat{Supp}_{C_\leq,P}(q) > 0$.
- $\widehat{Supp}_{C_\leq,P}(q) = 1$ means that q is a consequence for all $p \in P$.
- If $q_1 \leq q_2$, then $\widehat{Supp}_{C_\leq,P}(q_1) \leq \widehat{Supp}_{C_\leq,P}(q_2)$. That is the function $\widehat{Supp}_{C_\leq,P}$ is monotonic with respect to $\leq$.

Remark 3.5 *In order to know what happens if we calculate the support for the infimum or supremum, of two consequences, provided it exists and it is a consequence, weak boundaries are found,*

- $\widehat{Supp}_{C_\leq,P}(\sup\{q_1,q_2\}) \geq \max\{\widehat{Supp}_{C_\leq,P}(q_1), \widehat{Supp}_{C_\leq,P}(q_2)\}$
- $\widehat{Supp}_{C_\leq,P}(\inf\{q_1,q_2\}) \leq \min\{\widehat{Supp}_{C_\leq,P}(q_1), \widehat{Supp}_{C_\leq,P}(q_2)\}$

Obviously, if the operator is consistent, that is, if $q \in C_\leq(P)$, then $q' \notin C_\leq(P)$, it follows $\widehat{Supp}_{C_\leq,P}(q') = 0$.

Theorem 3.6 *Let $(L, \leq, ', \cdot, +)$ be a partial ordered set with first and last elements and $P \neq \{0\}$. Function $\widehat{Supp}_{C_\leq,P} : L \to [0,1]$ is a measure.*

Proof. It is monotonic as it is stated above. Since 0 is not a consequence $\widehat{Supp}_{C_\leq,P}(0) = 0$. Finally, it is obvious that $1 \in C_\leq(P)$ and $\widehat{Supp}_{C_\leq,P}(1) = 1$. □

From $\widehat{Supp}_{C_\leq,P}$ we can calculate $Supp_{C_\leq,P}$. If $\widehat{Supp}_{C_\leq,P}(q) = k$ and $|P| = n$, it is $|\{R \in \mathbb{P}(P); q \in C(R)\}| = k \cdot n$. Hence,

$$Supp_{C_\leq,P} = \frac{2^n - \sum_{i \in \{0,\ldots,n-k\cdot n\}} \frac{(n-k\cdot n)!}{i!(n-k\cdot n-i)!}}{2^n - 1}. \quad (4)$$

The numerator in (4) is the number of subsets of premises that contain at least one of the premises supporting q.

Corollary 3.7 *Let $(L, \leq)$ be a preordered set with first and last elements and $P \neq \{0\}$. Function $Supp_{C_\leq,P} : L \to [0,1]$ is a measure.*

Proof. It is monotonic as it is proven at the beginning of this section, and it verifies boundary conditions, since 0 is not a consequence $Supp_{C_\leq,P}(0) = \frac{2^n-2^n}{2^n-1} = 0$ and $Supp_{C_\leq,P}(1) = \frac{2^n-1}{2^n-1} = 1$. □

ISBN: 978-989-95079-6-8

Example 3.8 *Figure 1 represents a preordered set of medical symptoms and deseases for patients. Let's calculate the support for the consequences for a patient with* $P = \{antibody, bacterium\}$. *To such an end, let us notice that,* $C_{\leq}(P) = \{antibody, bacterium, fever, eruption, 1\}$. *Then,*

- $\widehat{Supp}_{C_{\leq},P}(antibody) = \frac{1}{2}$, *and* $Supp_{C_{\leq},P}(antibody) = \frac{2^2-(1+1)}{2^2-1} = \frac{2}{3}$.
- $\widehat{Supp}_{C_{\leq},P}(bacterium) = \frac{1}{2}$, *and* $Supp_{C_{\leq},P}(bacterium) = \frac{2}{3}$.
- $\widehat{Supp}_{C_{\leq},P}(fever) = 1$, *and* $Supp_{C_{\leq},P}(fever) = \frac{3}{3} = 1$.
- $\widehat{Supp}_{C_{\leq},P}(eruption) = \frac{1}{2}$, *and* $Supp_{C_{\leq},P}(eruption) = \frac{2}{3}$.
- $\widehat{Supp}_{C_{\leq},P}(1) = Supp_{C_{\leq},P}(1) = 1$.

Hence, the consequence with greatest support is 'fever'.

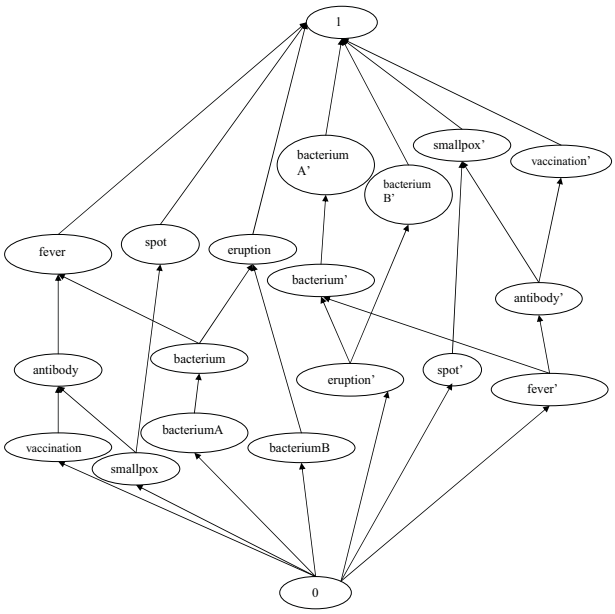


Figure 1: Preorder

4 Relevance for premises

This section introduces a measure to calculate the proportion of consequences that are gotten from a subset of premises, and what is more, it is shown how to reduce the set of premises using this ratio in order to give the same set of consequences.

Definition 4.1 *The relevance of a subset of premises* $R \in \mathbb{P}(P) - \{\emptyset\}$ *is the ratio of consequences deduced from a* R*, to all consequences.*

$$Rel_{C,P}(R) = \frac{|\{q \in L; q \in C(R)\}|}{|C(P)|} = \frac{|C(R)|}{|C(P)|}, \quad \text{if } R \in \mathbb{P}(P) - \{\emptyset\}, \tag{5}$$

and, $Rel_{C,P}(\emptyset) = 0$.

Since $|P| > 0$ *and* $P \subset C(P)$*, it is* $|C(P)| > 0$*, the quotient does exist.*

If a subset of premises allows to deduce all consequences, the set of premises can be reduced to it, since both have the same set of consequences.

There are many properties that $Rel_{C,P}$ verifies,

- If there exists $R \in \mathbb{P}(P)$, such that $Rel_{C,P}(R) = 1$, it means that all consequences for P are consequences of R. So, $C(P) = C(R)$.
- If $R_1 \subset R_2$, then $Rel_{C,P}(R_1) \leq Rel_{C,P}(R_2)$. That is function $Rel_{C,P}$ is monotonic.
- It is $Rel_{C,P}(P) = 1$, and $Rel_{C,P}(\emptyset) = 0$.

Theorem 4.2 *Function* $Rel_{C,P} : \mathbb{P}(P) \to [0,1]$ *is a measure.*

Proof. Straightforward, by the last properties. □

Remark 4.3 *In this case, the concept of fuzzy measure is defined in the preordered set* $(\mathbb{P}(L), \subset)$*, since relevance is defined for all subsets of premises and not only for single premises. Remember that the support is defined for each element.*

The ratio of relevance applying for each premise allows to define a partition into the set of premises, in classes whose elements have the same relevance, $[q] = \{p \in L;\ Supp_{C,P}(p) = Supp_{C,P}(q)\}$. Analogously, it can be built a partition in the set $\mathbb{P}(P)$, defining each class as $[S] = \{R \in \mathbb{P}(P);\ Supp_{C,P}(R) = Supp_{C,P}(S)\}$. The maximum number of classes that can exist is $|C(P)| + 1$.

4.1 Using the operator $C_{\leq}$

For the operator $C_{\leq}$ it is useful to calculate the relevance for each $p \in P$, since it is sufficient to get consequences for each premise and then join them. So, in this case is enough to deals with

$$Rel_{C_{\leq},P} : P \to [0,1]$$

Definition 4.4 *The relevance for a premise* $p \in P$ *is the proportion of consequences deduced from* p.

$$Rel_{C_{\leq},P}(p) = \frac{|\{q \in L; p \leq q\}|}{|C_{\leq}(P)|} = \frac{|C(p)|}{|C(P)|}. \tag{6}$$

If a premise allows to deduce all consequences, the set of premises can be reduced to that premise, since both give the same set of consequences.

There are many properties that $Rel_{C_{\leq},P}(p)$ verifies,

- For all $p \in P$, $Rel_{C_{\leq},P}(p) > 0$, since $p \in \{q \in L; p \leq q\}$ implies $|\{q \in L; p \leq q\}| > 0$.
- If there exists $p \in P$ such that $Rel_{C_{\leq},P}(p) = 1$, it means that all consequence for P is consequence of p. So, $C_{\leq}(P) = C_{\leq}(p)$.

ISBN: 978-989-95079-6-8

- If $p_1 \leq p_2$, then $Rel\,_{C_\leq,P}(p_2) \leq Rel\,_{C_\leq,P}(p_1)$. That is the function $Rel\,_{C_\leq,P}$ is anti-monotonic in this sense. Then, the function $1 - Rel\,_{C_\leq}$ is monotonic.
- Let L be endowed with an infimum operation. If $\inf P \in P$, as $Rel\,_{C_\leq,P}(\inf P) = 1$, because $\forall q \in C_\leq(P)$ there exists $p \in P$ such that $p \leq q$, and $\inf P \leq p \leq q$. Then, $C_\leq(P) = C_\leq(\inf P)$.
  A common consequence's operator is $C_\wedge$, defined by $C_\wedge(P) = \{q \in L; \inf P \leq q\}$, that can be defined as $C_\wedge(P) = C_\leq(\inf P)$, so in that case $C_\leq(P) = C_\wedge(P)$.

Example 4.5 *Using the same preset in figure 1. Let's calculate the relevance for premises. Here, we have an example that allows us to quantify the relevance of deseases of a patient.*

If the patient has $P = \{antibody, smallpox\}$*, then,* $C_\leq(P) = \{antibody, smallpox, spot, fever, 1\}$*. Hence,*

- $Rel\,_{C_\leq,P}(antibody) = \frac{|\{antibody,fever,1\}|}{|C_\leq(P)|} = \frac{3}{5}$.
- $Rel\,_{C_\leq,P}(smallpox) = \frac{|\{antibody,smallpox,spot,fever,1\}|}{|C_\leq(P)|} = \frac{5}{5} = 1$
- *Obviously,* $Rel\,_{C_\leq,P}(P) = 1$.

This example shows the case that a premise allows to deduce all consequences of P, since $C_\leq(P) = C_\leq(smallpox)$.

Theorem 4.6 *If* $(L, \leq, ', \cdot, +)$ *is a partial ordered set and* $P \subset L$ *such that* $\inf P \neq 0$.
There exists $p \in P$ *such that* $Rel\,_{C_\leq,P}(p) = 1$*, if and only if* $p = \inf P$.

Proof. If $Rel\,_{C_\leq}(p) = 1$, it is $p \leq q$ for all $q \in C_\leq(P)$. And since $C_\leq(P) \subset C_\wedge(P)$, it is $\inf P \leq q\ \forall q \in C_\leq(P)$. The infimum is the greatest lower bound of a set, then $p \leq \inf P$. It is also $\inf P \leq p$. Thus, $p = \inf P$, because L is a lattice, so verifies antisymmetric property and has an infimum for each subset.
On the other hand if $p = \inf P$, implies $\inf P \in P$, and as it is said, $Rel\,_{C_\leq,P}(\inf P) = 1$. □

In the theorem and in the above example, it is shown that the set of premises can be reduced to an only premise with relevance one, but if there is no one premise with relevance one, it could be found a subset of premises that allows to obtain the same consequences as the initial set of premises. When models deal with not a very big number of premises, a simple program can be used in order to find a minimal set of premises by calculating all combination of premises.

This algorithm is exponential in the number of premises. So, others algorithms can be designed in order to deal with a big number of premises.

,

The algorithm is as follows.

First of all, we look for premises with greatest relevance, we put one of these premises (p_1) into the set of reduced premises, then we calculate a relative relevance

$$Rel\,_{C_\leq,P-\{p_1\}}(p) = \frac{|\{q \in C_\leq(P) - C_\leq(p_1); p \leq q\}|}{|C_\leq(P) - C_\leq(p_1)|},$$

and we introduce a premise with the greatest relative relevance (p_2), and then we calculate other relative relevance,

$$Rel\,_{C_\leq,P-\{p_1,p_2\}}(p) =$$

$$\frac{|\{q \in (C_\leq(P) - C_\leq(p_1)) - C_\leq(p_2); p \leq q\}|}{|(C_\leq(P) - C_\leq(p_1)) - C_\leq(p_2)|},$$

and this process is repeated till the lowest r that verifies $C_\leq(P) = \bigcup_{i\in\{1,...,r\}} C_\leq(p_i)$. Then the reduced set searched in this way will be $\{p_1, ..., p_r\}$.

4.2 Using the operator $C_\bullet$

In this case we can particularize the definition of relevance for each subset of premises.

The relevance for a subsets of premises $R \subset \mathbb{P}(P) - \{\emptyset\}$ is the ratio of consequences deduced from R, to consequences deduced from P.

$$Rel\,_{C_\bullet,P}(R) = \frac{|\{q \in L; q \in C_\bullet(R)\}|}{|C_\bullet(P)|} =$$

$$\frac{|\{q \in L; \exists \tilde{R} \subset \mathbb{P}(R), \inf \tilde{R} \leq q\}|}{|C_\bullet(P)|}$$

Example 4.7 *Let* P *be* $\{c, d, a'\}$ *defined in the preorder in figure 2. So,* $C_\bullet(P) = \{c, d, f, g, b, e, d', a', 1\}$*. Relevance for all subset of premises are,*

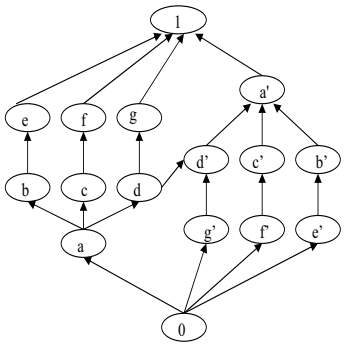


Figure 2: Preorder

- $Rel\,_{C_\bullet,P}(\{c\}) = \frac{3}{7}$, $Rel\,_{C_\bullet,P}(\{d\}) = \frac{5}{7}$, $Rel\,_{C_\bullet,P}(\{a'\}) = \frac{2}{7}$
- $Rel\,_{C_\bullet,P}(\{c, d\}) = 1$, $Rel\,_{C_\bullet,P}(\{c, a'\}) = 1$, $Rel\,_{C_\bullet,P}(\{d, a'\}) = \frac{5}{7}$

This example gives two reduced sets of premises, $\{c, d\}$ *and* $\{c, a'\}$*. Obviously* $C_\bullet(P) = C_\bullet(\{c, d\}) = C_\bullet(\{c, a'\})$.

5 Validating premises and partial hypotheses

Each premise is supported, or explained by hypotheses, so in this section, a support for each premise is defined. Then, we define the relevance of each partial hypothesis. This is analogously to the above section, because premises are consequences of hypotheses.

Let $Hyp^*{}_C(P) \neq \emptyset$.

ISBN: 978-989-95079-6-8

Definition 5.1 *The support of $p \in P$ is the ratio of subsets of hypotheses that allow getting p as consequence, to all subsets of partial hypotheses.*

$$Supp_{C,Hyp^*_C(P)}(p) = \frac{|\{H \subset Hyp^*_C(P); p \in C(H)\}|}{2^{|Hyp^*_C(P)|} - 1} \quad (7)$$

Since $|Hyp^*_C(P)| > 0$ the quotient in the definition is possible.

The bigger Support a premise has, the more hypotheses allow to deduce it.

The $Supp_{C,Hyp^*_C}$ verifies,

- If $P = \{p\}$, it is $Supp_{C,Hyp^*_C(P)}(p) = 1$.
- If $1 \in P$, then $Supp_{C,Hyp^*_C(P)}(1) = 1$.
- $Supp_{C,Hyp^*_C(P)}(p) = 1$ means that p is explained by all $h \in Hyp^*_C(P)$, in particular for all $h \in Hyp(P)$.
- If $p_1 \leq p_2$, then $Supp_{C,Hyp^*_C(P)}(p_1) \leq Supp_{C,Hyp^*_C(P)}(p_2)$. That is, function Sup is monotonic.
- $Supp_{C,Hyp^*_C(P)}(\sup\{p_1, p_2\}) \geq \max\{Supp_{C,Hyp^*_C(P)}(p_1), Supp_{C,Hyp^*_C(P)}(p_2)\}$.
- $Supp_{C,Hyp^*_C(P)}(\inf\{p_1, p_2\}) \leq \min\{Supp_{C,Hyp^*_C(P)}(p_1), Supp_{C,Hyp^*_C(P)}(p_2)\}$.

$Supp_{C,Hyp^*_C(P)}$, allows to compare premises in relation to hypotheses and to allocate different degrees to each premise.

Definition 5.2 *The relevance for a subset of partial hypotheses $H \subset Hyp^*_C(P)$ is the proportion of premises deduced from H.*

$$Rel_{C,Hyp^*_C(P)}(H) = \frac{|\{p \in P; p \in C(H)\}|}{|P|}, \quad (8)$$

if $H \in \mathbb{P}(Hyp^*_C(P)) - \{\emptyset\}$,

and, $Rel_{C,Hyp^*_C(P)}(\emptyset) = 0$.

Since $|P| > 0$, the quotient is possible.

There are many properties that $Rel_{C,Hyp^*_C(P)}$ verifies,

- If $P = \{p\}$, it is $Rel_{C,Hyp^*_C(P)}(H) = 1$, for all $H \subset Hyp^*_C(P)$.
- For all $H \subset Hyp^*_C(P)$, $Rel_{C,Hyp^*_C(P)}(H) > 0$.
- If $H_1 \subset H_2$, then $Rel_{C,Hyp^*_C(P)}(H_1) \leq Rel_{C,Hyp^*_C(P)}(H_2)$. That is, function $Rel_{C,Hyp^*_C(P)}$ is monotonic.
- If there exists $h \in Hyp^*_C(P)$ such that $Rel_{C,Hyp^*_C(P)}(\{h\}) = 1$, it means that h is a partial hypothesis that can explain all premises, so it is hypothesis.

Theorem 5.3 *If $Hyp_C(P) \neq \emptyset$, then $Rel_{C,Hyp^*_C(P)}$ is a measure.*

Proof. It is monotonic as it is stated above, $Rel_{C,Hyp^*_C(P)}(\emptyset) = 0$, and since there exists $h \in Hyp_C(P)$ $Rel_{C,Hyp^*_C(P)}(\{h\}) = 1$.
□

The measure, $Rel_{C,Hyp^*_C(P)}$, allows to compare partial hypotheses, and to distinguish which partial hypotheses are hypotheses in the classical sense, that are those one with relevance one.

6 Conclusions

In this paper, it is built a measure in the set of consequences, premises and partial hypotheses. That can be useful in decision problems in order to choose the consequence, premise or hypotheses with the biggest measure.

It should be also pointed out that by using a relevance measure, we can get rid of premises that do not add information, and still get the same set of consequences.

It is also introduced the concept of set of partial hypotheses, that contains the classical hypotheses one. The measure built allocate value one to that partial hypotheses that are hypotheses in the classical sense.

As future work it can be proposed to apply these measures to practical problems, for examples medical diagnosis problems much more bigger than the one that appears in this paper.

7 Acknowledgements

Authors thanks to the three anonymous reviewers for their hints and comments.

References

[1] J. L. Castro and E. Trillas. Sobre preordenes y operadores de consecuencias de Tarski. *Theoria*, 4(11):419–425, 1989. (In Spanish).

[2] M.K. Chakraborty. Use of fuzzy set theory in introducting graded consequence in multiple-valued logic. *Fuzzy Logic in Knowledge-Based Systems, Decision and Control*, (1):247–257, 1988.

[3] M.K. Chakraborty and S. Dutta. *Theory of graded consequence and fuzzy logics*. LATD, Italy, 2008.

[4] D. Qiu. A note on Trillas' CHC models. *Artificial Intelligence*, 171:239–254, 2007.

[5] A. R. de Soto, A. Álvarez, and E. Trillas. Short note: Counting conjectures. *Mathware and Soft Computing*, 14(2):165–170, 2007.

[6] E. Trillas and C. Alsina. A reflection on what is a membership function. *Mathware & Soft-Computing*, 6:201–215, 1999.

[7] E. Trillas, S. Cubillo, and E. Castiñeira. On conjectures in orthocomplemented lattices. *Artificial Intelligence*, 117:255275, 2000.

[8] E. Trillas, I. García-Honrado, and A. Pradera. Consequences and conjectures in peordered sets. *ECSC's Tecnical Report (available upon request)*.

[9] E. Trillas, M. Mas, and M. Monserrat. Conjecturing from consequences. *Forthcoming in Int. journal of General Systems*.

[10] Satosi Watanabe. *Knowing and guessing. A Quantitative Study of Inference and Information*. John Wiley and sons, New York, 1969.

ISBN: 978-989-95079-6-8

Flexible Retrieval of X-Ray Images Based on Shape Descriptors Using a Fuzzy Object-Relational Database

Juan Miguel Medina[1] Sergio Jaime-Castillo[1] Carlos D. Barranco[2] Jesús R. Campaña[1]

1.Department of Computer Science and Artificial Intelligence. University of Granada
C/ Periodista Daniel Saucedo Aranda s/n, 18071 Granada, Spain.
2.Division of Computer Science, School of Engineering. Pablo de Olavide University
Utrera Rd. Km. 1, 41013 Seville, Spain.
Email: {medina,sjaime}@decsai.ugr.es, cbarranco@upo.es, jesuscg@decsai.ugr.es

Abstract— *This paper presents a novel approach for medical image storage using a Fuzzy Object-Relational Database Management System (FORDBMS). The system stores medical images along with a set of parameters describing their content. Flexible queries can be performed over these parameters to retrieve images matching visually. To illustrate the capabilities of the FORDBMS, parameter curves are obtained from X-Ray images of patients suffering from scoliosis, and queries are performed when looking for images with a determined curve pattern. Results show that retrieved images visually match the condition established in the query.*

Keywords— Image Retrieval, Fuzzy Databases

1 Introduction

Images are a fundamental tool in health care for diagnosis, clinical studies, research and learning. Currently, there are multiple techniques to capture images from patients to help diagnostic tasks such as X-Ray images, Computer Tomography (CT), Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), Ultrasonography, etc. The diagnostic task generates a large amount of images which must be archived for future evaluations. Fortunately, most of these techniques produce digital images, which are more efficiently archived and handled through computer systems than physical ones. In medical imaging, these computer systems are called Picture Archiving and Communication Systems (PACS). PACS are computers or networks devoted to the storage, retrieval, distribution and presentation of images. These systems solve the problem of storaging digital images but do not provide mechanisms to retrieve them based on their content.

Content-Based image retrieval (CBIR) [1] is the application of computer vision techniques to the problem of digital image search in large databases. First image retrieval approaches were based on captions and textual descriptors collected by humans. Nowadays, image retrieval systems improve the textual-based ones using features such as color, texture or shape, which are automatically extracted from images [2]. In this regard, a very important point to take into account is the imprecision in the feature descriptions, as well as the storage and retrieval of that imprecise data. To deal with this vagueness, some interesting approaches introduce the use of fuzzy logic in the feature representation, as well as in the retrieval process [3, 4, 5]. These fuzzy approaches also allow to perform queries on the basis of linguistic terms, avoiding one of the drawbacks of the classical image retrieval systems, where the queries have to be defined on the basis of images or sketches similar to the one we are searching for.

Database Management Systems (DBMS) are crucial in CBIR for retrieval purposes and, as we have explained above, these systems must be able to represent fuzzy data as complex datatype structures to provide flexible content based retrieval. There is a wide variety of proposals for fuzzy data handling in databases [6, 7, 8, 9, 10, 11] but in general these models and/or implementations do not have enough modeling power and performance for image indexing applications. For these kind of applications we propose to use the Fuzzy Object-Relational Database System (FORDBMS) model introduced in [12, 13]. This model evolves classical fuzzy databases models to incorporate object-oriented features for a powerful representation and handling of data, fuzzy or not.

CBIR techniques get more semantical results when applied to a specific domain or application area, as knowledge on the domain and on the image characteristics helps the process to extract the relevant features for this specific area of application. Health care is an application area that may benefit from the CBIR techniques. If we focus the CBIR techniques on the analysis of a certain pathology we can get high level features processing certain types of images. For example, there are some proposals to extract vertebral and spinal shapes from X-rays [14, 15] (these algorithms need some user intervention). The paper [16] describes a technique to automatically measure Cobb angle [17] for scoliosis pathologies given the end vertebras of the curve.

This paper shows how our FORDBMS is suitable for easy image representation and retrieval, using the fuzzy descriptors obtained by means of computer vision algorithms [18] or provided by experts. To illustrate this, we show the process to represent a structure in the FORDBMS describing scoliosis measures (Cobb angles) obtained from anteroposterior X-rays. Aditionally, we will show that the use of fuzzy comparators implemented by the FORDBMS in queries let us retrieve images that match the expected visual characteristics searched.

The rest of the paper is organized as follows. Section 2 describes the scoliosis pathology and the use of X-rays for its diagnosis. Section 3 introduces the fuzzy object-relational database system used to store and retrieve the fuzzy data. Examples of queries are presented in Section 4 and, finally, the main conclusions and future works are summarized in Section 5.

ISBN: 978-989-95079-6-8

2 Idiopathic scoliosis and its X-ray based diagnosis

In order to show how the use of fuzzy databases can help in health care, we have focused on the study of representation and retrieval of images related with idiopathic scoliosis. In this section we will describe the most relevant characteristics of this pathology in relation with our purpose.

Scoliosis is a three-dimensional deformation of the spine that produces vertebral rotation and crushing, and lateral curvature. It is typically classified as congenital (caused by vertebral anomalies present at birth), idiopathic (sub-classified as infantile, juvenile, adolescent, or adult according to when onset occurred) or as having developed as a secondary symptom of another condition, such as cerebral palsy, spinal muscular atrophy or due to physical trauma. Depending on the severity and progression of the deformation may be necessary treatment, consisting of observation, orthotic (brace) treatment, or surgery. About 2-4% of the adolescent population has some degree of scoliosis. Approximately 2.2% of these adolescents will require treatment.

To diagnose and treat scoliosis it is necessary to perform measures of the spine deformity. There are physical examinations to initially detect the presence of the deformity, but a precise diagnosis and treatment needs the help of radiologic techniques. The most accurate technique to measure spinal deformity is Computed Tomography (CT), that provides a three-dimensional view of the spine. However this technique is expensive and exposes the patient to a high radiation. Taking into account that a patient having scoliosis may need observation and treatment for many years and many radiologic tests, the frequent use of this technique may be inappropriate. X-Rays expose the patient to a lower radiation and, because of this, full-length standing spine X-rays are the standard method for evaluating the severity and progression of the scoliosis. Anteroposterior X-rays (AP X-rays) project spinal deformities as curves. The standard method to quantitatively asses curvatures is the measurement of the Cobb angle on AP X-rays. The Cobb angle [17] can be manually measured by calculating the angle between the lines respectively drawn along the upper endplate of the superior end-vertebra and the lower endplate of the inferior end-vertebra, as shown in Fig. 1. Using this measure each curve present in the spine is characterized by means of four parameters: the side of the convexity of the curve (right or left), the superior end-vertebra, the inferior end-vertebra and the angle value. It is important also to identify the apical vertebra associated with the adjacent disc interspaces that have the greatest segmental angulation of all interspaces in the curve. This vertebra occurs at the horizon or apex of a curve (T9 in Fig. 1).

The manual measurement of Cobb angle depends on experience and personal judgment. Errors are due to selecting different end- vertebrae and estimating different slopes of the vertebrae. The standard measurement error is 3° to 5° for the same observer and 5° to 7° for different observers.

The use of AP X-rays is useful for diagnosis, clinical studies and health learning. To help these purposes, it will be interesting to perform storage and retrieval of X-rays from a database based on curves parameters present in the spine. The problem is that, as we have shown, measures are imprecise and the query parameters needs to be flexible. For this reason, classical databases are not suitable for this purpose. It is necessary a Database System that can store imprecise data and can perform flexible queries on them. Moreover, as spine parameter description is complex, the database system must handle complex structures to represent imprecise data and must provide flexible comparators on these structures. Our FORDBMS, as we will show in next sections, have those capabilities.

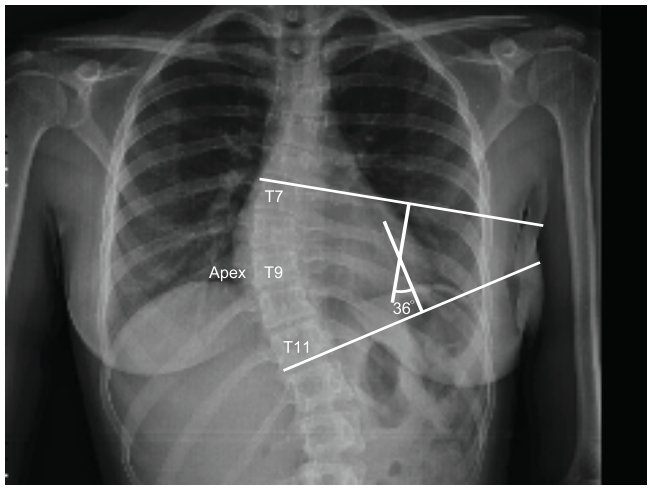


Figure 1: Cobb angle measurement.

3 The fuzzy object-relational database system

The recent proliferation of large image databases leads to the need for DBMS applied to multimedia libraries management to ensure high performance, scalability, availability with fault tolerance and distribution.

Nowadays, market leader DBMSs offer these required features transparently. However, the database models implemented by them, generally the relational model, are not suitable to manage fuzzy data, which is necessary for this kind of image description algorithms. In order to solve this drawback, some database models and DBMSs implementing them have been proposed. Nevertheless, the existing fuzzy DBMSs are in general research prototypes which do not match the high performance and other necessary requirements for this kind of applications.

In [12, 13] we introduce the strategy of implementation of our FORDMS model, that it is based on the extension of a market leader RDBMS (Oracle®) by using its advanced object-relational features. This strategy let us take full advantage of the host RDBMS features (high performance, scalability, etc.) and the ability for representing and handling fuzzy data provided by our extension, making this FORDBMS suitable to support systems for flexible content based retrieval of images.

3.1 Fuzzy datatype support

Our FORDBMS is able to handle and represent a wide variety of fuzzy datatypes, which allows to model any sort of fuzzy data easily. These types of fuzzy data, which are represented as classes with light gray background color in Fig. 2, are the following:

- Atomic fuzzy types (AFT), represented as possibility distributions over ordered (OAFT) or non ordered (NOAFT)

ISBN: 978-989-95079-6-8

domains.

- Fuzzy collections (FC), represented as fuzzy sets of objects with conjunctive (CFC) or disjunctive (DFC) semantics.
- Fuzzy objects (FO), whose attribute types could be crisp or fuzzy, and where each attribute is associated with a degree to weigh its importance in object comparison.

All fuzzy types define a Fuzzy Equal operator (FEQ) that computes the degree of fuzzy equality for each pair of instances. Each fuzzy datatype has its own implementation of this operator in accordance with its nature. Moreover, the FORDBMS provides parameters to adjust the fuzzy equality computation to the semantics of the data handled. For OAFTs the system implements other fuzzy comparators as FGT (fuzzy greater than), FLT (fuzzy less than), etc.

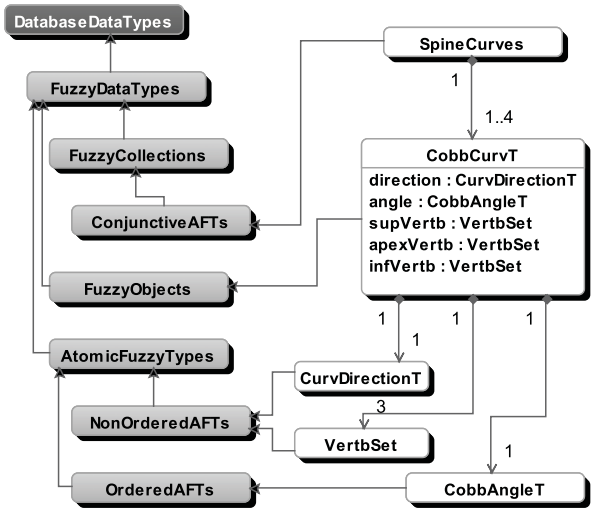


Figure 2: UML Class Diagram for Fuzzy types and SpineCurves Datatype.

3.2 Fuzzy datatype operators

Our FORDBMS can handle flexible comparisons over several types of fuzzy datatypes. Moreover, our system provides the possibility of adapting the behavior of some operators to adjust flexible comparisons to the semantics of the data modeled. The current section describes the most significant operators for the defined fuzzy datatypes, and their adaptation to handle flexible queries on complex objects like the spine curve description structure illustrated in this paper.

3.2.1 Fuzzy inclusion operator

The operator `FInclusion(A,B)` calculates the inclusion degree $A \subseteq B$, where A and B are instances of CFC. The calculus is done using a modification of the *Resemblance Driven Inclusion Degree* introduced in [19], which computes the inclusion degree of two fuzzy sets whose elements are imprecise.

Definition 1 (Resemblance Driven Inclusion Degree). *Let A and B be two fuzzy sets defined over a finite reference universe $\mathscr{U}$, μ_A and μ_B the membership functions of these fuzzy sets, S the resemblance relation defined over the elements of $\mathscr{U}$, $\otimes$ be a t-norm, and I an implication operator. The inclusion degree of A in B driven by the resemblance relation S is calculated as follows:*

$$\Theta_S(B|A) = \min_{x \in \mathscr{U}} \max_{y \in \mathscr{U}} \theta_{A,B,S}(x,y) \tag{1}$$

where

$$\theta_{A,B,S}(x,y) = \otimes(I(\mu_A(x), \mu_B(y)), \mu_S(x,y)) \tag{2}$$

We propose a modification that substitutes the minimum aggregation operator in equation 1 by a weighted mean aggregation operator, whose weight values are the membership degrees in A of the elements of $\mathscr{U}$, divided by the cardinal of A. This modification is made in order to obtain a less extreme resemblance inclusion degree, since it takes into account the importance of each included element. The *Modified Resemblance Inclusion Degree* is defined in equation 3.

$$\Theta_S(B|A) = \sum_{x \in \mathscr{U}} \frac{\mu_A(x)}{|A|} \cdot \max_{y \in \mathscr{U}} \theta_{A,B,S}(x,y) \tag{3}$$

with $|A| = \sum_{x \in \mathscr{U}} \mu_A(x)$. Our implementation of `FInclusion(A,B)` uses the minimum as t-norm, and the Gödel implication as the implication operator.

3.2.2 Fuzzy equality Operator

The operator `FEQ(A,B)` calculates the resemblance degree between two instances of a fuzzy datatype.

When `A` and `B` are two instances of CFC, this resemblance degree is calculated by means of the *Generalized Resemblance between Fuzzy Sets* proposed in [19], which is based on the concept of double inclusion.

Definition 2 *(Generalized resemblance between fuzzy sets). Let A and B be two fuzzy sets defined over a finite reference universe $\mathscr{U}$, over which a resemblance relation S is defined, and $\otimes$ be a t-norm. The generalized resemblance degree between A and B restricted by $\otimes$ is calculated by means of the following formulation:*

$$\beta_{S,\otimes}(A,B) = \otimes(\Theta_S(B|A), \Theta_S(A|B)) \tag{4}$$

Therefore, the implementation of the operator `FEQ(A,B)`, when A and B are instances of CFC, aggregates the results of `FInclusion(A,B)` and `FInclusion(B, A)` using a minimum t-norm.

If the operator `FEQ(A,B)` is applied when `A` and `B` are instances of the class *FO*, then, for the sake of flexibility, the resemblance degree between these objects is calculated as a weighted average of the resemblance degree of their attribute values. In the FORDBMS catalog we store some parameters for the attributes that belong to each fuzzy object. Specifically, the table *SDS_FO_ATTRIBUTES* stores into the column *relevance*, the relevance value of the considered attribute. This value attaches a weight to the attribute for the calculation of the `FEQ` operator. Moreover, if this value is set to -1 for an attribute, this attribute acts as determinant in the object comparison, in the way that, if the resemblance operator on this

ISBN: 978-989-95079-6-8

attribute returns 0, then the complete objects comparison returns 0. This parameter is useful, for example, in the case that we are searching for *"left curves"*, setting *Direction* attribute to -1, the query never retrieves a *"right curve"* even if resemblance for other object attributes is greater than 0. Another parameter that modifies the computation of the operator `FEQ` on fuzzy objects is the minimum percentage of attributes comparison greater than 0 (*min_pct*). This parameter, stored in the catalog table *SDS_FO_TYPES*, set the necessary number of attributes comparisons greater than 0 to get a value greater than 0 for the whole object comparison.

The following definition formalizes the ideas exposed above.

Definition 3 *(Parameterized Object Resemblance Degree). Let o_1 and o_2 be two objects of the class C, $obj.a_i$ the value of the i-th attribute of the object obj, $rel(a_i)$, relevance degree of the i-th attribute of the object obj, n the number of attributes defined in the class C, min_pct the necessary minimum percentage of attributes comparison greater than 0, and FEQ the resemblance operator.*

$$OR(o_1,o_2) = \begin{cases} 0 \text{ if} \\ (\exists i \in \mathcal{A} : FEQ(o1.a_i, o2.a_i) = 0 \wedge rel(a_i) = -1) \vee \\ (|\{a_i : i \in \mathcal{A} \wedge FEQ(o1.a_i, o2.a_i) > 0\}| < n.min_pct) \\ \\ \sum_{i\in\mathcal{A}}^{n} FEQ(o1.a_i, o2.a_i).|rel(a_i)| / \sum_{i\in\mathcal{A}}^{n} |rel(a_i)| \\ otherwise \end{cases} \tag{5}$$

Our FORDBMS also let us relax the resemblance operator `FEQ` applied on *OAFT* values. To do this, the user must use the static method `relax(kernel_pct,support_pct)`, where `kernel_pct` sets the percentage that the kernel increases for each pair of values being compared, and `support_pct` sets the percentage increase for the support of these values.

4 Retrieving X-Ray images from database using its fuzzy description

Our FORDBMS can handle complex fuzzy data structures to represent rich semantic problems. It can store instances of these structures and provides a powerful set of operators to retrieve data based on flexible parameters. We have chosen an example based on the representation of the characteristics of the spine curves taken from AP X-rays to illustrate how flexible queries can retrieve interesting results for the specialists. Let us show how to model the data structure of this example on our FORDBMS, then we will perform some queries and we will analyze the retrieved results.

4.1 Database modeling for spine curve description

Our FORDBMS is capable of representing curves structures present in the spine measured from AP X-Rays, taking into account that the measures can be imprecise due to errors. It also provides parameters to adjust the fuzzy comparators used to get good query results.

The classes whose background color is white in Fig. 2 integrate the datatype structure of the database that represents spine parameters. As this figure illustrates, we model the curve description of the spine as a fuzzy conjunctive collection datatype (*SpineCurves*) which includes up to four fuzzy object datatypes (*CobbCurveT*). *CobbCurveT* datatype represents a spine curve measure. This datatype has five attributes: *Direction*, *Angle*, *SupVertb*, *ApexVertb* and *InfVertb* that store: the side of the convexity of the curve (right or left), the angle measure and, the superior, apical and inferior vertebra of the curve, respectively. Attribute *Direction* is of NOAFT type, this allows to perform queries with a not determined value for the side of convexity of the curve. The attributes *SupVertb*, *ApexVertb* and *InfVertb* store values of type NOAFT. The domain of this attributes (*vertbSetT*) includes 24 vertebrae that are enumerated as follows: $v_1 = L5, v_2 = L4, v_3 = L3, v_4 = L2, v_5 = L1, v_6 = T12, v_7 = T11, v_8 = T10, v_9 = T9, v_{10} = T8, v_{11} = T7, v_{12} = T6, v_{13} = T5, v_{14} = T4, v_{15} = T3, v_{16} = T2, v_{17} = T1, v_{18} = C7, v_{19} = C6, v_{20} = C5, v_{21} = C4, v_{22} = C3, v_{23} = C2, v_{24} = C1$. We will define a nearness relation on this underlined domain to take into account the adjacency between each pair of vertebrae as follows:

$$\forall i,j \in Cardinal(VertbSett): \quad nearness(v_i,v_j) = \begin{cases} 1 & if\ i = j \vee i = j+1 \vee \\ & i = j-1 \\ 0.66 & if\ i = j+2 \vee i = j-2 \\ 0.33 & if\ i = j+3 \vee i = j-3 \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

The attribute *Angle* is of type OAFT, so we can store on this attribute numerical values, crisp or fuzzy (using a trapezoidal membership function). Measures obtained with errors for the angle can be represented as a fuzzy value for the attribute. This attribute allows flexible comparisons also. We can use the method `relax` described in the previous section on this attribute.

Once we have defined the type structure for the spine description, we can create a table that stores X-rays images together with their fuzzy descriptions as follows:

```
create table APXRay (
 image# number, xray bfile, SpineDescription SpineCurves);
```

We have inserted twenty images with their curve description to illustrate query examples; the statement for one image insertion is like this:

```
Insert into apxray values (
 313701 ,BFILENAME('APXRays','313701.gif'),
 SpineCurves(
     1,cobbCurvT(
        CurvDirectionT('RIGHT'),
        CobbAngleT(trapezoid(10.1)),
        VertbSetT('T2'),
        VertbSetT('T5'),
        VertbSetT('T10')
             ),
     1,cobbCurvT(
        CurvDirectionT('LEFT'),
        CobbAngleT(trapezoid(16.5)),
        VertbSetT('T9'),
        VertbSetT('T12'),
        VertbSetT('L2')
             ),
     1,cobbCurvT(
        CurvDirectionT('RIGHT'),
        CobbAngleT(trapezoid(12.5)),
        VertbSetT('L2'),
        VertbSetT('L4'),
        VertbSetT('L5')
             )
  )
);
```

ISBN: 978-989-95079-6-8

4.2 *Examples of Queries*

In this section we will illustrate some capabilities by performing a query that uses the `FInclusion` operator and a second one that uses the `FEQ` operator on `CFC` datatypes. To get flexible results on queries that involve whole fuzzy object comparison, it is advisable to relax comparison on angle attribute. To perform it we invoke the static method `relax` described in Section 3.2.2. In this case we relax 40% the kernel and 70% the support. This is the SQL sentence: `Execute CobbAngleT.relax(0.4,0.7);`

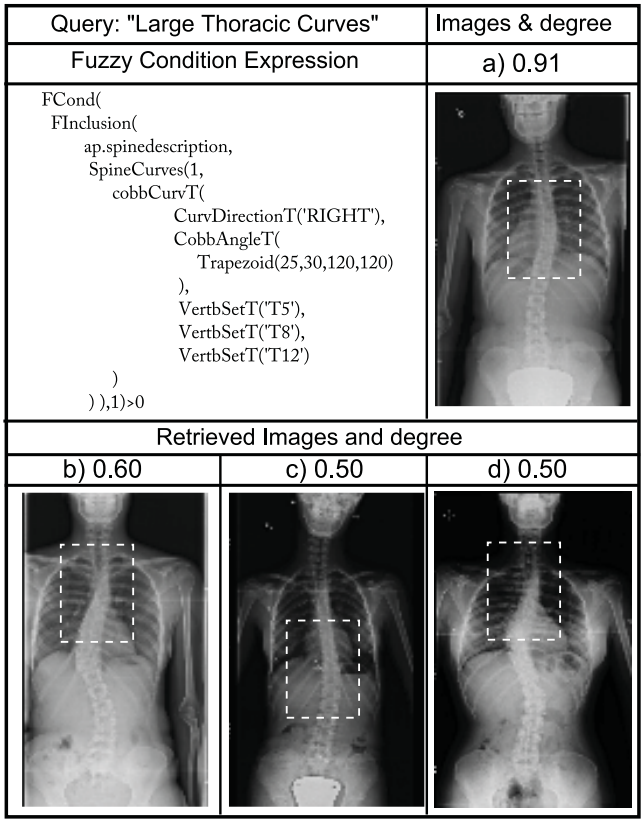


Figure 3: Searching images that include a thoracic pattern curve.

As shown in Fig. 3, we can perform queries that search images including a determined pattern of curves. In the example we illustrate the following query: *"Show X-Ray images and its compliance degree that include a large thoracic curve"*. The "thoracic curve" pattern proposed in [20] is described by the following approximate parameters: the convexity of the curve is on the right side, the superior end-vertebra is between T4 and T6, the apex vertebra is between T8 and T9 and, the inferior end-vertebra is between T11 and L2. The Fuzzy Condition Expression cell of the Fig. 3 shows the parameters used for the `WHERE` clause of the query statement, where the concept "large" is modeled as a trapezoidal distribution that matches Cobb angles that are approximately greater or equal than 25-30 degrees.

To perform the query we use the `FInclusion` operator, described in Section 3.2.1, which evaluates the similarity of the queried curve with respect to each curve included in the spine description. The query retrieves four images from the database. The image a), includes a curve, highlighted with a rectangular area, witch takes the values 'Right', 21.8, T5, T8, T10 for the parameters: orientation, angle size, superior, apex and inferior vertebrae, respectively. The compliance degree for this image with respect the query is 0.91. As we can see, this image includes a curve that appreciably matches the concept "large thoracic curve", this is the reason for the high degree computed. Visually, the marked curve in image b) weakly matches the concept "thoracic curve", this is the reason for a minor compliance degree. Image c) includes a curve that tends to a "thoracolumbar curve" concept; because of this, the retrieved compliance degree is low, 0.5. Finally, image d) includes a curve that tends to a "cervicothoracic curve" concept, the compliance degree (0.50) confirms this fact.

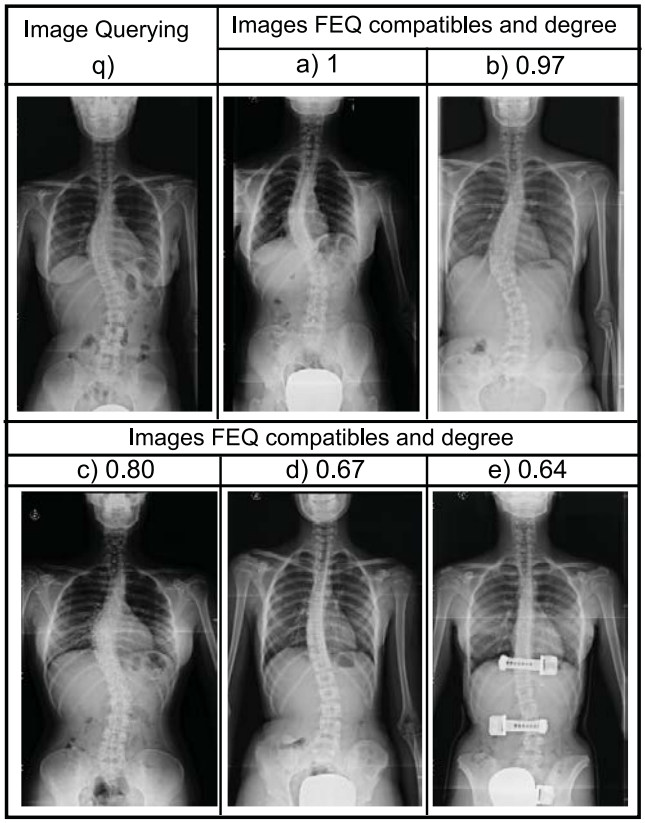


Figure 4: Searching images that present similar spine curvature that image q).

In some studies it may be interesting to find X-rays of patients that present a similar curve pattern to a given one. In our database, we model these curve patterns present in the spine as a *CFC*. To find similar curve patterns, our FORDBMS provides the operator `FEQ` which operates on *CFC* as has been described in Section 3.2.2. The query shown in Fig. 4 searches for X-rays of patients presenting similar curve patterns that the image query q). The SQL syntax is:

```
SELECT ap1.image#,ap1.xray,ap1.cdeg(1)
FROM apxray ap1, apxray ap2
WHERE ap1.image='q'
     AND FCOND(
          FEQ(ap1.spinedescription, ap2.spinedescription),1
         )>0 order by cdeg(1) desc;
```

As can be seen, higher compliance degree for an image denotes better visual matching with the queried image. Another interesting aspect of our results is the following: the images retrieved by the two kinds of queries evaluated hold that, the

ISBN: 978-989-95079-6-8

lower is the curve pattern matching, the lower the compliance degree computed for these one is.

5 Concluding remarks and future work

This paper shows that a FORDBMS is a powerful tool to represent flexible descriptors obtained from images by means of computer vision algorithms or provided by experts. The system also provides parameterized fuzzy comparators to retrieve images based on those descriptors in a flexible way. To the best of our knowledge, there is no system for flexible image retrieval that provides same capabilities on the subject considered in this paper. The prototype of FORDBMS used to show these results are implemented on the RDBMS Oracle® 10.2. Thereon, it is interesting to make the following remarks:

- The prototype has not implemented yet any index technique to accelerate the data retrieval based on fuzzy conditions. For this reason, the queries shown have been computed by means of a sequential search; moreover, the complex fuzzy datatype structure used for the spine curve representation more decreases the efficiency if none index technique is used. We are working on the implementation of the indexes techniques for fuzzy data proposed in [21, 22]; this will provide us mechanisms to optimize retrieval process.
- The scalability of our FORDBMS and CBIR system shown is guaranteed by the scalability of the host ORDBMS system, Oracle® 10.2.

Future work will involve too the enrichment of fuzzy features of the FORDBMS to enhance support for image content based retrieval. Finally we will study another application domains to perform content based retrieval using our FORDBMS.

Acknowledgment

This work has been partially supported by the Spanish "Ministerio de Ciencia y Tecnología" (MCYT) under grants TIN2006-07262 and TIN2007-68084-CO2-01, and the "Consejería de Innovación Ciencia y Empresa de Andalucía" (Spain) under research projects P06-TIC-01570 and P07-TIC-02611.

References

[1] Michael S. Lew, Nicu Sebe, Chabane Djeraba Lifl, and Ramesh Jain. Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2(1):1–19, 2006.

[2] J. Fuertes, M. Lucena, N. Perez de la Blanca, and J. Chamorro-Martínez. A scheme of colour image retrieval from databases. *Pattern Recognition Letters*, 22:323–337, 2001.

[3] Ju Han and Kai-Kuang Ma. Fuzzy color histogram and its use in color image retrieval. *IEEE Transactions on Image Processing*, 11(8):944–952, 2002.

[4] D. Sánchez, J. Chamorro-Martínez, and M.A. Vila. Modelling subjectivity in visual perception of orientation for image retrieval. *Information Processing and Management*, 39:251–266, 2003.

[5] B. Verma and S. Kulkarni. Fuzzy logic based interpretation and fusion of color queries. *Fuzzy sets and systems*, 147:99–118, 2004.

[6] S. Fukami, M. Umano, M. Muzimoto, and H. Tanaka. Fuzzy database retrieval and manipulation language. *IEICE Technical Reports, AL-78-85 (Automata and Language)*, 78(233):65–72, 1979.

[7] J.M. Medina, O. Pons, and M.A. Vila. Gefred. a generalized model of fuzzy relational databases. *Information Sciences*, 76(1):87–109, 1994.

[8] J.M. Medina, M.A. Vila, J.C. Cubero, and O. Pons. Towards the implementation of a generalized fuzzy relational database model. *Fuzzy Sets and Systems*, 75:273–289, 1995.

[9] H. Prade and C. Testemale. Generalizing database relational algebra for the treatment of incomplete or uncertain information and vague queries. *Information Sciences*, 34:115–143, 1984.

[10] M. Umano. Freedom-o: A fuzzy database system. In Gupta-Sanchez, editor, *Fuzzy Information and Decision Processes*. North-Holand, 1982.

[11] M. Zemankova-Leech and A. Kandel. *Fuzzy Relational Databases - A Key to Expert Systems*. Verlag TUV Rheinland, Koln, 1984.

[12] J.C. Cubero, N. Marín, J.M. Medina, O. Pons, and M.A. Vila. Fuzzy object management in an object-relational framework. In *Proc. 10th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU'2004*, pages 1767–1774, 2004.

[13] Carlos D. Barranco, Jesús R. Campaña, and Juan M. Medina. Towards a fuzzy object-relational database model. In José Galindo, editor, *Handbook of Research on Fuzzy Information Processing in Databases*, pages 435–461. IGI Global, 2008.

[14] B. Verdonck, R. Nijlunsing, F. A. Gerritsen, J. Cheung, D. J. Wever, A. Veldhuizen, S. Devillers, and S. Makram-Ebeid. Computer assisted quantitative analysis of deformities of the human spine. *Lecture Notes on Computer Sciences*, 1496:822–831, 1998.

[15] Mohammed Benjelloun and Sad Mahmoudi. Spine localization in x-ray images using interest point detection. *Journal of Digital Imaging. Published on-line, doi:10.1007/s10278-007-9099-3*, 2008.

[16] Junhua Zhang, Edmond Lou, Lawrence H. Le, Douglas L. Hill, James V. Raso, and Yuanyuan Wang. Automatic cobb measurement of scoliosis based on fuzzy hough transform with vertebral shape prior. *Journal of Digital Imaging. Published on-line, doi: 10.1007/s10278-008-9127-y*, 2008.

[17] JR Cobb. Outline for the study of scoliosis. *Am Acad Orthop Surg Inst Course Lect*, 5:261–275, 1948.

[18] J. Chamorro-Martínez, J.M. Medina, C.D. Barranco, E. Galán-Perales, and J.M. Soto-Hidalgo. Retrieving images in fuzzy object-relational databases using dominant color descriptors. *Fuzzy Sets and Systems*, 158:312324, 2007.

[19] N. Marín, J.M. Medina, O. Pons, D. Sánchez, and M.A. Vila. Complex object comparison in a fuzzy context. *Information and Software Technology*, 45(7):431–444, 2003.

[20] Ponseti and B. Friedman. Prognosis in idiopathic scoliosis. *J Bone Joint Surgery*, 32:381–395, 1950.

[21] Carlos D. Barranco, Jesús R. Campaña, and Juan M. Medina. A b+-tree based indexing technique for fuzzy numerical data. *Fuzzy Sets and Systems*, 159:14311449, 2008.

[22] Carlos D. Barranco, Jesús R. Campaña, and Juan M. Medina. B+-tree based indexing technique for necessity measured flexible conditions on fuzzy numerical data. In *Proc. 10th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU'2008 Torremolinos (Spain)*, pages 1717 – 1724, 2008.

ISBN: 978-989-95079-6-8

Minimal Combination for Incremental Grammar Fragment Learning

Nurfadhlina Mohd Sharef Trevor Martin Yun Shen

Artificial Intelligence Group, University of Bristol
Bristol, BS8 1TR, United Kingdom
Email: ennms@bris.ac.uk, trevor.martin@bris.ac.uk, yun.shen@bris.ac.uk

Abstract-*The ability to identify text fragments can benefit many information extraction tasks. The main challenge in this matter is to recognise the multiple patterns that exist across possibly heterogeneous sources. Traditional machine learning methods can generate grammars for a set of training examples but the disadvantage is their potential heavy computational overhead. The proposed incremental grammar learning method offers immediate learning of new examples by slight alteration of the existing grammar collections by measuring the changes needed to equalize two grammars. This minimal combination approach avoids overgeneralization and maintains the coverage of newly learnt and past examples without a need to repeat the training process. Our experimental results suggest that the proposed algorithm outperforms a fuzzy grammar learning algorithm implemented in normal supervised learning setting.*

Keywords- evolving fuzzy grammar, grammar fragments, incremental learning, grammar similarity, tagging, text mining.

1 Introduction

The "digital obesity" phenomenon is a pressing problem at both the individual and corporate level [1]. This is partly because information is recorded using different formats for similar purposes and according to different hierarchies and categories. However, it is likely for the objects placed within similar categories to have similar properties and with some degree of correspondence.

These information sources often contain text segments with additional structure that can be utilized to assist in the identification of matching items. For example, within a text segment such as: 'Burger King, 12, Meredith Lane, Ipswich IP2 3AX Ipswich' lies an embedded structure: business name, number, street name, placename, post code and county name while a brief comment such as: 'the Voyager has stopped receiving signal' contains information on the product name and problem description.

Works on lexical semantics and word disambiguation have studied text structures by focusing on linguistic patterns using formal language models i.e in the subject-verb-object structure. However, these methods are normally well-suited for texts that are in complete grammatical form. On the other hand, text segments such as addresses, dates and times, names of products, people, as well as simple sentence forms such as questions, complaints, and news are frequently contained within "casual" text such as email, posts on a web forum, text messages, etc which often does not follow formal grammar rules.

Humans are able to use clues that identify structured segments without necessarily following a strict pattern. There can be many variations of the patterns and humans are normally better at seeing the answer in a specific case than being able to define a pattern that is universally applicable. As such, this is an ideal problem for machine learning; the key subproblem is converting small sections of texts into a more structured form by using grammar fragments to add tags. Note that we are not attempting to parse the entire piece of free text, merely to locate sections within it that can be tagged with an additional structure. For this reason, we refer to grammar fragments, as they are not intended to be a complete language model.

We define grammar fragment learning as the extraction of text fragments that comply with the predefined grammar. In contrast to in-depth whole sentence understanding, the grammar fragments learning focus only on portions of text that are relevant to specific domains. For example, UK address grammars might extract the following patterns:

a. Chicago Rock Cafe Northgate St Ipswich IP1 3BX Suffolk
b. Lattice Lodge Guest House 499 Woodbridge Rd Ipswich
c. 25 St Nicholas Street Ipswich Suffolk IP1 1TW

while grammar collections for posts on a telecom provider forum will recognize the following question entries:

a. Can I use an apple with BT Broadband Talk and Softphone?
b. When does BT plan to develop and support a Softphone client for NON windows platforms such as Mac OSX or Linux?
c. What is Softphone and how much does it cost?

Several authors have proposed the use of genetic algorithm [2,3,4,5], fuzzy [6,7,8], semantic net [9] and hybrid probabilistic [10] to learn grammars mostly from structured examples but these follow the traditional route of training a fixed system on a set of examples, followed by application of that system. Adaptation to new examples (when the training set is found to be inadequate) will be computationally expensive and hence inefficient. The training set needs to be sufficiently diverse to cover all unseen examples or the training phase will need to be repeated in order to improve the training set. Thus, it is highly desirable for a method that can support the incremental learning of a new example by slight modification of the learned rules without reconsidering past examples.

The purpose of this study is to investigate an evolutionary grammar fragment learning with *incremental* feature i.e. it can be updated each time a new pattern is encountered (or each time a string is read) without extensive re-computation as compared to other supervised learning methods. The evolved grammar ensures the ability to encompass the newly learnt, as well as the previously seen, examples. Main issues in evolutionary grammar learning are also highlighted which are defining the fuzzy grammar learning operators and its representation aspect. This algorithm will benefit

ISBN: 978-989-95079-6-8

works in intelligent text pattern recognition and its relevant applications such as intelligent agent and recommendation system.
The paper is organised as follows: the first part gives the introduction to incremental grammar learning while section two describes some of existing works relating to text pattern learning. The third section details the proposed evolutionary algorithm setting followed by observed results in section four. The final section summarises the investigated issues and some future works related to the study.

2 Related Works

Text annotation and text tagging are among the active text extraction research areas. Information tagging has been applied for sentence structure learning in question answering system [11] as well as specific domains including finding medical terms in legal text [12,13], detecting contact information and address [14] and gene and protein interaction information identification [15,16]. Among the popular techniques that have been employed are CFG, statistical, and evolutionary method.
Viola and Narasimhan [14] reported the advantage of marrying statistical natural language processing to the analysis of non-natural language text through the Discriminative Context Free Grammar (Discriminative CFG) approach which has extracted contact information more accurately than a similar conditional Markov model. The hierarchically structured CFG method [3,7,14] is more powerful than the formal English language model such as the subject-verb-object and noun-adverb-adjective model because these models have limitations when more complex sentence forms are involved.
The CFG method provides a rich collection of features that can measure and tag the properties of a sequence of tokens. Another advantage is that the CFG can propagate long range token dependencies efficiently. CFG method can also result in the detection of semantic information (encoded in the form of a parse tree) within a sentence - for example the word *fish and chips* in the restaurant address entry can be classified as broad as a type of business, or a food business genre, or even as specific as a food name.
Backus-Naur-Form (BNF) is a simple method for specifying CFG. It consists of a set of production rules with various operators such as "[ ]" which means "*this part is optional*", "|" for *or*, etc. A BNF grammar is usually more easily deciphered than a regular expression. A method for generating BNF grammars from data was described in [17]. As with regular expressions, a BNF grammar does not generally allow partial matching and there are no methods for finding the similarity between two BNF grammars.
An alternative approach to this is the fuzzy grammar similarity algorithm [2,3,9] which can measure the degree of similarity between grammars. The incorporated fuzzy approach supports the uncertainty and redundancy notion in deriving the grammar. The fuzzy grammar method was applied in evolutionary grammar learning using genetic algorithm.
However, in common with other supervised learning algorithms, the grammars generated using GA method [2,3,4,5,6,10], can only recognize data belonging to the trained pattern and needs to be rebuilt to include newly added patterns. Our preliminary study on GA [6] shows that GA is unfeasible because of the slow convergence towards creation of a good parsing grammar. The genetic operators and random concept in GA had also ignored the semantic properties (sequence of the tokens that bears meaning) in the dataset.
The minimal combination approach discussed in this paper is a revised version from the maximal combination approach [6] which avoids overgeneralization. On the contrary, [6] has generated a lot of unseen examples which has also led to low accuracy. The preliminary experiment on genetic algorithm in [6] has shown that genetic algorithm method is unsuitable for the grammar learning task because it cannot support grammar similarity and is also burden by the slow convergence. Work described in this paper concentrates on an alternative minimal combination approach and demonstrates that different permutations of the training set have no significant effect on incremental evolutionary grammar learning.

3 Brief Introduction to Grammar Fragment Learning

The fuzzy parsing concept allows for the learning of the pattern in the strings and can be seen as an extension of sequence matching, in which: i) we allow one sequence to be defined by a grammar, so that the matching process is potentially more complex and ii) we allow the standard edit operations (insert, delete, substitute) on the string in order to make it conform to the grammar.
It is adapted from the Cocke-Younger-Kasami (CYK) algorithm to allow for:

a. fuzzy grammar fragments that can contain fuzzy sets of terminal symbols;
b. calculation of the overlap between fuzzy grammar fragments;
c. partial parsing, measured by the proportion of a string that has to be altered by Levenshtein-style operations in order to make it conform to a fuzzy grammar fragment.

We identify a grammar fragment with the set of strings that are parsed by the grammar fragment. Because we allow partial parsing, this is a fuzzy set. The fuzzy membership, $\mu_G(Si)$ calculates the similarity between string and grammar, estimated by the minimum number of edit operations needed to convert a string into one which is parsed by the grammar. A string's membership decreases as it needs more changes to make it conform to the grammar.

$$\mu_G(Si) = 1 - \min\left(1, \frac{\min_{P \in Ext(G)} (totalCost(Si, P))}{length(Si)}\right)$$

We define a cost to be a five-tuple *(I D S Rs Rt)* where *I*, *D* and *S* are nonnegative real numbers representing, respectively, the approximate number of insertions, deletions, and substitutions needed to convert a string parsed by the source grammar into one that satisfies the target grammar. *Rs* and *Rt* represent sequences of grammar elements remaining (respectively) in the source and target after the match; at least one of *Rs* and *Rt* is null in every valid cost. Table 1 shows a simple example of grammar similarity comparison between a source and a target grammar in a UK postal address. In this case, *Cost(sg,tg)=(0 1 1 countyName Null)* to substitute *word* with *placename*, and leave *word* and countyName unmatched.

ISBN: 978-989-95079-6-8

Table 1: Example of Grammar Edit Distance Operation
(*I:Insert, D:Delete, S:Substitute, Rs: remainder of source grammar, Rt: remainder of target grammar)

Source grammar, sg	Number	Word	Word	Streetending	Placename	
Target grammar, tg	Number	Placename		Streetending	Placename	Countyname
Edit distance*	0	S=1	D=1	0	0	Rt=1

The *totalCost* function defines a total order on costs by $C1 \le C2$ iff $totalCost(C1) \le totalCost(C2)$.

$$\text{totalCost}((I, D, S, Rs, Rt)) = I + D + S + \text{minLength}(Rs) + \text{minLength}(Rt)$$

Addition of costs is order dependent and is defined by

$$(I1, D1, S1, Rs1, Rt1) + (I2, D2, S2, Rs2, Rt2) = (I3, D3, S1 + S2, Rs2, Rt2)$$

where

$$I3 = I1 + I2 + \text{minLength}(Rt1)$$

$$D3 = D1 + D2 + \text{minLength}(Rs1)$$

For two grammars, *GS* and *GT*, *Overlap(GS, GT)* is the approximate degree to which an arbitrary string parsed by the source grammar *GS* is also parsed by the target grammar *GT*. We define the degree of overlap of a source grammar GS with a target grammar GT as

$$\text{Overlap}(GS, GT) = 1 - \min(1, \text{totalCost}(C)/\text{minLength}(GS))$$

where $C = CostGG(GS,GT)$ is an estimate of the cost of changing a string parsed by the grammar *GS* into one parsed by the grammar *GT*.

The *CostGG* function is defined as follows, where S and T are arbitrary sequences of grammar elements, *s*, *t* are terminal symbols, and *TSi* and *TSj* are (fuzzy) sets of terminal symbols.

$$\text{CostGG}(\text{nu}\mathbf{l}, T) = (0, 0, 0, \text{null}, T)$$

$$\text{CostGG}(S, \text{null}) = (0, 0, 0, S, \text{null})$$

$$\text{CostGG}(s, t) = (0, 0, \delta(s, t), null, null)$$

where $\delta(s, t)$ measures the cost of the symbol *s* replacing symbol *t*. By default $\delta(s, t) = 1$, if $t \neq s$ and $\delta(s, t) = 0$, if $t = s$.

A given grammar has full overlap with itself or any equivalent grammar, where we define equivalence as the ability to parse exactly the same set of strings. Conversely, two grammars have zero overlap if there is no string that both grammars can parse. For example, if *G1* defines postal address and *G2* defines email address then they have zero overlap as there are no strings that both can parse.

The degree of overlap is not symmetric, as can be seen if we consider two grammar fragments *G3* and *G4*, where the set of strings parsed by *G3* is a strict subset of those parsed by *G4* . In this case,

$$overlap(G3, G4) = 1$$

but

$$overlap(G4, G3) < 1.$$

Clearly, the degree of overlap could be determined by repeatedly generating and checking candidate strings—such an approach would be very computationally demanding.

Instead, we propose a method [2,3,9] based on estimating the number of edit operations needed to change an arbitrary string parsed by the first (source) grammar sequence into a string that would be parsed by the second (target) grammar. This will result the coverage of strings belonging to both grammars and is implemented through the grammar combination technique.

The former *maximal combination* technique [9] allows combination of grammars without filtering their cost. E.g, for the set of strings $S = \begin{Bmatrix} (a-b-c-d-e-f), \\ (a-b-a-c-d-e-f), \\ (b-c-d), \\ (a-d-e-f) \end{Bmatrix}$ the learned grammar is $G' = \{[a]-[b]-[a]-[c]-d-[e]-[f]\}$. Although this method can parse all the strings belonging to S, but it also allows for the parsing of strings which does not exist in S. Therefore our current method focuses on minimal combination; combination of grammars that have a difference of 1 and this have managed to avoid overgeneralization.

3.1 Incremental Evolutionary Grammar Learning Algorithm

As the fundamental of the study is to look at the pattern similarities within the strings, a parser that can tokenize and extract relevant symbols from each input is essential. This grammar derivation process will re-label the symbols with significant grammar markup (Fig. 2 and Fig. 3).

The incremental learning has allowed for the update of the available grammars when new pattern is triggered without needed to repeat the whole learning process (Fig.1). The employed minimal combination is governed by the cost between the source and the grammar; combination will only be executed when the change is one. This process contributes to the evolution of the whole grammar collection and the persistence of the parsing coverage.

A chosen grammar from the existing collection will be altered according to the grammar combination rules. Table 2 shows guidelines of grammar combination operations; insert, merge and create. These operations are based on the cost of changing the source to the target and the other way

Table 2: Grammar Combination Operation

Source Grammar	Target Grammar	Cost (Source,Target)	Cost (Target,Source)	Combination Operation	Combined grammar
a-b-c	a-b	0 1 0 - -	1 0 0 - -	Insert	a-b-[c]
a-b	a-b-c	1 0 0 - -	0 1 0 - -	Insert	a-b-[c]
a-b-c	a-B-c	0 0 0 - -	0 0 1 - -	Merge	a-B-c
a-B-c	a-b-c	0 0 0 - -	0 0 1 - -	Merge	a-B-c, B>b
a-F-c	a-G-c	0 0 1 - -	0 0 1 - -	Create	a-X-c , X:=F||G, $F \neq G$

ISBN: 978-989-95079-6-8

around. The *insert* operation is executed when the change based in the cost is a deletion or an insertion (Fig. 5). The combined grammar is always the altered version (with optional tags) of the longer grammar. When cost needs a substitute action and when generalization can be performed, *merge* operation is executed while *create* is executed when generalization is not suitable (Fig. 4). Note that the grammar combination is sensitive to the matched position of the grammar pairs. The notation '||' is only used as a convenience of the paper and not practised in the system development.

```
Algo. 1: Grammar Evolution
Input: S={s0,s1,…,sn}
Output: GT
Procedure:
initialise GT to empty set
gs = grammar derived from s[0] (Algo.3)
add gs to GT
for each string, s [i:1,...,n] in S[n]
  gt=grammar in GT that parses s with min cost C (Algo. 2)
  if C > 1 then
    Add gs to GT
  else
    if C = 1 and cost(gt,gs)=0 then
      replace gt with combine(gt,gs) (Algo.5)
    else if C = 0 and cost(gt,gs)=1 then
      replace gt with combine(gt,gs) (Algo. 5)
    else if C = 1 and cost(gt,gs)=1 then
      replace gt with combine(gs,gt) (Algo. 4)
    end if
  end if
```

Figure 1: Grammar Evolution Algorithm

```
Algo. 2: minimum parsing cost, C
Input: string,s, grammar gt[0... n]
Output: minGrammar and C
Procedure:
Initialise C to cost(s, gt[0] )
Initialise minGrammar to gt[0]
for each  [i:1,...n]
  if( C > cost(s, gt[i]) then
      C = cost(s, gt[i])
     minGrammar = gt[i]
  end if
end for
```

Figure 2: Maximum Parsing Cost Algorithm

```
Algo. 3: Grammar Derivation
Input:string, s
Input: terminal symbols representing S,
S ∈ {word, placename, streetending, number,...}
Output: gs, the grammar derived for s
Procedure:
initialise gs to empty string
initialise maxMem and mem to zero
for each token in s, s[i], [i:0,...n]
maxMem= fuzzy membership score of s[i] with t[0]
  for each terminal symbols, t[j], [j:1,...m]
    mem= fuzzy membership score of s[i] with t[j]
    if mem>maxMem and t[j] is less general than t[j-1] then
      append the terminal token into gs
    end if
  end for
end for
```

Figure 3: Grammar Derivation Algorithm

```
Algo. 4: Combine For Substitute Operation
Input: gs,gt
Output: combinedGrammar, subGrammar
Procedure:
initialise combinedGrammar and subGrammar to empty
string
Initialize n=lengthOf(gs)=lengthOf(gt)
for i=[0,…n]
  if gs[i] is equal to gt[i] then   combinedGrammar[i] = gs[i]
  else
     create  subGrammar= gs[i], subGrammar=  gt[i]
     append subGrammar into combinedGrammar[i[
end for
```

Figure 4: Combine (Substitute) Algorithm

```
Algo. 5: Combine For Insertion or Deletion Operation
Input: gs,gt
Output: combinedGrammar, subGrammar
Procedure:
initialize combinedGrammar, subGrammar  to empty string
if gs.length>gt.length then loop=gs.length
else loop=gt.length
for [i:0…loop]
   if gs[i] is equal to gt[i] then
      append gs[i] to combinedGrammar
  else
      if gs.length>gt.length then
         append gs[i] with optional tag to combinedGrammar
      else loop=gt.length
         append gt[i] with optional tag to combinedGrammar
  end if
end for
```

Figure 5: Combine (Insertion or Deletion) Algorithm

The impact of the grammar learning when the training set is prepared in various orders is being observed. Fig. 6 shows an example of a collection of grammars while Table 3 shows the grammar generated when the grammars are learned in the different orders as follows:

Order1: G_2-G_8-G_3-G_5-G_4-G_7-G_6-G_1.
Order 2: G_6-G_3-G_2-G_8-G_4-G_7-G_5-G_1
Order 3: G_2-G_5-G_3-G_8-G_4-G_7-G_6-G_1

Our observation has shown that the final set of evolved grammars can parse the same set of strings, with the same accuracy. Section 5 gives some remarks on this issue. However, further work is ongoing to investigate this in more detail from a theoretical and practical perspective.

$$G = \{G_1, G_2, ..., G_8\}$$
$$g_n = \{a, b, c, d, e, f, j, k\}, a > k > j$$
$$G_1 = w - n - w - se - pn - pc - cN$$
$$G_2 = w - w - n - w - se - pn - pc - cN$$
$$G_3 = w - w - acc - w - se - w - pn - pc - cN$$
$$G_4 = w - bt - w - se - pn - pc - cN$$
$$G_5 = w - w - n - w - se - w - pn - pc - cN$$
$$G_6 = w - w - acc - w - se - pn - pc - cN$$
$$G_7 = w - acc - w - se - pn - pc - cN$$
$$G_8 = w - w - bt - w - se - pn - pc - cN$$

Figure 6: Examples of training grammars

ISBN: 978-989-95079-6-8

Table 3: Evolved Grammars in Various Training Data Orders

Order 1	Order 2	Order 3
w-w-(bt||n)-w-se-pn-pc-cN ($G_2 \cup G_8$)	w-w-acc-w-se-[w]-pn-pc-cN ($G_6 \cup G_3$)	w-w-n-w-se-[w]-pn-pc-cN ($G_2 \cup G_5$)
w-w-(acc||n)-w-se-w-pn-pc-cN ($G_3 \cup G_5$)	w-w-(n||bt)-w-se-pn-pc-cN ($G_2 \cup G_8$)	w-w-acc-w-se-w-pn-pc-cN (G_3)
w-(bt||acc)-w-se-pn-pc-cN ($G_4 \cup G_7$)	w-(bt||acc)-w-se-pn-pc-cN ($G_4 \cup G_7$)	w-w-bt-w-se-pn-pc-cN (G_8)
w-w-acc-w-se-pn-pc-cN (G_6)	w-w-n-w-se-w-pn-pc-cN (G_5)	w-(bt||acc)-w-se-pn-pc-cN ($G_4 \cup G_7$)
w-n-w-se-pn-pc-cN (G_1)	w-n-w-se-pn-pc-cN (G_1)	w-w-acc-w-se-pn-pc-cN (G_6)
		w-n-w-se-pn-pc-cN (G_1)

4 Results and Discussion

The underlying ideas of the experiment are to examine approximate parsing, fuzzy overlap of grammar fragments, and their use in the incremental evolution of fuzzy grammars. To demonstrate the benefit of incremental evolution an experiment was carried out to compare the parsing scores of the grammars generated through fuzzy grammar learner (FG) versus the proposed incremental evolutionary learning grammar (IEG).

The term incremental is coined so that the grammar will always be able to parse all supplied string patterns (identified by grammars) including the triggered new string pattern. This simulates interaction with a human expert who would be able to correct errors—although clearly this disadvantages the FG method, which is not designed to undergo multiple training phases. On the other hand, the FG method is performed according to the normal supervised learning setting; the performance of the training dataset is measured by its parsing on the test dataset. No additional learning is allowed even the parsing score is lower than 1. Table 4 shows the terminal sets used in this experiment.

Table 4: Terminal Sets

Word (w)	Any token composed entirely of alphabetical characters
Number (n)	Any token composed of digits
Streetending (se)	One of {road, rd, avenue, ave, street, st, lane, ln} or similar
Placename (pn)	A handcrafted set of UK placenames
CountyName (cN)	A handcrafted set of UK district names
BusinessType (bt)	One of {café, restaurant, bar} or similar
Accommodation (acc)	One of {hotel, lodge} or similar
Alphanumeric (aN)	A token composed of alphabetical characters and digits

Two sets of address dataset in UK format are prepared; restaurant.xml and yellowpages.xml. The restaurant dataset contains a small set of restaurant addresses in the UK while the yellowpages dataset contains domestic addresses. Each dataset is split into 25% for training and 75% for testing. The restaurant dataset contains 85 entries while the yellowpages dataset contains 396 entries. Both have no repeated entries but naturally consist of similar or repeated string patterns. This experiment was repeated using three different orders of the training and test dataset, to test the validity of the results and the sensitivity and reliability of the algorithm. The same result was produced in the different data permutations. Future work on this study will focus on the formal theoretical proving of order independent learning to support this hypothesis.

Table 5 shows the comparison between the mean of maximum parsing score of the grammars learned. The average of exact parsing is a measure of the average of the maximum parsing score of the grammars on each string in the test dataset. Score of 100% means that all strings have been recognized by the grammar.

The IEG method generates new rules immediately and shows good incremental learning behavior as the grammar fragments evolve; meaning that the grammar fragments may be altered according to the change required to transform the source and target grammars so that they will be equally parsed with increased coverage. Thus, the IEG method will always give an optimal parsing score of 1 for each string it has encountered, as shown in Table 5. For each string in the dataset, there is either a grammar that can parse it partially or accurately; and a grammar can thoroughly parse a minimum of 1 string. This indicates that for every string in the dataset, there is at least a grammar that can parse it accurately (100% parsing). As opposed to the IEG method, the average of exact parsing in FG is much lower. Observation in the restaurant dataset shows that the minimum parsing score of the entries in FG is 0.667 and 0.75 in the yellowpages dataset. Both share a maximum of score=1.

As discussed in section 2, supervised learning method is constrained by its training stage; the training dataset should contain all the patterns expected to be seen in the test dataset. As shown in Table 5 the performance of the grammars generated in FG is lower than of IEG due to its insufficient training. The FG score in yellowpages dataset is higher than the restaurant dataset because the yellowpages dataset is larger than the restaurant dataset so during training a better coverage has been provided.

Table 5: Average of Exact Parsing

Setting	Dataset	Average of Exact Parsing (%)
FG	Restaurant	89.5
IEG	Restaurant	100
FG	Yellowpages	97
IEG	Yellowpages	100

This shows that the evolutionary learning method has successfully provided immediate and gradual learning to the grammars even when the training is found to be insufficient. The set of evolved grammars have successfully maintained the parsing coverage of the existing and newly combined

ISBN: 978-989-95079-6-8

patterns. The goal has been achieved where the adaptations are performed in light of new examples while preserving its current knowledge.

5 Summary and Conclusion

Incremental evolutionary grammar learning (IEG) is a method that can cope with newly found text pattern. It is an advantage over the expensive retrain-retest setting in traditional machine learning approaches when examples in the training phase are found to be insufficient. Empirical result shows that the IEG method has 100% parsing of the dataset compared to 89.5% and 97% achieved by the fuzzy grammar method. The incremental feature in IEG has also surpassed the genetic algorithm method both in the processing time and accuracy.

The underlying structure (grammar) of the text is first detected and corresponding grammar that represents the text is derived. The novel fuzzy grammar similarity is utilized to find similarity of the grammars. The evolution concept featured in this algorithm is reflected by the enrichment of the learnt patterns; where its incremental learning characteristic gradually updates its knowledge on the various patterns. Each time a new pattern is encountered, the adaptation is executed by a minimal adaptation which provides a *fit* coverage corresponding to the newly found and the altered grammar as well as all the strings that they belong. Over time, the set of evolved grammars become more concise, wider coverage yet sustaining its persistence.

Further work is ongoing to study the theoretical proving of the order-dependency effect on the evolved grammars, investigating a method to create more compact grammars and to expand the application of fuzzy grammar fragments to softer structures (e.g., identification of questions, negative comments, positive reviews, etc.).

References

[1] Martin, T. P. "Fuzzy sets in the fight against digital obesity," *Fuzzy Sets Syst.*, vol. 156, pp. 411–417, 2005.

[2] Petasis., G., (2004) eg-GRIDS: Context Free Grammatical Inference from Positive Examples using Genetic Search, International Colloquium on Grammatical Inference 2004, *Lecture Notes in Artificial Intelligence* 3264, pp. 223-234.

[3] Sakibara, Y., and Marumatsu. H., (2000) Learning Context-free Grammars from Partially Structured Examples, International Colloquium on Grammatical Inference 2000, *Lecture Notes in Artificial Intelligence* 1891, pp. 229-240.

[4] Smith, T.,C., and Witten, H., (1996) Learning Language Using Genetic Algorithm, *Lecture Notes in Computer Science* 1040 pp. 132-1445.

[5] Unold, O., and Cielecki, L.,(2007) Learning Context-Free Grammars from Partially Structured Examples: Juxtaposition of GCS with TBL, *Seventh International Conference on Hybrid Intelligent Systems*, pp. 348 – 352

[6] Mohd .S., N., Martin, T., Shen, Y. (2008) Incremental Evolutionary Grammar Fragments, *Proceedings of the UK Workshop on Computational Intelligence (UKCI 2008)*, pp. 89-94.

[7] Martin, T. and Azvine, B. (2006) Evolution of Fuzzy Grammars to Aid Instance Matching,. IEEE *International Symposium on Evolving Fuzzy Systems,* pp. 163 – 168.

[8] Martin, T, Shen., Y., Azvine, B. (2008) Incremental Evolution of Fuzzy Grammar Fragments to Enhance Instance Matching and Text Mining, *IEEE Transactions On Fuzzy Systems,* Vol. 16, No. 6, December 2008, pp. 1425-1438.

[9] Yuhua L., David M.L., Zuhair A. B., James D. O., and Keeley Crockett (2006) Sentence Similarity Based on Semantic Nets and Corpus Statistics,. *IEEE Transaction on Knowledge and Data Engineering*, Vol 18, No. 8 pp.1138-1150.

[10] Bosman, P. A. N. (2004) Learning Probabilistic Tree Grammars for Genetic Programming, Parallel Problem Solving from Nature - PPSN VIII, *Lecture Notes in Computer Science* Vol. 3242 pp. 192-201.

[11] Ravichandran, D., and Hovy, E.(2002) Learning Surface Text Patterns for a Question Answering System, *Proceedings of the 4th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 41-47.

[12] Dozier, C., et. al (2007) Fast Tagging of Medical Terms in Legal Text, *Proceedings Of The 11th International Conference On Artificial Intelligence And Law*, pp. 253-260.

[13] Serban, R., et al (2007). Extraction and Use of Linguistic Patterns for Modelling Medical Guidelines, *Artificial Intelligence in Medicine (2007) 39*, pp. 137-149.

[14] Viola, P. and Narasimhan, M., (2005) Learning to Extract Information from Semi-structured Text using a Discriminative Context Free Grammar,. *8th Annual International ACM SIGIR Conference*.

[15] Joshua, M. T, and Gilder, M. R., (2003) Extraction Of Protein Interaction Information From Unstructured Text Using A Context-Free Grammar, *Bioinformatics 19(16) Oxford University Press.* Vol. 19. No. 16. pp. 2046-2053.

[16] Tanabe, L., and Wilbur, W. J., (2002). Tagging Gene and Protein Names in Biomedical Text, *Bioinformatics Vol. 18, no. 8 2002*, pp. 1124-1132

[17] Mernik, M., Gerlic, G. , Zumer, V. and Bryant, B. R. (2003) Can a Parser be Generated from Examples, *Proceedings of the 2003 ACM Symposium on Applied Computing,* pp. 1063 –1067

ISBN: 978-989-95079-6-8

Managing imperfect temporal metadata in the Catalog services of Spatial Data Infrastructures compliant with INSPIRE

Gloria Bordogna[1] Paola Carrara[2] Marco Pagani[1] Monica Pepe[2] Anna Rampini[2]

1 CNR-IDPA,
Unità di Dalmine (Bg), Italy
2 CNR-IREA
Unità di Milano, Italy
Email: {gloria.bordogna, marco.pagani}@idpa.cnr.it, {carrara.p, pepe.m, rampini.a}@irea.cnr.it

Abstract—*In this paper we analyze the limitations of current recommendations of the INSPIRE (Infrastructure for Spatial Information in Europe) Directive as far as the temporal metadata definition for discovery purposes, and propose its extension so as to allow the representation and management of imperfect spatio-temporal metadata. We propose to extend the metadata in order to cope with the requirements of both metadata producers, who often are unable to specify precise values, and users who submit queries to catalog services for discovering interesting data, who may express soft selection conditions on metadata values. The proposal is illustrated and explained through an example taken from an active Spatial Data Infrastructure (SDI).*

Keywords— Geo-data, INSPIRE, metadata, Spatial Data Infrastructure, spatio-temporal imperfection.

1 Introduction

Infrastructures are complex systems in which a network of interconnected but autonomous components is used for the exchange and mobility of goods, persons, information. Their successful exploitation requires technologies, policies, investments in money and personnel, common standards and harmonised rules. Typical examples of infrastructures which are critical for society are transportation and water supply. In Information Technology, the term infrastructure could be related to communication channels through which information can be located, exchanged, accessed, and possibly elaborated. Since the last decade of the past Century, creation of infrastructures for spatial information emerged as an issue in some countries in order to prevent wasting of money in geo-data multiple creation, to favour their shares at multiple levels in the society, and to support decision making in various fields, such as the environment [1]. In 2007, the INSPIRE Directive of the European Parliament and of the Council entered into force [2] to trigger the creation of a European Spatial Data Infrastructure (ESDI) that delivers to the users integrated spatial information services*. These services should allow users to discover and possibly access spatial or geographical information from a wide range of sources, from the local to the global level, in an inter-operable way for a variety of uses. Discovery is performed through services that should follow INSPIRE standards and can be implemented through some products (either proprietary or not) that declare their compliance. Users' conditions, expressed through discovery service clients, are matched against archived metadata, describing associated geo-data, in order to give an answer to the above conditions. The matching result is a list (sometimes empty) of records of metadata satisfying the match. Discovery services currently provide an exact matching: users cannot express flexible selection conditions, enabling partial matching mechanisms between the ideal metadata and the archived metadata.

On the metadata side, their creation is usually in charge of data/services providers: metadata fields, their meaning, and their types/ranges of values are defined in INSPIRE through recommendation documents aimed at easing discovery [3], [4]. With respect to the temporal characterisation, current recommendations for metadata specification are inadequate for metadata providers: some needed fields are missing; the semantics of the recommended fields is ambiguous; they do not consider nor manage representation of uncertain or vague information due to incomplete knowledge.

In this paper we take into account this last drawback and propose to let imperfect temporal values be used by both metadata providers in describing the temporal validity of the geodata, and SDI users in expressing discovery conditions. Moreover we suggest mechanisms to introduce partial matching in discovery services.

2 Metadata extension proposal within SDI architecture of INSPIRE

2.1 SDI Architecture

An overview of the current understanding of the technical architecture of INSPIRE is depicted in figure 1; it has been adopted by the ESDI [5].

Discovery services are the element aimed at discovering information of interest to the users. They are connected to metadata catalogues, necessary due to the intrinsically distributed nature of an SDI, where information repositories

* Following the Directive, components of European SDIs include: metadata, spatial data themes (as described in Annexes I, II, III of the Directive), spatial data services; network services and technologies; agreements on data and service sharing, access and use; coordination and monitoring mechanisms, processes and procedures.

ISBN: 978-989-95079-6-8

are owned and maintained on the Web servers of different organisations. Services managing catalogues could have the twofold function of archives of metadata, which describe either geo-data or geo-services available in the SDI, and of discovery tools, able to match users' search conditions with archived metadata. Catalogues services should be fitted to let users express selection conditions that specify "what" is of interest through content keywords (e.g., landslides), "why" they are searched (e.g., to detect recurrence of landslides), "where" the interesting features should be located (e.g., in a bounding box surrounding the Alps), and "when" these features should have been observed (e.g., the date/s of the observations).

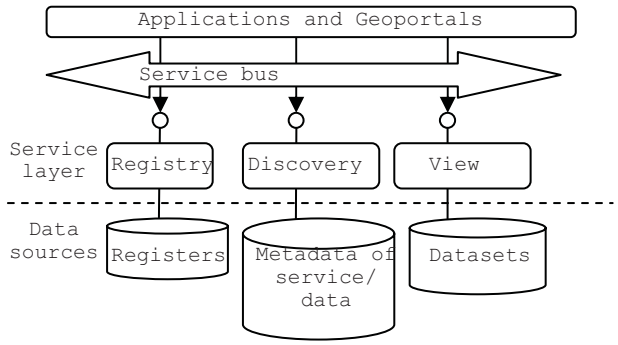


Figure 1: INSPIRE technical architecture overview; discovery service and metadata

2.2 Temporal metadata in INSPIRE SDIs

In order to ensure that the SDIs participating in the ESDI are compatible and usable in a trans-boundary, multicultural context such as Europe, the INSPIRE Directive requires that common Implementing Rules (IR) are adopted in a number of specific areas, in particular in metadata definition. The INSPIRE Metadata IR Version 3 of 26th October 2007 [4] defines four elements of temporal information, i.e.:

1. *Temporal extent*, the time period covered by the content of the resource
2. *Date of publication* of the resource
3. *Date of last revision* of the resource if the resource has been revised
4. *Date of creation* of the resource if it has not been revised

Only the first element (Temporal reference-5.1 Temporal extension) can be used to describe a temporal aspect of the content of geo-data itself, since the following three elements are referred to geo-data life cycle as a document. Nevertheless it plays an unavoidable role in numerous use cases, such as searching for events occurring in a particular period of time, looking for multiple observations of the same geographic area, evaluating if data can be used for a particular process, etc.

A specific study has been commissioned in order to improve temporal metadata specification [6]. It introduces some recommendations that rise attention to the role of 'Temporal extent' element; specifies the type and format of time values to be adopted either in expressing dates or time periods; suggests how to preserve precision in exchange or conversion processes.

2.3 Proposal of extension of temporal metadata

In our opinion and based on the experience as partners in the European Project IDE-Univers (http://www.ideunivers.eu), some further extensions should be adopted to improve temporal characterisation of geo-data in an SDI. In fact, the current specification of INSPIRE does not allow to specify the time relative to the occurrence of the event that is observed in one, several or a series of spatial data. Furthermore, current formats do not allow to specify imperfect values for metadata, both temporal, spatial, and thematic, which are the only kind of values that metadata producers can sometime provide, given that they often do not have enough information to specify precise values.

For example, if one has to create the metadata of a series of remote sensing images of a landslide that occurred in 1986 taken in distinct periods during its monitoring after its occurrence, the current specification just allows to indicate the time of the observations and not the time of the event, i.e. of the landslide occurrence. Nevertheless, this can be a very important information in an SDI context, because one can be interested in comparing the status of a landslide in different periods of time in the same place, as they appear in distinct images. One then, must be sure that the compared images refer to the same event, and not to distinct ones.

Moreover, in real situations, it often happens that the date or time of occurrence of an event is ill-known. This is due to several reasons: either because the event was not observed during its occurrence, and thus it can only be guessed by experts, or because it occurred long time ago so that its date can be deduced only imprecisely by the dating methods available (e.g., this is the case or archeological sites, geologic and paleontological findings).

In summary we propose:

- To introduce the possibility to include in metadata one or more events/processes/phenomena of reference for the geo-data: a satellite image can be an observation of fires in a region; a set of meteo records are measures of a rainfall; some thematic maps can be subsequent representations of urban growth, etc. A temporal extent element should be defined also for the reference event(s), of course.
- To allow the definition in metadata of soft temporal values in the temporal extent element of both geodata and related event(s); soft values should be allowed in the expression of single time points (by ex. the landslide occurred in the night of 15th December, 2005), of time intervals (by ex. snow precipitation of three days starting from 14th December 2008), and of subsequent either periodic or aperiodic dates/intervals (ex. fires occurring each Summer).

ISBN: 978-989-95079-6-8

3 Basic framework of the proposal

3.1 Extension of temporal metadata in INSPIRE

Notwithstanding we are aware that further extensions are possible, in this contribution we propose to include the following temporal metadata:

Instant in which the observation/event occurred:

- Date, time (e.g. the landslide occurred the 11-07-2008 at 8:00:00)

Period of validity, or duration of an observation/event:

- period (interval of dates, times) (e.g. the duration of a fire was from 11-07-2008 to 13-07-2008)

Sequence of instances of occurrences/events:

- multiple instants of time or dates (e.g. the dates of the distinct fires were 11-07-2008 , and 12-07-2008)

Sequence of periods of validity or duration of occurrences/events:

- multiple periods or durations (e.g. the times of the distinct fires were from 11-07-2008 to 13-07-2008 and from 20-08-2008 to 23-08-2008)

We also extend the time metadata format so as to allow the specification of ill-known instants, intervals and time series by means of fuzzy instant, fuzzy intervals and fuzzy time series associated with any kind of observation/event.

First of all we describe a formal framework in which all these fuzzy temporal indications are modeled, adopting the proposal of [7]. In order to express such fuzzy temporal indications into an eXtended Markup Language (XML)-type language, that is required for INSPIRE metadata, we adopt TimeML specification language [8].

The use of TimeML is motivated by the fact that it is a textual meta language, thus easy to read and to index by common Information Retrieval techniques, that can be employed in a catalog service context in order to represent the metadata contents for a successive search and discovery. It is enough flexible to allow the annotation (description) of the kind of event/observation and its temporal information, possibly imprecise and vague.

In the following subsections, let us first describe the representation within fuzzy set theory and possibility theory of time expressions, then introduce TimeML and specifically the tags we adopt, and finally, the partial matching mechanism we propose.

3.2 Modeling flexible time indications within the fuzzy set and possibility framework

The representation of temporal information requires the choice of a numeric scale to describe the order of events and phenomena; this scale is defined by an origin (time t=0) and a time unit, so that events occurred before t=0 have a negative time values and those occurred after have positive values. The choice of the time unit depends on the cycles of an observed periodic phenomenon, such as the movement of the sun. Nevertheless almost every time scale shares the basic unit *second*, and other units with lower granularity such as *minute*, *hour* and *day*. Besides these units of time, there are other natural language time indications such as *week*, *month*, *season*, *year*, *century* that are artificial notions, defined within a calendar such as the Julian and Gregorian ones. Other notions of time, related to human experience such as *now*, *soon*, *recent*, *often*, *end of the year*, are approximate time indications.

Several time models have been proposed [9], their common limitation is that they deal with crisp time indications.

A flexible framework has been defined within fuzzy set and possibility theory to express approximate hierarchical time indications close to natural language at distinct level of granularity [7]. It takes inspiration from the time hierarchy proposed in [10] and the time granularity defined as a mapping from the set of positive integers to the absolute time proposed in [11], and builds up a multi-granular hierarchical time structure in which also vague and imprecise time granules can be defined.

We assume this approach at the basis of our proposal. In this way it is possible to express temporal indications in several time units with distinct granularities, the less refined ones obtained by grouping units of the higher level granularity.

A *basic domain* G_0, consists of granules of time points below which the elements are not discernable. For example if G_0 is *hour* we cannot discern the *minutes* within an hour. Notice that, a granule of a domain G' (e.g. G'=*week*) that is not the basic domain G_0 (e.g. G_0=*hour*) can be defined by recursively grouping granules of its parent domains G=*day* and G_0 in the hierarchy. For example G'=*week*=7**day* =7*24*G_0=168*hours*.

The set of temporal specifications that we adopt are listed here following.

A *time point* indication is defined as a pair [t, G] in which t is an ordinal indicating the position with respect to the time origin on a domain G of granules. An example is: [t=2, *day*] that indicates the second day from the time origin; a fuzzy example is [t={0.8/3, 1./4, 0.7/5}, *day*] that means around the fourth day after the time origin.

A duration in time, i.e. a *time span*, is a pair [Δt, G] and can be denoted by either a set or a range of time points. A fuzzy time span example is [Δt ={0.8/3, 1./4, 0.7/5}, *year*] that means a duration of about 4 years.

A temporal distance from the origin, i.e. a *time distance*, is defined as a pair [d, G] in which d is a positive or negative value, indicating the distance in time granules on G from the origin. In this case [d=2, *day*] means two days after the origin. As t, also d can be a fuzzy set indicating a fuzzy time distance.

A *time interval* is a triple [t, Δd, G]; in a crisp case [t=1991, Δd=3, *year*] means 3 years from 1991.

A *composite span* is a union of spans $\cup[\Delta t_i, G_i]$, not necessarily adjacent and on the same basic domain G.

An *aperiodic time element* is a union of time intervals $\cup[t_i, \Delta d_i, G_i]$. The crisp example [t=1-11-2008, Δd=28, *day*] ∪ [t=30-11-2008, Δd=31, *day*] means 28 days from 1-11-2008 and 31 days from 30-11-2008.

A *periodic time element* is a union of a time interval and a time distance: $\cup[t_i, \Delta d_i, G_i], [d_k, G_k]$. By example ∪[t=1-8-2000, Δd=31, *day*], [d=1, *year*] means every August from

ISBN: 978-989-95079-6-8

year 2000. An example of vague periodic time element is ∪ [t=1-2000, Δd={0.2/1, 0.8/2, 1./3, 0,8/4}, *week*], [d=1, *year*] that means around the third week of every January from year 2000.

Since in the context of metadata compilation we may have time series that are related to finite repetitions of observations or events a *finite periodic time element* is defined as a composition of a periodic time element and a time point: $\cup[t_i, \Delta d_i, G_i], [d_k, G_k], [t, G]$ in which the time point t specifies the end of the repetition. An example of finite periodic time element is "*every autumn from 2000 to 2005*" that is formally expressed as: ∪[22-09-2000, 90, *day*], [1, *year*], [21-12-2005, *day*].

3.3 TimeML specification language for flexible time indications

TimeML is a markup language of the XML family for describing events, signals and temporal expressions into a text written in natural language. It is designed to address four situations:

1. Time stamping of events (identifying an event in time, instant or interval of validity).
2. Ordering events with respect to one another (relative ordering).
3. Reasoning with contextually underspecified temporal expressions (temporal functions such as 'last week' and 'two weeks before').
4. Reasoning about the persistence of events (how long does an event or the outcome of an event lasts).

The tags in TimeML[†] that we adopt to extend the INSPIRE metadata and that we model within the fuzzy framework previously described, are listed and illustrated in Table 1.

Table 1: TimeML tags adopted in this proposal

EVENT
Tag used to annotate the semantics of the event described. Syntactically, EVENT can be a verb (such as "raining"), but also a nominal, such as "fire"
MAKEINSTANCE
It indicates different instances (observations) of a given event. Different instances can have different attribute values, and every EVENT introduces at least one corresponding MAKEINSTANCE
TIMEX3
This tag is central to our objectives since it is primarily used to mark up explicit temporal expressions, such as times, dates, durations, etc. TIMEX3 allows marking up the following types of temporal indications specified by the attribute `type`: *Durations* such as "four weeks", "half a year"; *Calendar dates* (points in time equal or bigger than a day) both precise such as "13 August 2007" and imprecise or vague such as "few days ago", "end of July", "at the beginning of summer"; *Times of day* (smaller than a day) both precise such as "at 9.50.00 a.m." and imprecise or vague such as "before noon"; *Sets* (Recurring time expressions) such as "every month", "twice a week" The `value` attribute can assume XML datatypes based on the 2002 TIDES guideline, which extends the ISO 8601 standard for representing dates, times, and durations. E.g. "twelve weeks" becomes "P12W" and "21 February 2008 at 8.30.00 a.m." becomes "2008-2-21T8:30:00" The `mod` attribute allows specifying temporal modifiers that cannot be expressed either within `value` proper, or via links or temporal functions, such as "before", "after", "equal or less", "end".
TLINK
one of the TimeML link tags which encodes the relations that exist between two temporal elements (e.g., `BEGINS, HOLDS, INCLUDES, AFTER`)

For example the expression "*every autumn from 2000 to 2005*" is formulated in TimeML as follows:

```
<TIMEX3 tid="t10" type="SET" value="R6/2000-09-
22/P1Y0M0D">
every autumn from 2000 to 2005
</TIMEX3>
```

Finally, in TimeML it is possible to mark confidence values to be assigned to any tag and to any attribute of any tag. The confidence value associated with the `value` attribute of TIMEX3 expresses the uncertainty that the metadata provider has in assigning the temporal indication to an event or observation. For example, we can add the confidence annotation to TIMEX3 so as to specify the uncertain date of an observation that might have occurred the first or second of January 2000 as follows:

```
<TIMEX3 tid="t1" type="DATE" value="2000-01-01">
On January 1st, 2000
</TIMEX3>
<CONFIDENCE tagType="TIMEX3" tagID="t1"
confidenceValue="1."/>

<TIMEX3 tid="t2" type="DATE" value="2000-01-02">
On January 2nd, 2000
</TIMEX3>
<CONFIDENCE tagType="TIMEX3" tagID="t2"
confidenceValue="0.80"/>
```

4 Management of the proposed temporal metadata extension

The frameworks and assumptions of section 3 can be used to define the temporal fields associated with events and observations in the metadata of a SDI catalogue.

The metadata provider defines the time indications of events and observations by means of a metadata editor. We propose an example of the metadata of two thematic maps representing subsequent observations from satellite of the same event, i.e. the melting of the Lys Glacier (a glacier of the Italian Alps), during Summer 2007. In this example, the temporal indications of the occurrence of the observations

[†] TimeML vers. 1.2.1 <http://www.timeml.org>

ISBN: 978-989-95079-6-8

are known precisely. Following the extension we propose, in the metadata of the first map we could have fields such as:

Metadata 1
…
Event="Lys Glacier melting"
Occurrence="observation of Lys Glacier melting"
Time Position of Occurrence="1.7.2007"
…

In the metadata of the second map we could have:

Metadata 2
…
Event="Lys Glacier melting"
Occurrence="observation of Lys Glacier melting"
Time Position of Occurrence="3.9.2007"
…

Metadata are translated into TimeML sentences like in Table 2.

To allow partial matching with respect to flexible selection conditions specified by a user a parser translates the external TimeML definitions of temporal metadata into their internal fuzzy set representation μ_t. We can have fuzzy sets defined on distinct domains (G). The hierarchy of basic temporal concepts defining the temporal domains with distinct granularity, is represented by a graph that contains in each edge connecting node i to node j a mapping function $F_{i,j}$ that allows converting the current granule of the i-th node in terms of aggregation of granules of the j-th node (see Fig 2 which is a simplification to illustrate how the concept works).

This function $F_{i,j}$:G'→G, with G⊂G', and i,j identifiers of the i-th and j-th nodes, with i-th node defined on domain G' and j-th defined on G, associates a granule g'∈G'→$F_{i,j}$(g')∈G, where $F_{i,j}$(g') is a fuzzy set of granules on G. A temporal indication *t*, defined on G' (e.g., 2002 year) can be converted on another domain G (e.g., day) by repeatedly applying the mapping functions $F_{i,j}$ associated to the edges on the path from node G*'* to node G as proposed in [12].

Table 2: TimeML example

Lys Glacier
<EVENT eid="e10" class="OCCURRENCE">
is melting
</EVENT>
<MAKEINSTANCE eiid="ei1" eventID="e10"
pos="VERB" tense="PRESENT" aspect="PROGRESSIVE"
/>
<TIMEX3 tid="t2" type="DATE" value="2007-07-01">
On July 1st, 2007
</TIMEX3>
<TLINK eventInstanceID="ei1" relatedToTime="t2"
relType="DURING"/>
<MAKEINSTANCE eiid="ei2" eventID="e10"
pos="VERB" tense="PRESENT" aspect="PROGRESSIVE"
/>
<TIMEX3 tid="t3" type="DATE" value="2007-09-03">
On September 3rd, 2007
</TIMEX3>
<TLINK eventInstanceID="ei2" relatedToTime="t3"
relType="DURING"/>
…

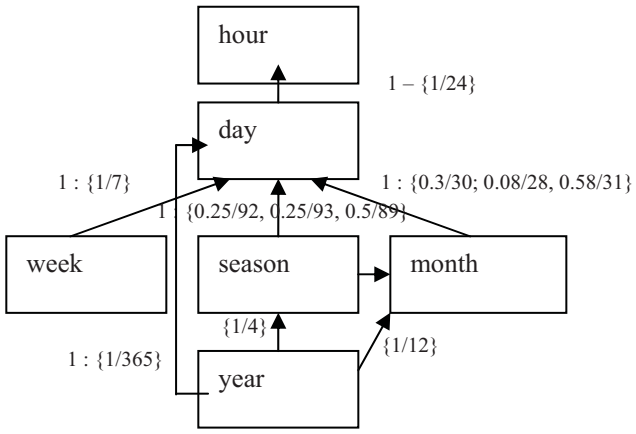


Figure 2: Simplified example of temporal graph: on each edge the fuzzy set defining the conversion function F of granules.

However, since there can be more than a single path P_1, ..P_k, connecting two nodes (e.g., P_1=*year*, *season*, *day* and P_2= *year, month, day*), multiple definitions of *t* on the domain G can be obtained t_{P1}, .. t_{Pk} In order to reconcile these definitions we choose the one obtained by considering the shortest path connecting the two nodes. If there exists more than one of such short paths with same length we generate the maximum bound of their definitions:

$$t_{P1 \oplus P2 \oplus Pk}(g) = \max(t_{P1}(g) .. t_{Pk}(g))$$

The reason for this choice is that the maximum bound comprehends all possible definitions of the temporal indication on the domain G.

For example, the sentences in Table 2 correspond to the two *time points* [t_1=1-7-2007, *day*], and [t_2=3-9-2007, *day*], with crisp pointwise membership functions.

On the other side, the user specifies her/his temporal soft selection conditions Q within a catalog service interface. They can be expressed by specifying soft constraints, i.e. with a desired membership function μ_Q, defined on a time line with a given granularity, chosen among one of the available in the temporal hierarchy (e.g. see fig. 3). An example of user selection condition for the Lys Glacier melting case reported in table 2 could be: *"Search for occurrences of glacier melting observations, occurring close to late Summer 2007"*.

Users' temporal query specifications are converted into the internal fuzzy set representation and, if necessary, transformed into the granularity of the data they have to be matched. The soft constraint of the example corresponds to a vague *time interval* with trapezoidal membership function μ_Q such as [t=15-6-2007, Δd={ 1-8-2007, 15-8-07, 23-9-2007, 10-10-2007}, *day*].

The internal fuzzy representation of the temporal metadata μ_t are then matched with the fuzzy representations of the soft query constraint μ_Q by applying a representation-based matching function as proposed in [13][14].

In this representation-based framework, both the metadata values μ_t, possibly uncertain, and the soft query condition μ_Q are interpreted as soft constraints and one can match

ISBN: 978-989-95079-6-8

them to obtain a degree of satisfaction $RSV(t,Q) \in [0,1]$, by computing either a measure of similarity [14] or a fuzzy inclusion measure [13] between the two fuzzy sets μ_Q and μ_t.

In the specific case of matching temporal constraints, it makes sense to allow the choice of both matching functions. One could be interested in selecting observations of an event taken in a date *close* to another date. This corresponds to select the similarity matching function [14]:

$$RSV(t,Q)=Similarity(\mu_t,\mu_q)=\frac{\sum_{i\in G} min(\mu_t(i),\mu_Q(i))}{\sum_{i\in G} max(\mu_t(i),\mu_Q(i))}$$

in which $\mu_t(i)$ and $\mu_Q(i)$ are the membership degrees of a time point i in a fuzzy metadata value t and in the query constraint Q.

Another case is when one wants to select observations that occurred *within* a period. This corresponds to select a matching function defined as a fuzzy inclusion [13]:

$$RSV(t,Q)=Inclusion(\mu_t,\mu_q)=\frac{\sum_{i\in G} min(\mu_t(i),\mu_Q(i))}{\sum_{i\in G}(\mu_t(i))}$$

Other matching functions could be defined corresponding with other temporal relations such as "*close before*" "*close after*" "*recent*" and so on.

The retrieved metadata can be ranked in decreasing order on the basis of *RSV* values, thus avoiding empty answers and suggesting an access order to the referred geo-data. In the example, both metadata are retrieved: Metadata 2 has a *RSV* score "1" being situated in late Summer, while Metadata 1 is also retrieved since it partially satisfies the query condition "*close*", thus meaning that it is associated to an observation of glacier melting that is in proximity of the limits of the temporal range 'late Summer'.

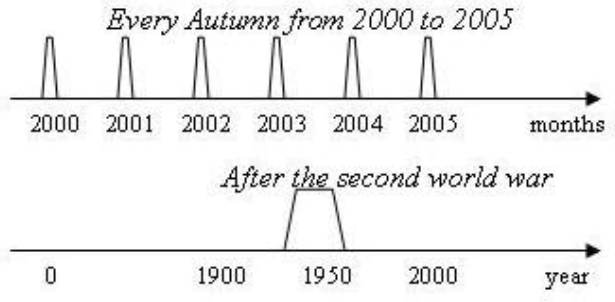


Figure 3: Examples of two soft temporal constraints defined on two timelines with distinct granularity (months and years, respectively). The constraint "*every Autumn from 2000 to 2005*" is defined as a fuzzy finite periodic time element, while "*after the second world war"* as a fuzzy time interval.

5 Conclusions

Spatial Data Infrastructures are becoming a common practice for discovering and accessing distributed heterogeneous geo-data. Nevertheless, the catalogue services on which SDIs are based are still founded on old database paradigms, that do not allow partial matching mechanisms, nor the representation and management of ill-defined metadata.

The present paper represents a first step in the proposal of a fuzzy framework to model catalogue services functionalities to manage ill-defined metadata and flexible queries. Specifically, we proposed a solution to express, represent, and manage temporal metadata, possibly imperfect, in a flexible way even if the revising policy within INSPIRE is long and needs to overcome several reviewing steps.

References

[1] Executive Order 12906, Coordinating Geographic Data acquisition and access: The National Spatial Data Infrastructure, edition of the U.S.A. Federal Register, Volume 59, Number 71, pp. 17671-17674, April 13, 1994

[2] Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 (INSPIRE), online: <www.ec-gis.org/inspire/directive/l_10820070425en00010014.pdf>, 2007

[3] Commission Regulation (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata, 2008

[4] Draft Implementing Rules for Metadata (version 3), European Commission, 2007-10-26, 2007

[5] INSPIRE Technical Architecture Overview, INSPIRE Cross Drafting Teams Report, 05-11-2007, online: < http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/network/INSPIRETechnicalArchitectureOverview_v1.2.pdf >, 2007

[6] M. Dekkers, Temporal Metadata for Discovery, M. Craglia (ed.), EUR 23209 EN, online: <http://inspire.jrc.ec.europa.eu/reports/TemporalMetadataforDiscovery>, 2008

[7] R. De Caluwe, F. Devis, P. Maesfranckx, G. De Trè, B. Van der Cruyssen, Semantics and Modelling of Flexible Time Indication, in Computing with words in Information/Intelligent Systems 1, Foundation, Zadeh L.A., Kacprzyk J., Physica Verlag, 1999

[8] TimeML, Markup Language for Temporal and Event expression, online:<http://www.timeml.org/site/index.html>

[9] V.J Tsotras., A. Kumar, Temporal database bibliography update, ACM SIGMOD Record, 25(1), 41-51, 1996

[10] R. Maiocchi, B. Pernici, F. Barbic, Automatic deduction of temporal indications, ACM Trans. On Database systems, 17(4), 647-668, 1992

[11] X.S. Wang, C. Bettini, A. Brodsky, S. Jajodia., Logical design for temporal databases with multiple granularities, ACM Trans. On database Systems, 22(2), 115-170, 1997

[12] R. De Caluwe, G. de Tré., B. Van der Cruyssen, F. Devos, P. Maesfranckx, Time management in fuzzy and uncertain object-orineted databases, in Knowledge management in fuzzy databases, O. Pons, M.A. Vila, J. Kacpzzyk (eds), Physica Verlag, 2000

[13] G. Bordogna, G. Pasi, A Flexible approach to evaluate soft conditions with unequal preferences in fuzzy databases to appear in Int. J. of Intel. Syst., 2007

[14] P. Bosc, L. Duval, O. Pivert., Value-based and representation-based Querying of Possibilistic Datases, in Recent Issues on Fuzzy Databases, G. Bordogna & G. Pasi eds, Studies in Fuzziness and soft computing n.53, Physica Verlag, 2000

ISBN: 978-989-95079-6-8

On Some Fuzzy Categories of Many-valued Topological Spaces

Ingrīda Uļjane
Department of Mathematics, University of Latvia
Institute of Mathematics and Computer Sciences, University of Latvia (IMCS UL)
Riga, Latvia
Email: uljane@mail.com

Abstract— *The aim of this paper is to construct some new fuzzy categories related to many-valued sets and many valued topologies and to consider their properties.*

Keywords— Fuzzy categories, GL-monoids, L-valued equalities, L-topologies and L-fuzzy topologies.

1 Introduction

Since the inception of the notion of a fuzzy set by Zadeh [19] and its generalization as an (L)-fuzzy set where L is a complete lattice by Goguen [3], the efforts of many researchers were forced to introduce the appropriate fuzzy counterparts of the classical mathematical concepts and to develop substantial corresponding theories. However from the categorical point of view in most cases the work was done in the context of ordinary, that is crisp categories. To explain our idea we give some examples

1. Following the works of Chang [2] and Goguen [4], an L-fuzzy topological space is a pair (X, τ) where X is a set and τ is a subfamily of L^X satisfying axioms which are natural analogues of classical topological axioms. The continuity of mappings between L-fuzzy topological spaces is defined just by "translating" the definition of continuity in fuzzy enviroment.

2. In [18] Mingsheng Ying introduced the concept of a fuzzifying topological space by semantical analysis of usual topological axioms, cf also [6]. Namely a fuzzifying topological space is a pair $(X, \mathcal{T})$ where X is a set and $\mathcal{T} : L^X \to L$ is a mapping satisfying the axioms which are obtained by logical interpretation of the standard topological axioms. Continuity of mappings between two fuzzifying topological spaces is defined by logical interpretation of the axiom "the preimage of an open set is open".

3. In [11] A. Rosenfeld defined a fuzzy subgroup of a group $(G, \cdot, e)$ as a mapping $H : G \to [0, 1]$ such that $H(x \cdot y) \geq H(x) \wedge H(y)$ and $H(x) = H(x^{-1})$ for all $x, y \in G$. A homomorphism between fuzzy subgroups H_1 and H_2 is defined as a homomorhism between the underlying groups $f : G_1 \to G_2$ such that $H_1(x) \leq H_2(f(x))$ for every $x \in G_1$.

Thus in all these cases as well as in many other situations, the authors work in the following context: they consider a (usual) category whose objects are certain mathematical structures involving fuzzy sets as objects and mappings satisfying certain properties between these objects as morphisms.

On the other hand in [13] the concept of a fuzzy category was introduced. A fuzzy category is some category-type conglomerate whose objects and morphisms may be such only to a certain degree, see the next section for the precise definition. The foundations of the theory of fuzzy categories were developed in a series of subsequent papers, see e.g. [14], [10], [15], etc. These works contain also description of some concrete fuzzy categories related to algebra and topology, which are obtained by a certain fuzzification of classical categories.

The aim of this paper is to construct some new fuzzy categories related to many-valued sets and many valued topologies and to consider their properties. They can be regarded as fuzzy counterparts of the category $\mathbf{SET}(L)$ [7] and categories $\mathbf{TOP}(L)$ and $\mathbf{FTOP}(L)$ introduced in our work [16]. The fuzzy categories considered in this paper were first defined in our talk at the Czech-Latvian Seminar *"Advanced methods in Soft Computing",* in Trojanovice, Czech Republic, November 19-22, 2008.

2 Fuzzy categories

First we recall the concept of an (L)-fuzzy category (where L is a GL-monoid) in a special form appropriate for our merits.

Recall (see [7]) that a $GL-$monoid is an infinitely distributive lattice $(L, \leq, \wedge, \vee)$ [1] enriched with a monotone, commutative and associative binary operation $*$ such that

1. $a * 1 = a$ and $a * 0 = 0$ for all $a \in L$ where 0 and 1 are the bottom and the top elements of L respectively;

2. $a * \left(\bigvee\limits_{i \in \mathcal{I}} b_i \right) = \bigvee\limits_{i \in \mathcal{I}} (a * b_i)$ $\forall a \in L$, $\forall \{b_i : i \in \mathcal{I}\} \subseteq L$;

3. If $a \leq b$, then there exists $c \in L$ such that $a = b * c$. (The last property of a GL-monoid is called divisability.)

Such operation $*$ will be refered to as the conjunction in the GL-monoid. Important examples of GL-monoids are Heyting algebras (in this case $* = \wedge$) and Łukasiewicz MV-algebra (in this case L = [0,1] and $a * b = \max\{a + b - 1, 0\}$) [5].

It is known that every $GL-$monoid is residuated, i.e. there exists a further binary operation " $\mapsto$ " (residuation) on L linked to $*$ with the Galois condition:

$$a * b \leq c \Longleftrightarrow a \leq (b \mapsto c) \qquad \forall a, b, c \in L.$$

Explicitely implication is given by

$$a \mapsto b = \bigvee \{\lambda \in L \mid a * \lambda \leq b\}.$$

ISBN: 978-989-95079-6-8

Let $L = (L, \leq, \wedge, \vee, *)$ be a GL-monoid. *An (L)-fuzzy category* [13] is a pair $(\mathcal{C}, \mu)$ where $\mathcal{C}$ is an ordinary category with the class of objects $\mathcal{O}b(\mathcal{C})$, the class of morphisms $\mathcal{M}or(\mathcal{C})$, and where $\mu : \mathcal{M}or(\mathcal{C}) \to L$ is an L-subclass of the class of morphism such that:

(i) $\mu(g \circ f) \geq \mu(g) * \mu(f)$ whenever composition $g \circ f$ is defined in the category $\mathcal{C}$;

(ii) for each $X \in \mathcal{O}b(\mathcal{C})$ $\mu(id_X) = 1$ where id_X is the identity morphism.

3 Category SET(L) and fuzzy category $\mathcal{F}$-SET(L)

Category **SET**(L) was first defined in [7], see also [8]. The objects of the category **SET**(L) are L-valued sets, that is pairs (X, E) where X is a set and E is an L-valued equality, that is a mapping $E : X \times X \to L$ such that

1) $E(x, x) = 1$;

2) $E(x, y) = E(y, x)$;

3) $E(x, y) * E(y, z) \leq E(x, z)\ \forall x, y, z \in X$.

An L-subset A of an L-valued set (X, E) (that is a mapping $A : X \to L$) is called extensional if $A(x) * E(x, x') \leq A(x')$ for all $x, x' \in X$. The family of all extensional L-subsets of (X, E) will be denoted L^X_E.
The morphisms $f : (X, E_X) \to (Y, E_Y)$ in the category **SET**(L) are extensional mappings, that is mappings $f : X \to Y$ such that

$$E_X(x, x') \leq E_Y(f(x), f(x')) \quad \forall x, x' \in X.$$

Consider the category

$$\mathcal{C}(L) = (\mathcal{O}b(\mathbf{SET}(L)), \mathcal{M}or(\mathbf{SET})),$$

that is $\mathcal{C}(L)$ has objects from **SET**(L) and morphisms from **SET**, that is all mappings between the corresponding sets.
We define the measure μ of extensionality for a mapping $f : X \to Y$ by

$$\mu(f) = \bigwedge_{x,x' \in X} (E_X(x, x') \mapsto E_Y(f(x), f(x'))).$$

Theorem 3.1 *The triple $(\mathcal{O}b(\mathbf{SET}(L)), \mathcal{M}or(\mathbf{SET}), \mu)$ is an L-fuzzy category. It will be denoted by $\mathcal{F}$-**SET**(L) and called the L-fuzzification of the category* **SET**(L).

Proof: Let $id_X : (X, E) \to (X, E)$ be the identity mapping. Then, obviously, $\mu(id_X) = 1$. Hence, to prove the theorem, we have to show that if $f : (X, E_X) \to (Y, E_Y)$ and $g : (Y, E_Y) \to (Z, E_Z)$ are potential morphisms (that is mappings between the corresponding sets), then $\mu(g \circ f) \geq \mu(g) * \mu(f)$. Indeed,

$$\mu(g \circ f) =$$
$$= \bigwedge_{x,x' \in X} (E_X(x, x') \mapsto E_Z((g \circ f)(x), (g \circ f)(x'))) \geq$$
$$\bigwedge_{x,x' \in X} (E_X(x, x') \mapsto E_Y(f(x), f(x'))) *$$
$$\bigwedge_{x,x' \in X} (E_Y(f(x), f(x')) \mapsto E_Z(g(f(x)), g(f(x')))) \geq$$
$$\mu(f) * \bigwedge_{y,y' \in Y} (E_Y(y, y') \mapsto E_Z(g(y), g(y'))) =$$
$$\mu(f) * \mu(g).$$

□

4 Category TOP(L) and fuzzy category $\mathcal{F}$-TOP(L).

Definition 4.1 *Let $L = (L, \wedge, \vee, *)$ be a GL-monoid and (X, E) be an L-valued set. A family $\tau \subseteq L^X$ is called an L-topology on an L-valued set (X, E) if*

1. *$0_X \in \tau; 1_X \in \tau$ (where $0_X, 1_X$ are constant L-sets with values 0 and 1 respectively;*
2. *if $U, V \in \tau$, then $U \wedge V \in \tau$;*
3. *if $U_i \in \tau \quad \forall i \in \mathcal{I}$, then $\bigvee_{i \in \mathcal{I}} U_i \in \tau$;*
4. *if $U \in \tau$, then $U(x) * E(x, x') \leq U(x')\ \forall x, x' \in X$.*

The last condition means that all L-sets in τ are extensional and hence $\tau \subseteq L^X_E$.
The triple (X, E, τ) where τ is an L-topology on an L-valued set is called an L-valued L-topological space. Respectively, the elements $U \in \tau$ are called open L-sets in this L-valued L-topological space.

Definition 4.2 *A mapping $f : (X, E_X, \tau_X) \to (Y, E_Y, \tau_Y)$ is called continuous if*

1. *$E_X(x, x') \leq E_Y(f(x), f(x'))$ for all $x, x' \in X$,* that is f is an extensional mapping between the corresponding L-valued sets (X, E_X) and (Y, E_Y), and
2. *$f^{-1}(V) \in \tau_X$ whenever $V \in \tau_Y$.*

Obviously, if $f : (X, E_X, \tau_X) \to (Y, E_Y, \tau_Y)$ and

$$g : (Y, E_Y, \tau_Y) \to (Z, E_Z, \tau_Z)$$

are continuous, then the composition

$$g \circ f : (X, E_X, \tau_X) \to (Z, E_Z, \tau_Z)$$

is continuous and the identity mapping

$$id_X : (X, E_X, \tau_X) \to (X, E_X, \tau_X)$$

is continuous. Hence L-valued L-topological spaces and continuous mappings between them form a category **TOP**(L).
One can easily see that in case E is crisp, that is $E(x, x') = 0$ if $x \neq x'$ and $E(x, x) = 1$, the L-valued toplogical space is just an L-topologcal space in the sense of [2], [4] and the category of L-topological spaces is actually a complete subcategory of **TOP**(L).

Notice that, by lower-semicontinuity of conjunction, for extensional L-sets U_i, we have

$$\left(\bigvee_{i \in \mathcal{I}} U_i(x)\right) * E(x, x') \leq$$
$$\bigvee_{i \in \mathcal{I}} (U_i(x) * E(x, x')) \leq \bigvee_{i \in \mathcal{I}} U_i(x') \quad \forall x, x' \in X.$$

Thus the supremum of extensional L-subsets of an L-valued set is extensional itself. Therefore, in an analogy with classical topology we can define the interior $int(A)$ of an L-subset A of an L-valued L-topological space (X, E, τ) as the largest $(\geq)$ one of all open L-subsets of (X, E, τ) contained $(\leq)$ in A. Equivalently, it can be defined by the formula

$$int(A) = \bigvee \{U \in \tau \mid U \leq A\}.$$

ISBN: 978-989-95079-6-8

One can easily verify that the resulting operator

$$int : L^X \to L^X$$

satisfies all properties, analogous to the properties of the interior operator in classical topology.

The interior operator allows us to measure "the degree of openess" of an L-set A in an L-valued L-topological space by evaluating, to what extent an L-set A is contained in its interior. To realize this we need to extend an L-valued equality E from a set X to its extensional L-powerset L_E^X. This can be done as follows (see [17]):
Given $A, B \in L_E^X$ let

$$\mathcal{R}(A,B) = \bigwedge_{x,z\in X} ((E(x,z) * A(x)) \mapsto B(z)).$$

It is shown in [17] that $\mathcal{R} : L_E^X \times L_E^X \to L$ is an L-fuzzy order relation on L_E^X, that is it is refelexive ($\mathcal{R}(A,A) = 1$ for all $A \in L_E^X$) and transitive ($\mathcal{R}(A,B) * \mathcal{R}(B,C) \le \mathcal{R}(A,C)$ for all $A, B, C \in L_E^X$). By setting $\mathcal{E}(A,B) = \mathcal{R}(A,B) \wedge \mathcal{R}(B,A)$ an L-valued equality on L_E^X is obtained [17].
Now to fuzzify the category **TOP**(L) we consider the category

$$(\mathcal{O}b(\mathbf{TOP}(L)), \mathcal{M}or(\mathbf{SET}), \circ),$$

where objects are L-valued L-topological space (X, E, τ), but morphisms are from the category **SET**, that is mapings between corresponding sets. For a given mapping

$$f : (X, E_X, \tau_X) \to (Y, E_Y, \tau_Y)$$

we define the degree of extensionality by

$$\mu_1(f) = \bigwedge_{x,x'\in X} (E_X(x,x') \mapsto E_Y(f(x), f(x')),$$

the degree of continuity by

$$\mu_2(f) = \bigwedge_{V\in\tau_Y} \mathcal{R}(f^{-1}(V), int(f^{-1}(V)))$$

and finally the degree of being a morphism by

$$\mu(f) = \mu_1(f) \wedge \mu_2(f).$$

Theorem 4.3 *$(\mathcal{O}b(\mathbf{TOP}(L)), \mathcal{M}or(\mathbf{SET}), \circ, \mu)$ is a fuzzy category.*

Proof Let $f : (X, E_X, \tau_X) \to (Y, E_Y, \tau_Y)$ and

$$g : (Y, E_Y, \tau_Y) \to (Z, E_Z, \tau_Z).$$

The inequality

$$\mu_1(g \circ f) \ge \mu_1(g) * \mu_1(f)$$

was established in the previous section. To establish the inequality

$$\mu_2(g \circ f) \ge \mu_2(g) * \mu_2(f)$$

we are reasoning as follows:

Let $f : (X, E_X, \tau_X) \to (Y, E_Y, \tau_Y)$,
$g : (Y, E_Y, \tau_Y) \to (Z, E_Z, \tau_Z)$.
We fix $W \in \tau_Z$. Then by transitivity of $\mathcal{R}$

$$\mathcal{R}(f^{-1}(g^{-1}(W)), int_X(f^{-1}(g^{-1}(W)))) \ge \mathcal{R}(f^{-1}(g^{-1}(W)), f^{-1}(int_Y(g^{-1}(W))))* \mathcal{R}(f^{-1}(int_Y(g^{-1}(W))), int_X(f^{-1}(g^{-1}(W)))).$$

We estimate each member separately:

$$\mathcal{R}(f^{-1}(int_Y(g^{-1}(W))), int_X(f^{-1}(g^{-1}(W)))) \ge \mathcal{R}(f^{-1}(int_Y(g^{-1}(W))), int_X(f^{-1}(int_Y g^{-1}(W)))) \ge \mu_2(f).$$

$$\mathcal{R}(f^{-1}(g^{-1}(W)), f^{-1}(int_Y(g^{-1}(W)))) = \bigwedge_{x,x'} ((E(x,x') * g^{-1}(W)(f(x))) \mapsto int_Y(g^{-1}(W)(f(x')))) \ge$$

(by extentionality of f)

$$\ge \bigwedge_{x,x'} ((E(f(x),f(x')) * g^{-1}(W)(f(x))) \mapsto int_Y(g^{-1}(W)(f(x')))) \ge \bigwedge_{y,y'} \left((E(y,y') * g^{-1}(W)(y)) \mapsto int_Y(g^{-1}(W)(y'))\right) \ge \mu_2(g).$$

Thus

$$\mathcal{R}(f^{-1}(g^{-1}(W)), int_X(f^{-1}(g^{-1}(W)))) \ge \mu_2(g) * \mu_2(f).$$

Since this is valid for any $W \in \tau_Z$ we obtain the requested

$$\mu_2(g \circ f) \ge \mu_2(g) * \mu_2(f).$$

Finally referring to the properties of a GL-monoid, we get

$$\mu(g \circ f) = \mu_1(g \circ f) \wedge \mu_2(g \circ f) \ge (\mu_1(f) * \mu_1(g)) \wedge (\mu_2(f) * \mu_2(g)) \ge (\mu_1(f) \wedge \mu_2(f)) * (\mu_1(g) \wedge \mu_2(g)) = \mu(f) * \mu(g).$$

We conclude the proof by noticing that obviously $\mu(id_X) = 1$ for the identity morphism

$$id_X : (X, E_X, \tau_X) \to (X, E_X, \tau_X).$$

□

5 Category FTOP(L) and fuzzy category $\mathcal{F}$-FTOP(L)

The concept of an L-valued L-fuzzy topological space as a generalization of the concept of an L-fuzzy topological space [12], [9] and the corresponding category **FTOP**(L) were introduced in [17]. The objects of this category are triples $(X, E, \mathcal{T})$ where (X, E) is an L-valued set and $\mathcal{T} : L^X \to L$ is an L-fuzzy topology, that is an extensional subset of L_E^X such that

1) $\mathcal{T}(1) = \mathcal{T}(0) = 1$;

2) $\mathcal{T}(U \wedge V) \ge \mathcal{T}(U) \wedge \mathcal{T}(V) \quad \forall U, V \in L_E^X$;

3) $\mathcal{T}\left(\bigvee_{i\in I} U_i\right) \ge \bigwedge_{i\in\mathcal{I}} \mathcal{T}(U_i) \quad \forall \{U_i \mid i \in \mathcal{I}\} \subseteq L_E^X$.

and morphisms are extensional mappings

$$f : (X, E_X) \to (Y, E_Y)$$

such that $\mathcal{T}(f^{-1}(V)) \ge \mathcal{T}(V)$.
One can easily see that in case E is crisp, that is $E(x,x') = 0$

ISBN: 978-989-95079-6-8

if $x \neq x'$ and $E(x,x) = 1$, the L-valued L-fuzzy toplogical space is just an L-fuzzy topological space as it was defined in [12], [4] and the category of L-fuzzy topological spaces is actually a complete subcategory of $\mathbf{FTOP}(L)$.

Let $(X, E_X, \mathcal{T}_X)$, $(Y, E_Y, \mathcal{T}_Y)$ be L-valued L-fuzzy topological spaces and $f : X \to Y$ be a mapping of the corresponding underlying sets. The measure of the extensionality μ_1 of the mapping $f : (X, E_X) \to (Y, E_Y)$ is defined as above. The measure of continuity μ_2 of a mapping $f : (X, \mathcal{T}_X) \to (Y, \mathcal{T}_Y)$ is defined by

$$\mu_2(f) = \bigwedge_{V \in L_E^Y} \left(\mathcal{T}_Y(V) \mapsto \mathcal{T}_X(f^{-1}(V))\right).$$

Further, for a mapping $f : (X, E_X, \mathcal{T}_X) \to (Y, E_Y, \mathcal{T}_Y)$ we set $\mu(f) = \mu_1(f) \wedge \mu_2(f)$.

Theorem 5.1

$$\mathcal{F}-\mathbf{FTOP}(L) = (\mathcal{O}b(\mathbf{FTOP}(L)), \mathcal{M}or(\mathbf{SET}), \mu)$$

is an L-fuzzy category.

Proof: Since for the identity morphism

$$id_X : (X, E_X, \mathcal{T}_X) \to (X, E_X, \mathcal{T}_X)$$

obviously $\mu_1(id_X) = 1$ and $\mu_2(id_X) = 1$, we have $\mu(id_X) = 1$. Therefore to prove the theorem we have to establish that given two functions $f : (X, E_X, \mathcal{T}_X) \to (Y, E_Y, \mathcal{T}_Y)$ and $g : (Y, E_Y, \mathcal{T}_Y) \to (Z, E_Z, \mathcal{T}_Z)$, viewed as the mappings of the corresponding underlying sets it holds

$$\mu(g \circ f) \geq \mu(g) * \mu(f).$$

The inequality $\mu_1(g \circ f) \geq \mu_1(g) * \mu_1(f)$ was established in subsection 3. The inequality

$$\mu_2(g \circ f) \geq \mu_2(g) * \mu_2(f)$$

was established in [14], see also [15]. Now, referring to the properties of a GL-monoid, we get

$$\begin{gathered}\mu(g \circ f) = \mu_1(g \circ f) \wedge \mu_2(g \circ f) \geq \\ (\mu_1(g) * \mu_1(f)) \wedge (\mu_2(g) * \mu_2(f)) \geq \\ (\mu_1(g) \wedge \mu_2(g)) * (\mu_1(f) \wedge \mu_2(f)) = \mu(g) * \mu(f).\end{gathered}$$

□

References

[1] G. Birkhoff, *Lattice Theory*, 3rd ed., AMS Providence, RI, 1967.

[2] C.L. Chang, *Fuzzy topological spaces*, J. Math. Anal. Appl., vol. 24, (1968), pp. 182-190.

[3] J.A. Goguen, *L-fuzzy sets*, J. Math. Anal. Appl., vol. 18, (1967) 145-174.

[4] J.A. Goguen, *The fuzzy Tychonoff theorem*, J. Math. Anal. Appl., vol. 43, (1973) 737-742.

[5] J. Łukasiewicz, *Selected works: Studies in Logic and the Foundations of Mathematics*, Borkowski eds., North Holland, Amsterdam, 1970.

[6] U. Höhle, *Upper semicontinuous fuzzy sets and applications*, J. Math. Anal. Appl. vol.78, (1980), 659–673.

[7] U. Höhle, *M-valued sets and sheaves over integral commutative cl-monoids*, In: *Applications of Category Theory to Fuzzy Subsets* S.E. Rodabaugh, E.P. Klement and U. Höhle eds., Kluwer, Dodrecht, Boston, 1992, pp. 33 – 72.

[8] U. Höhle, *Commutative residuated l-monoids*, In: *Non-Classical Logics and their Applications to Fuzzy Subsets - A Handbook of the Mathematical Foundations of Fuzzy Set Theory*, U. Höhle and E.P. Klement eds., Kluwer, Dodrecht, Boston, 1994, pp. 53-106.

[9] T. Kubiak, On Fuzzy Topologies, Ph. D. Thesis, Adam Mickiewicz Univ., Poznań, Poland (1985).

[10] T. Kubiak, A.Šostak, *A fuzzification of the category of M-valued L-topological spaces*, Applied General Topology, vol. 5, (2004), 137–154.

[11] A. Rosenfeld, *Fuzzy groups*, J.Math.Anal.Appl., vol. 35, (1971), 512-517.

[12] A. Šostak, *On a fuzzy topological structure*, Suppl. Rend. Circ. Matem. Palermo, Ser II, vol. 11, (1985), 89-103.

[13] A. Šostak, *On a concept of a fuzzy category*, In: *14th Linz Seminar on Fuzzy Set Theory: Non-Classical Logics and Their Applications*, Linz, Austria, 1992, 63-66.

[14] A. Šostak, *Fuzzy Categories Related to Algebra and Topology* Tatra Mount. Math. Publ., vol. 20, (1999), 159 - 186.

[15] A. Šostak, *On some Fuzzy Categories Related to the Category L-TOP of L-topological spaces* Chapter 12 In: Topological and Algebraic Structures in Fuzzt Sets, E.P. Klement, S.E. Rodabaugh eds., Trends in Logic, vol. 20, Kluwer Academic Publ., 2003.

[16] I. Uļjane, A. Šostaks,*On a category of L-valued L-topological spaces*, Acta Univ. Latv., vol. 2005, pp. 99-107.

[17] I.Uļjane, *On the Order Type L-valued Relations on L-powersets*, Mathware and Soft Computing, Vol. XIV, n.3 (2007), pp. 183. - 199.

[18] Ying Mingsheng, *A new approach to fuzzy topology*, Part I, Fuzzy Sets and Syst., vol. 39, (1991), 303-321.

[19] L.A. Zadeh, *Fuzzy Sets*, Inform. and Control, vol. 8, (1965), 338-365.

ISBN: 978-989-95079-6-8

Valued Constraint Satisfaction Problems Applied to Functional Harmony

Nele Verbiest, Chris Cornelis and Yvan Saeys
Ghent University, Ghent, Belgium
Email: {Nele.Verbiest, Chris.Cornelis, Yvan.Saeys}@UGent.be

***Abstract*— *Harmonization with four voices is a musical problem which is subject to hard constraints, which absolutely need to be fulfilled, as well as to soft constraints, which preferably hold, but are not mandatory. In this paper, we model this problem as a valued constraint satisfaction problem (VCSP): costs are assigned to possible solutions based on the constraints they violate. We design an algorithm that finds a minimal-cost solution, thus solving the harmonization problem, and we present initial results obtained by this algorithm.*

Keywords— functional harmony, valued constraint satisfaction problems, soft constraints

1 Introduction

Ever since people have had computers at their use, they have attempted to make them acquire human skills like rational thought and creativity. While the prospect of computers actually creating art might still seem out of reach, in musical composition some successful efforts have already been made; in [1], Truchet and Codognet give an overview of musical problems that were solved using computers.

In this paper, we discuss the use of Valued Constraint Satisfaction Problems (VCSPs) in functional harmony with four voices (soprano, alto, tenor, bass). Solving harmonization problems is a basic skill that every composer has to acquire before attempting more serious work. These problems range from very simple to advanced, with many gradations in between. Their solution is governed by a set of strict rules that have to be verified, but also by a variety of "softer" guidelines with varying, subjective importance, imposed to make the result more interesting, more varied, ... Valued constraints can be constructed and enforced sensibly to model them.

The structure of this paper is as follows: in Section 2, we discuss related work, and in particular solvers that have already been proposed for this problem, and position our own contribution. In Section 3, we briefly explain what functional harmony is about, and which constraints govern harmonization problems. In Section 4, we give the necessary background on VCSPs, while in Section 5 we translate the musical composition problem and its constraints to this framework. In Section 6 we describe the algorithm designed for solving our VCSP. In Section 7 we present some results obtained by this algorithm, and in the last section we discuss its practical use and suggest future research directions.

2 Related work

Programs to compose music have been designed since the 50s. We are specifically interested in harmonization problems with four voices, a problem that can be stated as a constraint satisfaction problem [2]. To solve this problem, two main classes of techniques have been described in the literature: backtracking and genetic algorithms.

Backtracking algorithms tackle the problem in the following way. The harmonization of a piece of music proceeds in a left to right fashion, analyzing every subsequent note. For each note, one seeks a possible chord for that note, and if there is no solution for this note, the algorithm goes back to a preceding note to try other possibilities. The standard work in this class is due to Ebcioğlu [3]: he introduced an expert system for harmonizing chorales with four voices, based on a set with about 350 rules. Apart from this, there are also backtracking algorithms where possible chords for a note are ranked by musical suitability. At every step, one tries the best chord. If this chord does not lead to a solution, the chord is removed from the ranking and a solution with the second best chord is sought [4]. Existing backtracking solvers look for one solution and then stop.

Genetic Algorithms (GAs) use an evaluation function to judge the musical quality of a set of solutions, and proceed in an iterative way to optimize these solutions [5, 6], using evolutionary reproduction operators like mutation and crossover.

Both approaches have their advantages and disadvantages. Backtracking algorithms are very efficient in finding a single solution, but little can be said about its quality: any solution found by the algorithm is guaranteed to meet the imposed hard constraints, but soft constraints are generally not considered. Ranking possible chords for a note deals only partially with this problem; this tactic will find a good solution, but it is not guaranteed to be the best one. By virtue of the evaluation function, genetic algorithms are better at catering for soft constraints, but they also suffer from the fact that their solutions are often suboptimal.

The solver we describe in this paper combines the use of soft constraints (which can be seen as the analogon of the evaluation function in GAs) with the potential to find an optimal solution in the search space.

3 Functional harmony with four voices

We briefly describe the essentials of functional harmony, and refer the interested reader to [7, 8] for more details. The problem we consider is a simplification of the general harmonization problem: given a melody for a soprano, we aim to find corresponding melodies for alto, tenor and bass. We suppose that only quarter notes are used in all the melodies and we work in C major. In tonal music, one speaks about grades, which —for a particular tuning— coincide with a set of chords. The most important grades that are always considered are I, IV and V. Other grades that often appear are VI and sometimes II. In C major tuning, the first note is c which results in grade I being the chord built on c; the fourth note is f,

ISBN: 978-989-95079-6-8

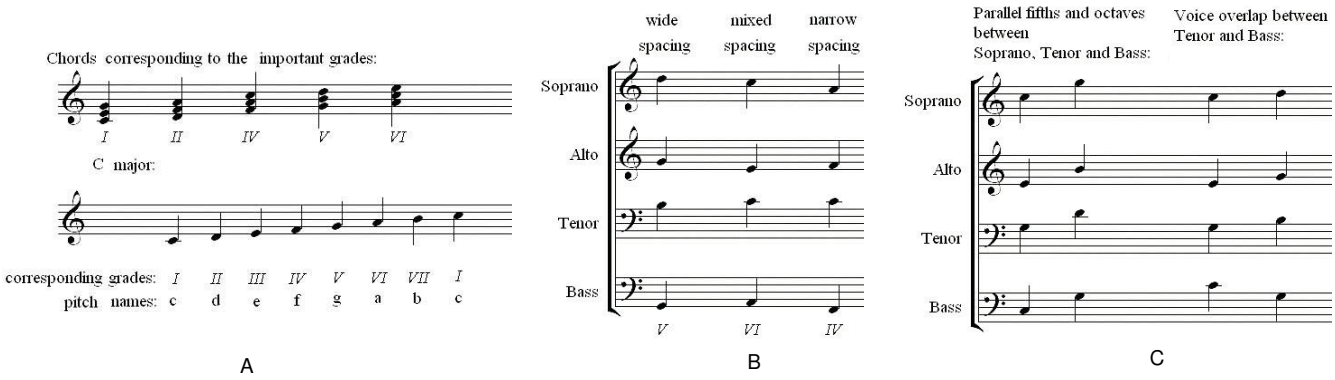


Figure 1: Essentials of functional harmony, illustrating the grades and chords of C Major (A), the concepts of narrow, mixed and wide spacing (B), and some examples of constructs that are not allowed (C).

so the chord corresponding to grade IV is the chord built on f. Fig. 1A illustrates the grades and chords of C major.

Now consider a note in the soprano. This note is in one or more of these chords. In harmonizing the soprano, one must choose a chord, and attribute the corresponding notes in the other voices to a note of this particular chord. The bass must always be the ground note (the lowest note) of the chord. For example, if IV is chosen, then the bass must be f. The soprano, alto and tenor have to be different from each other, so each note of the chord appears once, except for the ground note which appears twice.

We now introduce the concepts *wide* and *narrow* spacing. For every note in the soprano, we still have some freedom for the notes in alto, tenor and bass. We now restrict the possibilities a bit more. For each chord we choose, we have to decide whether we write it in wide or in narrow spacing. Narrow spacing means that between soprano and alto, and between alto and tenor there is no other note of the chord. Wide spacing means that there is always one note of the chord between soprano and alto, and between alto and tenor. This is illustrated in Fig. 1B. If the previous chord is written in narrow spacing, then the next chord must be narrow spaced as well. If the previous chord is wide spaced, then so must be the next chord.

The VIth grade chord is written like the other chords, except for the case when it is preceded by a Vth grade. In this case, the second note of the chord is written twice. This means that it has neither wide nor narrow spacing; we refer to this situation as *mixed* spacing. When we encounter a sequence V-VI it is possible to go from wide to narrow spacing or vice versa using mixed spacing on the VIth grade. Fig. 1B shows a good V-VI-IV sequence where we switch spacing.

Now we know how to write the chords we must consider the problem of choosing the right grade. To this end, a number of rules have been defined in the domain of functional harmony:

- The first and the last grade of a music piece must be I. The other grades depend on the previous grades.
- Some transitions between grades are not allowed, others are considered very good, while still others are not considered wrong but should be avoided as much as possible.
- Each of the different voices has its own range of notes that it can sing. The soprano can sing the highest notes, the bass sings the lower notes, and care has to be taken that these ranges are respected.
- It is forbidden to write parallel fifths or parallel octaves, to write the same grade before and after a bar or to overlap voices.

Some of these forbidden constructs are shown in Fig. 1C.

4 Valued constraint satisfaction problems

A constraint satisfaction problem (CSP) involves assigning values to variables that are subject to some constraints [9]. If there are constraints that are preferred but not essential, we can consider the problem of assigning values to the variables in such a way that the more important constraints are fulfilled and the less important ones are fulfilled to the extent possible. The latter yields a valued CSP (VCSP), which is characterized by a set of hard constraints (that must be fulfilled), and a set of so-called *valued constraints*. If the valued constraints are treated as mandatory, the problem is often unsolvable. If the valued constraints are ignored, we get solutions of bad quality. Our interest is therefore in the solution that best respects the set of constraints. To express what the best solution is, we assign costs to valued constraints. Every time a valued constraint is violated, we count the cost of this violation. In the end, we are interested in the solution with the lowest cost. We recall a formal definition of VCSP's.

Definition 1. *A Valued Constraint Satisfaction Problem (VCSP) is a quadruple* (X, D, C, V)*, where*
$X = \{X_1, ..., X_n\}$ *is a set of variables,*
$D = \{D_1, ..., D_n\}$ *is a set of finite domains:* D_i *is the set of possible values for* X_i*,*
$C = \{C_1, ..., C_r\}$ *is a set of hard constraints and*
$V = \{V_1, ..., V_s\}$ *is a set of valued constraints.*

A hard constraint C_i on the ordered set of variables $var(C_i)$ is the relation of the allowed combinations of values for the variables in $var(C_i)$; C_i is a mapping from the cartesian product of domains of variables in $var(C_i)$ to $\{0, 1\}$.

A valued constraint V_i on the ordered set of variables $var(V_i)$ specifies the cost of the combination of the values for the variables in $var(V_i)$; V_i is a mapping from the cartesian product of domains of variables in $var(V_i)$ to $[0, 1]$.

For example, if V_i pertains to the variables X_1, X_2 (i.e.

ISBN: 978-989-95079-6-8

$Var(V_i) = (X_1, X_2)$) and we assign values x_1 and x_2 to X_1 and X_2 respectively, the cost of this combination is given by $V_i(x_1, x_2)$. If C_i pertains to the variables X_1, X_2, X_3 (i.e. $Var(C_i) = (X_1, X_2, X_3)$) and we assign values x_1, x_2 and x_3 to these variables, the combination (x_1, x_2, x_3) is allowed if $C_i(x_1, x_2, x_3) = 1$ and forbidden if $C_i(x_1, x_2, x_3) = 0$.

5 Modeling the problem

We can now translate the musical problem of harmonizing a soprano to a purely mathematical problem. Suppose we have a soprano that lasts n times, with every time divided in 4 counts. All the times contain four quarter notes, except for the last time which contains one whole note. This makes a total of $l := (n-1) * 4 + 1$ notes. We call l the length of the problem. For every note, we want to store the notes for the bass, tenor, alto and soprano and we want to store the grade we use. This can be done using a $5 \times l$ matrix X.
The grades and notes are represented by a positive integer, and the central c is represented by the number 24. We assign integers to the other notes as follows: if the note is higher/lower than the central c, we count the number of half tones between the central c and this note and add/subtract it to 24. In this way, we fix the domain and variables of our VCSP. The variables are the entries $x_{i,j}, i = 1, ..., 5, j = 1, ..., l$ of the matrix X and the domain is the set of natural numbers $\mathbb{N}$.

5.1 Hard constraints

We will first translate the hard constraints to mathematical formulas. We only use grades I, II, IV, V and VI, which entails:

$$x_{5,j} \in \{1, 2, 4, 5, 6\}\ \forall j = 1, ..., l.$$

The first and last grade must be I, so

$$x_{5,1} = x_{5,l} = 1.$$

Every voice has a specific range of notes that it can sing, which can be expressed as:

$$\text{alto} : 19 \leq x_{2,j} \leq 38\ \forall j = 1, ..., l$$

$$\text{tenor} : 12 \leq x_{2,j} \leq 28\ \forall j = 1, ..., l$$

$$\text{bass} : 5 \leq x_{2,j} \leq 26\ \forall j = 1, ..., l$$

The grades before and after a bar must be different:

$$\forall j = 1, ..., l :$$

$$\text{if } j \bmod 4 \equiv 1 \text{ and } j \neq 1$$

$$\text{then } x_{5,j} \neq x_{5,j-1}$$

No parallel fifths or octaves are allowed:

$$\forall i, k, j : i, k = 1, 2, 3, 4, j = 1, ...l - 1 :$$

$$\text{if } (x_{i,j} - x_{k,j} \bmod 12 \equiv 0) \text{ and } k \neq i$$

$$\text{then } (x_{i,j+1} - x_{k,j+1} \bmod 12 \not\equiv 0)$$

and

$$\text{if } (x_{i,j} - x_{k,j} \bmod 12 \equiv 7) \text{ and } k \neq i$$

$$\text{then} (x_{i,j+1} - x_{k,j+1} \bmod 12 \not\equiv 7).$$

It is forbidden to write voice overlap (strictly between soprano, alto and tenor, not strictly between tenor and bass):

$$\forall j = 1, ..., l :$$

$$x_{1,j} > x_{2,j} > x_{3,j} \geq x_{4,j},$$

$$\forall j = 1, ..., l - 1 :$$

$$x_{2,j+1} < x_{1,j}$$

$$x_{3,j+1} < x_{2,j}$$

$$x_{4,j+1} \leq x_{3,j}.$$

The possibilities for the notes in the bass, tenor and alto and the forming of the chords (with wide and narrow spacing) are also hard constraints. We will not explicitly state these constraints here as mathematical formulas due to space constraints.

5.2 Valued constraints

We are now ready to introduce some valued constraints. There are five aspects of functional harmony we model using valued constraints. In the following subsections we propose techniques for calculating the cost of certain combinations. The results are five costs in $[0, 1]$ that correspond to the transition of one grade to another (c_1), the frequency of grades (c_2), the range of notes a voice can sing (c_3), the distance between two succeeding notes (c_4) and the use of contrary motion between bass and soprano (c_5). To allow a composer to determine the relative importance of each type of valued constraints, we attach weights $\{w_1, w_2, ..., w_5\}$ and define the total cost as $\sum_{i=1}^{5} w_i c_i$ where all weights sum to one. Assigning these weights depends on personal taste. Some composers might think it is important to have a lot of variation in the use of grades, while others might prefer to write music that is in the right range for the different voices. The weights we propose are as follows: $w_1 = 0.3, w_2 = 0.2, w_3 = 0.1, w_4 = 0.2, w_5 = 0.2$.

5.2.1 Grade transition constraints

As stated before, we only work with grades I, II, IV, V and VI. The most important grades are I, IV and V. V can be seen as a chord that builds up tension, I creates rest and IV is an advancing chord (to V). In music, it is important to have enough tension but there are points of rest needed as well.
II is a grade that can be used as a replacement grade for IV. It is also an advancing grade to V. VI is a grade that can replace I or IV, depending on the context. Generally speaking, when followed by V or preceded by I, VI is a replacement grade for IV, while when preceded by V or followed by IV, VI is a replacement grade for I. (This is also valid if we replace IV by II.) So depending on the context, VI can bring rest in music or can advance to tension. II and VI are used to bring more variation in music. For example, instead of writing I-IV-V-I (a very good sequence) all the time, we can also write I-II-V-VI. To express the aptitude of grade sequences, we introduce a valued constraint G. It is a mapping from the cartesian product of the possible grades to the interval $[0, 1]$ that represents the cost of the sequence:
$G : \{1, 2, 4, 5, 6\} \times \{1, 2, 4, 5, 6\} \rightarrow [0, 1]$.
To determine the cost of a sequence, we only look at two consecutive grades. This is sufficient since every bad sequence is

ISBN: 978-989-95079-6-8

the result of one or more separate bad successions in the sequence.
The assignment of the costs is an intuitive and subjective task. However, we try to bring some structure in it. The model sequence is I-IV-V-I. Therefore, we assign cost zero to every subsequence of length two that appears in it. All sequences that can be obtained from those subsequences by substituting a grade by its replacement grade (we call it *derivative sequences*) also have no cost.
Going from advancing to tension to rest is not very interesting. Therefore, we assign cost 1 to the sequence IV-I and all derivative sequences. Going from rest to tension without advancing should be avoided but is not particularly bad, so we assign cost 0.5 to the sequence I-V. Using the same sequence two times is not always interesting. We assign cost 0.3 to sequences I-I, V-V and derivatives, and cost 0.4 to sequences IV-IV and derivatives. An overview of the assignments is given in Table 1.

5.2.2 Grade frequency constraints

As stated before, in music it is important to maintain sufficient variation. This is not covered by the previous constraints: writing I-IV-V-I the whole time would give zero cost, but it would be boring to listen to. To express this we introduce additional valued constraints.
The algorithm designed in Section 6.1 has three options: it can generate a solution with only grades I, IV and V, only grades I, IV, V and VI, or using all grades I,II, IV, V and VI. The number of times a grade should ideally be used depends on this option, so we introduce three different valued constraints. First consider the option where only grades I, IV and V can be used. The best sequence here is I-IV-V-I. We want some variation so we sometimes write I-I, I-IV-I or I-V-I. Say that ideally, every three times we write I-IV-V, we should write I-I, I-IV-I or I-V-I once. The corresponding grade occurence percentages are given in Table 2.
We now define the valued constraint $H_1 : \{1, 4, 5\}^l \rightarrow [0, 1]$ as follows: for the whole piece of music, count how many times grades I, IV or V are used and calculate the percentages. Now for every grade, add the absolute difference between the ideal and real percentage, and divide the result by three to obtain a value in $[0, 1]$.
Since grade VI is a replacement grade for I, in the option where I, IV, V and VI can be used, we consider the same percentages as in the first option for IV and V, but we split the ideal percentage for grade I. Analogously, as grade II has the same function as grade IV, we split the ideal percentage for grade IV in the third option. The percentages are presented in Table 2. The valued constraints H_2 and H_3 are defined as before.
Again, these ideal percentages are constructed subjectively and intuitively. Different composers would propose different percentages, but they should be similar. Moreover, as for the previous valued constraints, the exact percentages are not very important, more significant are the gradations.

5.2.3 Voice range constraints

The problem we consider is to write a music piece for four voices. Of course, these voices cannot sing all existing notes. Theoretically, a bass can sing notes presented by integers from 5 to 26, yet there exist very good basses that can not reach the very low note presented by 5 or the very high note presented by 26. It is not wrong to write these extreme notes but it should be avoided, because even if the bass can sing these notes, they will not sound as good as the notes in the middle of their register. To express this mathematically, we introduce new valued constraints R_a, R_b and R_t as follows: for every note in the reach of a voice, R_b, R_t and R_a express how bad this note is for bass, tenor or alto respectively. Fig. 2 presents the values of R_a, R_b and R_t. The total cost is given by $(R_a + R_b + R_t)/3 \in [0, 1]$.

5.2.4 Distance constraints

In tonal music, it is important to have fluent melodies. Modeling this requires counterpoint techniques, which we do not consider here. Instead, we model the soundness of the melodies by considering the distances between the notes. Furthermore, it is hard to sing large intervals correctly. Therefore, we introduce a valued constraint D that limits the distance between two notes. We assign cost 1 to all intervals larger than an octave (12 half tones). Intervals smaller than an octave get cost $c_4 = d/12$, proportional to the distance d between the notes. An exception is made for the octave, which gets the same cost as a fifth because it sounds equally well and is equally easy to sing. The costs are presented in Table 3.

5.2.5 Contrary motion constraints

Contrary motion is the general movement of two melodic lines in opposite directions. That is, when one of the lines moves up, the other line moves down. In tonal music, contrary motion is important to maintain independence of melodic movement. To express this as a valued constraint we assign a cost

Table 1: The mapping G that assigns a cost to every allowed succession of two grades

GRADE1-GRADE2	COST=G(GRADE1,GRADE2)
II-V, V-I, I-II, IV-V	0
VI-V, VI-II, I-IV, V-VI	0
VI-IV, V-I, I-VI	0
I-I, V-V, VI-I, VI-VI	0.3
IV-II, IV-IV, II-II	0.4
I-V	0.5
IV-I, II-I, II-VI	1.0

Table 2: valued constraint H compares the count of grades to the ideal percentages.

GRADE	IDEAL PERCENTAGES		
	OPTION 1: I, IV AND V	OPTION 2: I, IV, V AND VI	OPTION 3: I, II, IV, V, VI
I	0.394	0.263	0.263
II	-	-	0.075
IV	0.303	0.303	0.228
V	0.303	0.303	0.303
VI	-	0.131	0.131

ISBN: 978-989-95079-6-8
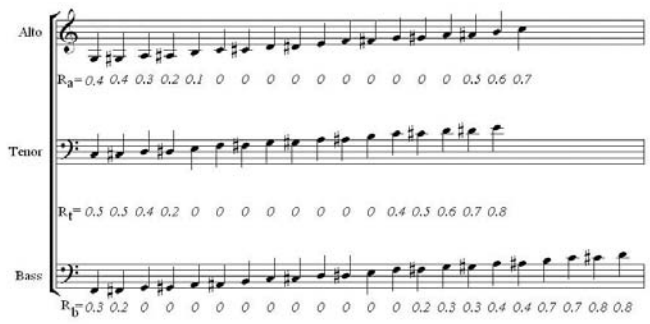


Figure 2: Costs that are incrurred when notes at the extreme ends of a voice's scale are used.

$c_5 = 1$ if there is no contrary motion between bass and soprano.

6 Solving the four voices harmonization VCSP

Now we translated the musical problem to a purely mathematical problem, we can design an algorithm that solves it. There exist good solvers for VCSPs [9] but we choose to design and implement a separate algorithm for the harmonization problem. The algorithm follows the thinking of a composer that solves the exercise: from left to right, try out all the possibilities and return when a solution gets stuck.

6.1 A backtracking algorithm

Algorithm 1 shows the outline of the main part of the algorithm. First note that it returns a single minimal-cost solution; if there are several such solutions, our algorithm returns the first one encountered. At each stage, max_cost and total_cost represent the cost of the optimal solution so far, and of the currently investigated solution, respectively. They are initialized in lines 1 and 2.

From Section 2, we know that the grade on the first note must be I. Either narrow or wide spacing can be used for the first chord; the former option is considered in lines 3–5, the second one in line 6–8.

The central part of the algorithm is in the recursive procedure Calculate_Solution. For every note i except the last one, all possible solutions are generated (line 2), that is, all permissible combinations of grades for note i, and corresponding notes for bass, tenor and alto. Assuming no hard constraints are violated (line 3), the cost of a solution w.r.t. the valued constraints is evaluated (line 4); all costs can be calculated directly, except from c_2 which is calculated considering the problem from the beginning until position i. c_2 is only calculated when $i > 4$. If the total cost for the current solution is smaller than the new costs added up to the total cost, one has to try the next possibility, because from here we will never find a solution that is cheaper than the solution we already found (line 5-6). Otherwise, one adds the costs except from c_2 up to the total cost and goes further looking for a solution (lines 8-9). c_2 is not added up to the total cost since this cost is calculated every step and it counts for the whole piece.

When we have reached the final note, we change the current solution and the maximum cost and we return in the recursion to find a cheaper solution (lines 16-17).

This can lead to a solution or to failure. In both cases we return to try the next posibilities. If there are no posibilities left, the algorithm does not go further in this branch of the recursion.

Table 3: valued constraint D is a measure for the soundness of an interval.

INTERVAL (IN HALF TONES)	COST c_4	INTERVAL (IN HALF TONES)	COST c_4
> 12	1.000	6	0.500
12	0.583	5	0.417
11	0.917	4	0.333
10	0.833	3	0.250
9	0.750	2	0.167
8	0.667	1	0.083
7	0.583	0	0.000

Algorithm 1: Algorithm that solves the problem of harmonizing a melody for a soprano, l is the length of the problem.

```
1  max_cost=Double.max;
2  total_cost=0;
3  Fill in the first chord, using narrow spacing;
4  total_cost=cost of this first chord (only w3c3);
5  Calculate_Solution(2,l);
6  Fill in the first chord, using wide spacing;
7  total_cost=cost of this first chord (only w3c3);
8  Calculate_Solution(2,l);
```

Procedure `Calculate_Solution`(**note position** *i*, **total length** *l*)

```
1  if (i < l) then
2      foreach possibility for note i do
3          if no hard constraints violated then
4              calculate costs ;
5              if total_cost+Σ_{i=1}^{5} w_i c_i > max_cost then
6                  go to next possibility;
7              else
8                  total_cost=total_cost+Σ_{i=1}^{5} w_i c_i − w_2 c_2;
9                  Calculate_Solution(i+1);
10             end
11         else
12             go to next possibility;
13         end
14     end
15 else
16     max_cost=total_cost+w_2 c_2;
17     save solution;
18 end
```

7 Results

In this section we present some test results. We tested the program for several problems but we will show only one representative example here.

We illustrate the suitability of our approach by showing four

ISBN: 978-989-95079-6-8
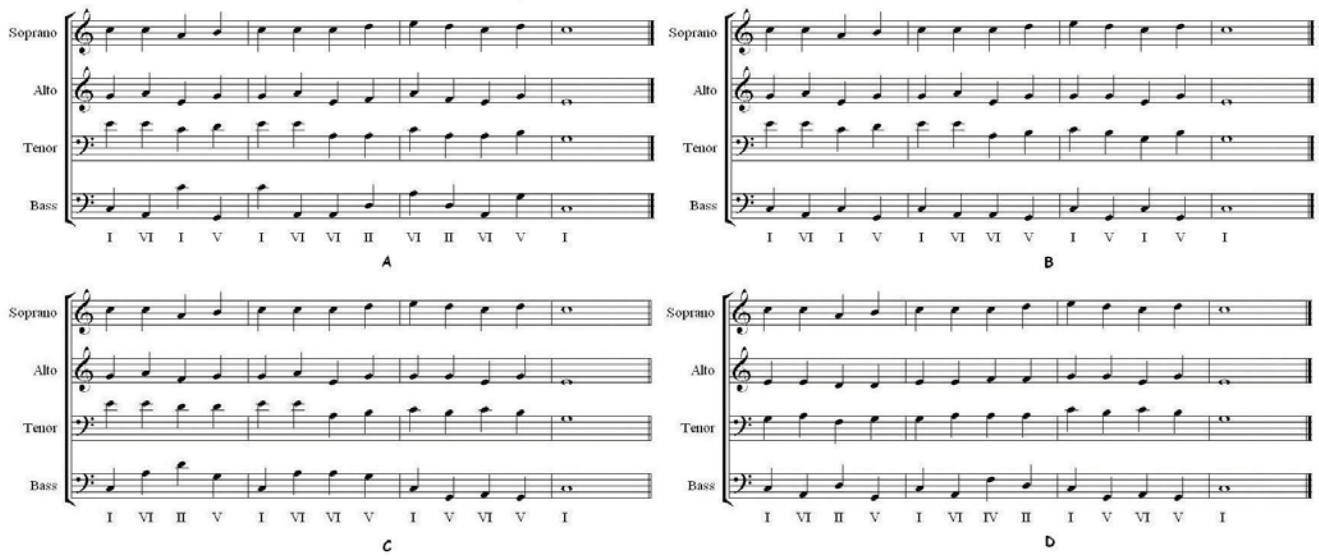


Figure 3: A representative harmonization exercise, showing four solutions: (A), (B), (C) show solutions with weights as proposed in Section 5. (A) has cost 3.14, (B) has cost 1.12 and (C) is the optimal solution with cost 0.65. (D) shows the optimal solution with weights $w_1 = 0.3, w_2 = 0.2, w_3 = 0.1, w_4 = 0.3$ and $w_5 = 0.1$. It has cost 0.47.

solutions with different costs. In Fig. 3A we show the solution with the highest cost (3.14). In this solution there are unacceptable distances in the bass. Furthermore, the melodies in the alto and tenor are not fluent, there is almost no contrary motion between soprano and bass and the grade sequence is bad.

Fig. 3B shows a solution with cost 1.12. This solution has a better bass line. The grade sequence is slightly better since we now have tension (grade V) in the second bar. This solution sometimes lacks contrary movement and there is no second grade used.

The optimal solution is shown in Fig. 3C. The grade sequence and grade frequencies are very well but there are many large distances between the bass notes. This can be explained by the weights we assigned to the costs: contrary motion is considered equally important as the distances between the notes. If we raise the weights corresponding to the distances between notes and we lower the weight corresponding to the use of contrary movement we obtain the result in Fig. 3D. Except from the fact that there is no tension in the second bar this is a good solution.

8 Practical Use And Further Work

The algorithm we presented is designed to solve harmonizing problems. It can be used as a didactic tool: teachers can use it to test quickly if problems have solutions, if these solutions are interesting enough and if they are not too difficult to solve. For instance, if a problem has only one solution where only grades I and IV are used, this problem cannot be solved well. Furthermore, students can use this program to compare their solutions to an optimal solution.

The model we presented can be extended in various ways. First of all, we can overcome the simplification of the harmonization problem by allowing modulations, chord inversions, non-chord tones, rhythm and ornaments, seventh chords and so on. This requires an extension of the domains and the introduction of new hard constraints. Furthermore, we can also introduce more soft constraints. For instance, some counterpoint rules can be expressed as soft constraints.

As illustrated in the previous section, the assignment of the weights is important. Fine-tuning of these parameters is needed and should be done more profoundly.

Acknowledgment

Chris Cornelis and Yvan Saeys thank the Research–Foundation Flanders for funding their research.

References

[1] C. Truchet and P. Codognet. Musical constraint satisfaction problems solved with adaptive search. *Soft Comput.*, 8(9):633–640, 2004.

[2] François Pachet and Pierre Roy. Musical harmonization with constraints: A survey. *Constraints*, 6(1):7–19, 2001.

[3] Kemal Ebcioğlu. An expert system for harmonizing chorales in the style of J. S. Bach. pages 294–333, 1992.

[4] Stéphanie Vanhove. Backtrackalgoritmen voor regelgebaseerde vierstemmige harmonisatie. Master's thesis, Ghent University, 2007.

[5] Somnuk Phon-amnuaisuk, Andrew Tuson, and Geraint Wiggins. *Evolving Musical Harmonisation*. 1999.

[6] Somnuk Phon-amnuaisuk and Geraint A. Wiggins. The four-part harmonisation problem: a comparison between genetic algorithms and a rule-based system. In *Proceedings of the AISB99 Symposium on Musical Creativity*, pages 28–34. AISB, 1999.

[7] Allan Forte. *Tonal Harmony in Concept and Practice*. 1962.

[8] Imogen Holst. *An ABC of music: a short practical guide to the basic essentials of rudiments, harmony, and form. With a foreword by Benjamin Britten*. 1963.

[9] Pedro Meseguer, Noureddine Bouhmala, Taoufik Bouzoubaa, Morten Irgens, and Martí Sánchez. Current approaches for solving over-constrained problems. *Constraints*, 8:9–39, 2003.

ISBN: 978-989-95079-6-8

Detecting and Reacting on Drifts and Shifts in On-Line Data Streams with Evolving Fuzzy Systems *

Edwin Lughofer[1] Plamen Angelov[2]

1. Department of Knowledge-Based Mathematical Systems, Johannes Kepler University
Altenbergerstrasse 69, A-4040 Linz, Austria
2. Department of Communication Systems, Infolab 21, Lancaster University
Lancaster, LA1 4WA, UK
Email: edwin.lughofer@jku.at, p.angelov@lancaster.ac.uk

Abstract— *In this paper, we present new approaches to handle* drift *and* shift *in on-line data streams using evolving fuzzy systems (EFS), which are characterized by the fact that their structure is not fixed and not pre-determined. When dealing with* drifts *and* shifts *in data streams one needs to take into account two major issues: a) automatic detection of, and b) automatic reaction to this. To address the first problem we propose an approach based on the concepts of* age *and* utility *of fuzzy rules/clusters. The second problem itself is composed of two sub-problems concerning the influence of the* drifts *and* shifts *on: 1) the antecedent parts (fuzzy set and rule structure) and 2) the consequent parts (parameters) of the fuzzy models. To address the latter sub-problem we propose an approach that introduces a gradual forgetting strategy in the local learning process. To address the former sub-problem we introduce two alternative methods: one that is based on the evolving density-based clustering, eClustering (used in* eTS*); and one that is based on the automatic adaptation of the learning rate of the evolving vector quantization approach (eVQ) (used in* FLEXFIS*). The paper is concluded with an empirical evaluation of the impact of the proposed approaches in (on-line) real-world data sets where* drifts *and* shifts *occur.*

Keywords— drifts and shifts in data streams, evolving fuzzy systems, detection and reaction to drifts and shifts, age of a cluster/fuzzy rule, gradual forgetting

1 Introduction

1.1 Motivation and State of the Art

Nowadays data-driven fuzzy systems enjoy a great attraction in many industrial applications, as opposed to expert-based fuzzy systems. They can be automatically generated from process data such as measurements, images (features) or signal streams. Furthermore, they are proven universal approximators [27], i.e. being able to model a given problem to (theoretically) any degree of accuracy. Moreover, they also allow an insight in the form of linguistically and visually interpretable rules to be gained [8].

During the last decade, the research field of 'evolving fuzzy systems' (EFS) emerged as an important part of the fuzzy systems research [4] [11], as they are capable to include new information on demand into the models and on-the-fly without necessarily using prior knowledge. EFS can also permanently learn from their environment and are applicable in fast on-line identification [4] and modelling processes as well as huge data bases which cannot be loaded into the virtual memory at once [18]. Often, there are no enough data in advance (off-line) to build reliable models with high predictive quality which can also require application of EFS as in [23].

Various approaches for EFS have been set up during the last years, one of the most popular and pioneering approach is the *eTS* family which comes with a regression [4], [3] and a classification variant [5]. Another approach for adaptation and evolution of clusters inspired by the evolving vector quantization (eVQ) [19] is the so-called *FLEXFIS* family [21], in particular *FLEXFIS* for regression [20] and for classification [22] (denoted as *FLEXFIS-Class*). A range of other alternative approaches includes ePL [16], SAFIS [12], and evolving fuzzy neural networks such as SOFNN [15] or GDFNN [28].

All these methods have a common denominator: they all are life-long learning approaches, which means that they incorporate all the data samples into the fuzzy models with equal weights in the same order as they are coming in during the on-line process. Hence, fuzzy models reflect a compact information of all the samples seen so far with equal importance. This is a beneficial to adapt the models, especially when a convergence to an optimality criterion or stable state of the model structure is achievable [20]. However, this benefit is only true in case of data streams which are generated from the same underlying data distribution respectively which do not show any *drift* or *shift* to other parts of the input/output space [26]. *Drift* (respectively *shift*) indicate the necessity of out-dating of previously learned relationships (in terms of structure and parameters) as these are not valid any longer and hence should be eliminated from the model. In order to cope with this problem, *drift* handling was already applied in other machine learning techniques, e.g. in connection with SVMs [14], ensemble classifiers [24] or instance-based (lazy) learning approaches [7] but to the best of our knowledge this concept was not yet applied to fuzzy systems.

*This work was partially supported by the Upper Austrian Technology and Research Promotion and by The Royal Academy, UK. This publication reflects only the authors' views.

ISBN: 978-989-95079-6-8

1.2 Our Approach

Hence, in this paper we are dealing with approaches for an appropriate handling of *drifts* and *shifts* by EFS. However, parts of the concepts, especially the detection of, and reaction to *drifts* in the consequent parts of the rules, can be applied to a wider range of EFS approaches including *eTS* and *FLEXFIS*. A more detailed description will be given in Section 2. For achieving an automatic approach, we propose 1.) the detection of a *drift* respectively *shift* based on the concepts of fuzzy rule *age* and *utility* function, and 2.) the reaction to a detected *drift* respectively *shift*, both in rule antecedent (for *eClustering* as applied in *eTS* and *eVQ* as applied in *FLEXFIS*) as well as consequent parts of Takagi-Sugeno fuzzy models [25].

The paper is concluded with an empirical evaluation of the impact of the proposed approaches on predictive accuracy of the evolving fuzzy models when applying to real-world data streams where *drifts* and *shifts* occur (Section 5). Note, that no benchmark data from the Internet or well-known data bases can be applied for empirical evaluation as these are usually all smooth in the sense that no *drift* occurs therein.

2 Problem Statement

In machine learning literature, they distinguish between different types of 'concept change' of the underlying distribution of (on-line) data streams: a) *drifts*, and b) *shifts*, see [26]. *Drift* refers to a *gradual* evolution of the concept over time. The concept *drift* concerns the way the data distribution slides smoothly through the data/feature space from one region to another. For instance, one may consider a data cluster moving from one position to another. This concept is closely related to the time-space representation of the data streams. While the concept of (data) density is represented in the data space domain, *drift* and *shift* are concepts in the joint data-time space domain.

An (artificial) example of a *drift* is demonstrated in Figure 1. There, the original data distribution (in a 2-D data space) is marked by circular samples, which changes over time into a data distribution marked by rectangular samples. If a conventional clustering process is applied by weighting all new incoming samples equally, the cluster center would end up exactly in the middle of the whole data bunch, averaging old and new data. Such a clustering approach is applied in EFS approaches to identify the local regions that can be used to form (the antecedent parts of the) fuzzy rules.

On the other hand, the concept *shift* refers to a sudden, abrupt change of the underlying concept to be learned. A *shift* in the input space opens up a data cloud in an unexplored region and hence usually automatically causes a new rule to be evolved by the incremental/evolving clustering algorithm. In Figure 2 a case of a *shift* in the output variable is shown: the original trajectory (consisting of dense data samples) on the right-hand side is marked with a light line, whereas the *shift* is represented by the dark data samples forming a trajectory above the other one. In case if only apply the usual adaptation

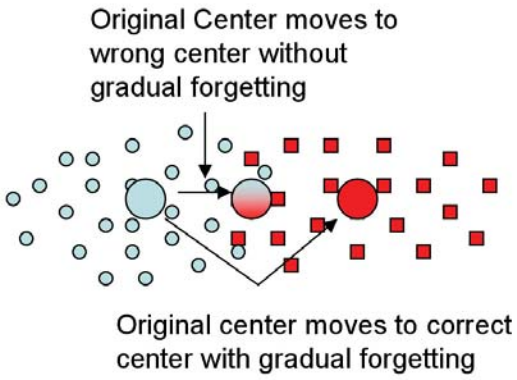


Figure 1: A *drift* in an evolving cluster (used for learning the antecedent parts of EFS), the distribution before the drift shown in circular, after the drift shown in rectangular samples.

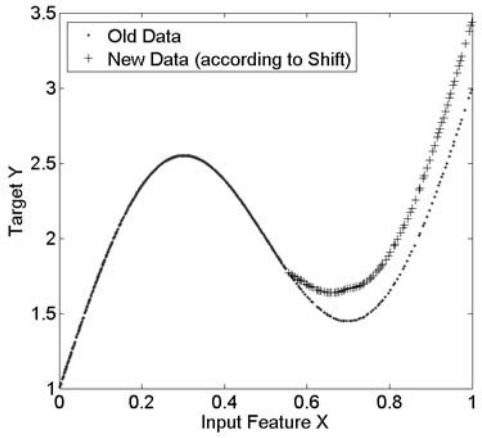


Figure 2: Example of a *shift* in the output variable; compare light dots (original data distribution) with dark dots (data distribution after the *shift*) on the right-hand side of the image (also marked by arrows)

of the consequent parameters with weighted recursive least squares without forgetting the old data due to the *shift* the approximation curve of the fuzzy model will end up exactly in-between these two.

In this paper, we demonstrate novel approaches for autonomous *drift* and *shift* detection and handling when learning fuzzy rule-based systems of Takagi-Sugeno type from on-line data streams in an evolving manner [4]. We focus on EFS approaches exploiting the Takagi-Sugeno model architecture [25].

3 Autonomous Detection of Drifts and Shifts in Data Streams by EFS

In this section the method for autonomous detection of *shifts* and *drifts* in data streams based on the *age* and *utility* of the cluster/fuzzy rule [1], [10] is described. This is an important step in the process of handling non-stationarity in data streams and for building au-

ISBN: 978-989-95079-6-8

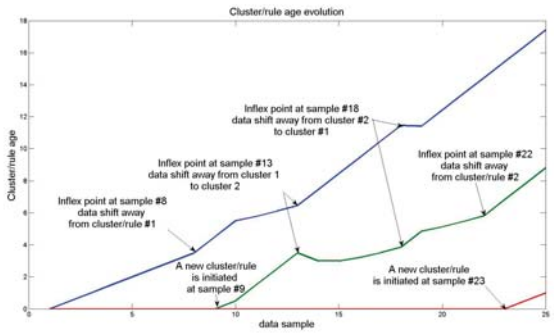


Figure 3: Evolution of the *age* of the clusters/fuzzy rules; *drifts* are related to the inflexion points; *shifts* are related to forming new clusters/fuzzy rules.

tonomous self-developing, self-learning, and evolving models and systems.

3.1 Detecting Drifts by Age of Clusters (Rules) [1]

In [1], the concept of cluster (respectively fuzzy rule) *age* was introduced. We extend this by including the membership degrees of the rules ($\Phi_{i,l}$ in case of the ith rule for the lth sample)

$$age^i = k - \frac{\sum_{l=1}^{n_i} I_l \Phi_{i,l}}{n_i} \quad (1)$$

where i is the rule index; n_i denotes the support of rule i; I_j denotes the time instance when the data sample was read; k is the current time instance. Since $\frac{\sum_{l=1}^{n_i} I_l}{n_i}$ can also be accumulated recursively, the *age* can be easily calculated when necessary. *Age* of the cluster/rule changes with each new data sample being read. If the newly read data sample does not fall into that cluster (supports that fuzzy rule with a low or even 0 value of $\Phi_{i,l}$ in case of fuzzy sets with finite support) the *age* grows by the rate of one sample at a time. That means, if a certain cluster (fuzzy rule) is not supported by any future samples after being initiated (say in the time instant, t_i) then its *age* at any time instant, k will be $k - t_i$. However, if any new data sample (between t_i and k) falls into that cluster (supports that fuzzy rule with a high value of $\Phi_{i,l}$), the *age* does not grow with the same rate, but with a smaller one and one can say that the cluster (fuzzy rule) is being *refreshed*.

The fuzzy rule *age* is important and closely linked to the data streams (which are sequences of data in time) and to the concept *drift*. We propose to analyze the *age* of clusters (respectively, fuzzy rules) on-line by using the gradient of the *ageing* curve as well as its second derivative which indicate a change in the slope of the *ageing* curve. When there is a **significant** change of the *ageing* which results in a significant change of the slope, then obviously the second derivative of the *age* curve will be indicative of this inflection points. An example is demonstrated in Figure 3, where the *drift* and *shift* are clearly marked on the *age* evolution curves.

3.2 Detecting Shifts by Utility Function

Shift in the data streams is a more significant, abrupt and sudden change of the data concept. Therefore, the distinction between *drift* and *shift* is the pace, the degree and speed of the changes, while both indicate a change in the data distributions with time. Essentially, the reaction to a detected *shift* has to be reflected in the structure evolution of the fuzzy rule base. While *drift* is related more to a smooth forgetting of parameters, learning and forgetting, *shift* is closely related to changes in the structural components.

Shift can be detected by the *utility* of the fuzzy rule. *Utility* is a parameter of the quality of the fuzzy rule that is defined [2] as the accumulated firing level of each rule given by Ψ_i summed over the life of each rule:

$$U_k^i = \frac{1}{k - I_k^i} \sum_{l=1}^{I_k^i} \Psi^l \quad (2)$$

$Utility$, U accumulates the weight of the rule contributions to the overall output during the life of the rule (from the current time instant back to the moment when this rule was generated). It is a measure of importance of the respective fuzzy rule comparing to the other rules (comparison is hidden in the relative nature of Ψ).

If the *utility* of a fuzzy rule is *low* then this rule becomes obsolete or not used very much. A *shift* away from or to a cluster/fuzzy rule can be detected by the derivative of the *utility* — if the value of the first derivative of the *utility* is *large* then there is a *shift* towards the cluster. The *shift* is away from a cluster if the derivative of the *utility* is *large* negative. Then this fuzzy rule is deemed obsolete (does not contribute significantly to the overall prediction/estimation/classification) and can be removed from the rule base [10].

4 Reactions to Drifts and Shifts in Data Streams by EFS

4.1 Reaction to Drift and Shift in the Antecedents

In this section, we introduce methods for addressing both *drifts* and *shifts* in data streams by adapting learning mechanisms for consequent parts and by evolving antecedent part. We demonstrate this on the example of the popular *eTS* method [4] and on *FLEXFIS* [20].

4.1.1 *By* eClustering

Reaction to a detected *shift* is by either a) forming a new rule around a new data sample which becomes an attraction point for the global data distribution, or b) replacement of a fuzzy rule which itself consists of; i) forming a new rule around the new point, and ii) removal of the rule which has lower density and is close to this newly added one [2] — see Figure 4. If locally optimal learning is being applied then removing a cluster and respectively a locally valid Kalman filter/RLS does not affect the overall learning significantly (only through the fuzzy weight Ψ [4]). If globally optimal learning is being applied, removal of a cluster (respectively fuzzy rules) does affect n columns and n rows of the covariance matrix directly and the remaining values of the covariance

ISBN: 978-989-95079-6-8

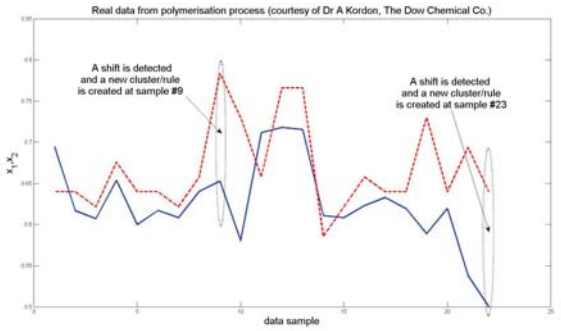


Figure 4: Shift in the time domain representation and two new rules evolved as a reaction to it.

matrix indirectly [4]. Deleting *old* fuzzy rules was applied in classification in [5] and in prediction [2].

4.1.2 *By* eVQ

When *shifts* in the data stream occur, usually new clusters are evolved automatically in *eVQ*, as they cause new data clouds in previously unexplored regions of the data space (in regions further away than a fraction of the space diagonal in the input space to which the vigilance parameter ρ is usually set, i.e. 0.1 to 0.3 of the space diagonal, see [19]. Synchronously, older clusters can be deleted by the concept of *utility* [9] [10] as discussed in Section 3.2. This does not affect the learning of consequent parameters for the other rules if local learning is applied.

For reacting to *drifts* in the antecedent parts of the rules we propose to re-adjust the parameter in the eVQ clustering algorithm η [19] which steers the learning gain. We define the tracking concepts for the ith rule throughout this section, which can be generalized to any rule in a straightforward manner. Currently, in *eVQ* the learning gain is defined by the following formula [18]:

$$\eta_i = \frac{0.5}{n_i} \tag{3}$$

If a *drift* occurs in a data stream, the centers and widths of the cluster(s) should change to the new data distribution (as shown in Figure 1). For re-activating the converged clusters, i.e. re-animating them for stronger movements in a drifting case, we suggest a sudden increase for the first sample in the *drift* phase, followed by a gradual decrease for the next samples in order to balance in the new sample distribution in the same manner as is done for original ones.

Here, we propose the following mechanisms for the learning gain η: first we transform the forgetting factor λ, used in the gradual out-weighting when doing consequent learning (see next section) and denoting the intensity of a *drift*, to a value in $[0, 1]$. Hereby, 0.9 (minimal value for λ) is mapped to 0.99, whereas 1 is mapped to 0, hence:

$$\lambda_trans = -9.9\lambda + 9.9 \tag{4}$$

Then, when a *drift* occurs, we re-set the number of samples forming the ith cluster (n_i) (used in the denominator of the calculation of η_i) by

$$n_i = n_i - n_i * \lambda_trans \tag{5}$$

This means that the stronger the drift is, the more n_i is decreased and hence the stronger the forgetting effect will be. In Figure 5 it is demonstrated how η_i develops (lines) in usual (non-drift) case (for the first 100 samples), then a drifting scenario is artificially caused with three intensities leading to the three λ values (in different line styles). After the *drift* indicator (at sample 100), it is decreased in usual way such that the jumped center can converge to the new data distribution.

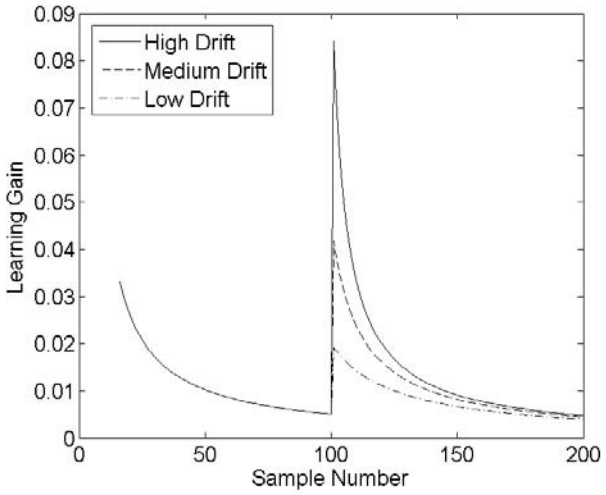


Figure 5: Abrupt increasing learning gain η during in case of a *drift* (after 100 samples) when applying different values for the forgetting factor λ.

4.2 *Reaction to Drifts and Shifts in the Consequents*

For the rule consequents *drifts* and *shifts* can be handled in one sweep as it is just a matter of the setting of the forgetting factor as we will see below. Whenever a *drift* in the output variable occurs (as shown in Figure 2) and is detected, it is necessary to apply a specific mechanism in the sample-wise incremental learning steps of the consequent parameters in Takagi-Sugeno fuzzy systems. When locally optimal fuzzily weighted Recursive Least Squares (wRLS) learning approach [4] is used this can easily be accommodated as detailed later on.

If we do not take *drift* into account all newer samples lying in the same local positions relative to the rules as the older samples are included with the same rule weights in the update process. Then the fuzzy model will end up with approximation curve as shown in the left image of Figure 6 with dotted lines. Obviously, the approximation ends up in the middle of the two trajectories (the newer after the *drift* shown in light font, the older shown in darker font), as trying to minimize the quadratic errors (least squares) of all samples to the curve. Hence, it is necessary to include a parameter in the update process, which forces older samples to be out-dated over time. Gradualism is important here in

ISBN: 978-989-95079-6-8

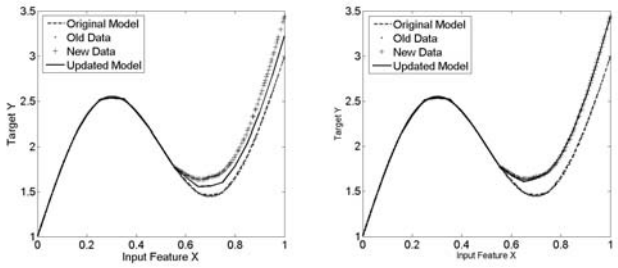


Figure 6: Left: The adapted model (dotted line) with the new incoming samples (dark dots) when applying conventional recursive weighted least squares approach; right: the adapted model (dotted line) by including a forgetting factor of 0.95 and using (7)-(9), the approximation surface lying exactly on the trajectory of the data samples denoting the novel data distribution

order to guarantee a smooth forgetting and to prevent abrupt changes in the approximation surface. For doing so, we re-define the least squares optimization function for the ith rule by

$$J_i = \sum_{k=1}^{N} \lambda^{N-k} \Psi_i(\vec{x}(k)) e_i^2(k) \longrightarrow \min_w \quad (6)$$

with $e_i(k) = y(k) - \hat{y}(k)$ the error of the ith rule in sample k. Assuming N the number of samples loaded so far, this function out-dates the sample processed i steps ago by λ^{N-k}. Usual values of λ lie between 0.9 and 1, where a value near 1 means a slow forgetting and a value near 0.9 a fast forgetting of former loaded data samples and the exact choice depends strongly on the strength of the *drift* (see below). Following a similar recursive deduction scheme as in conventional recursive least squares (RLS) [17], we obtain the following incremental update formulas for consequent parameters $\vec{w}_i$ for the ith rule:

$$\hat{\vec{w}}_i(k+1) = \hat{\vec{w}}_i(k) + \gamma(k)(y(k+1) - \vec{r}^T(k+1)\hat{\vec{w}}_i(k)) \quad (7)$$

$$\gamma(k) = \frac{P_i(k)\vec{r}(k+1)}{\frac{\lambda}{\Psi_i(\vec{x}(k+1))} + \vec{r}^T(k+1)P_i(k)\vec{r}(k+1)} \quad (8)$$

$$P_i(k+1) = (I - \gamma(k)\vec{r}^T(k+1))P_i(k)\frac{1}{\lambda} \quad (9)$$

with $P_i(k) = (R_i(k)^T Q_i(k) R_i(k))^{-1}$ the covariance matrix and $\vec{r}(k+1) = [1 \;\; x_1(k+1) \;\; x_2(k+1) \;\; \ldots \;\; x_p(k+1)]^T$ the regressor values of the $k+1$th data sample, Ψ_i the fulfillment degree of the ith rule, serving as weight in the recursive least squares algorithm.

The final question is how to set the parameter λ in order to guarantee an appropriate *drift* tracking. We propose a strategy to deduce it directly from the *age* curves analysis [10] since they are indicative of the speed of a *drift* (see Section 3). In Section 3, it was mentioned that the age of a rule always lies in $[0, k]$. Hence, we normalize the *age* of the ith rule to $[0, 1]$ by $age_i_norm = \frac{age_i}{k}$ in order to achieve gradients of the normalized rule ages Δage_i_norm also lying in $[0, 1]$. Whenever the change of the gradient in the rule *age* curve is significant, recursive weighted RLS with forgetting (wRLSf) should be triggered. We use the following estimation for λ:

$$\lambda = 1 - 0.1\Delta^2 age_i_norm \quad (10)$$

This guarantees a λ between 0.9 (strong forgetting) and 1 (no forgetting), according to the degree of the gradient change (1 = maximal change, 0 = no change). The forgetting factor is then kept for a while at this level (otherwise only one single sample would cause a gradual forgetting) and set back to 1, after a stable gradient phase is achieved (usually after around 20 to 30 samples showing a moderate value of $\Delta^2 age_i_norm$). Setting back to 1 is necessary, as otherwise the forgetting will go on inside the new data distribution. This cause the drift phase in the antecedents to stop.

5 Evaluation

This section deals with the evaluation of the impact of reacting on *drifts* and *shifts* in case of data streams where actually *drifts* and *shifts* occur. This is done by implementing the approaches discussed throughout this paper. One application example is coming from a rolling mill, where the task was to identify a prediction model on-line for the resistance value of a steel plate at a rolling mill. The other application concerns a polymerization process in chemical industry.

5.1 On-Line Prediction Models at Rolling Mills

The task was to identify a prediction model for the resistance value of a steel plate at a rolling mill. This should be done in a first step with some off-line (precollected) measurement data in order to obtain a feeling about the achieved quality of the fuzzy model and then to refine the prediction model with newly recorded online data. The later step was possible as first a prediction for the resistance is given, influencing the whole process at the rolling mill, whereas a few seconds later (after the steel plate is passed), the real value for the resistance is measured, which can then be incorporated into the model adaptation process. In this sense, the correct measured value not the predicted (which might have been wrong)one is taken for learning. In fact, an improvement in terms of predictive power could be achieved when updating the fuzzy models trained in batch mode with 6000 samples during on-line operation mode with further 6600 samples. For details of the experimental setup and results see [20] (and also Table 1 below summarizing all the results).

Now, in this paper we want to examine whether a reaction onto drifts by a gradual forgetting of older samples during the on-line process may further improve the quality of the models. A justification of an application of reaction onto *drifts* is that the operation process at rolling mills is divided into different "stitches". One stitch represents one closed cycle in the rolling process. In the on-line mode the measurements come in continuously from stitch to stitch. However, for the current processed stitch, the previous stitch should play only little or even no role. However, the measurements from the

ISBN: 978-989-95079-6-8

Table 1: Comparison of prediction with and without applying gradual forgetting

Method	**MAE**	**Max MAE Too High / Max MAE Too Low / # MAEs > 20**
Analytical	7.84	63.47 / 87.37 / 259
Static fuzzy models	6.76	45.05 / 81.65 / 176
FLEXFIS	5.41	38.05 / 78.88 / 159
FLEXFIS with forg.	4.65	31.99 / 74 / 68

previous stitch are already included in the fuzzy models as updated by their samples. Thus, this means that older samples from the previous stitch should be forgotten when including samples from the current stitch. Another aspect is that here we do not need any drift/shift detection, as the drift/shift is indicated by the beginning of a new stitch. As no *drift* detection was carried out, λ was set to 0.97, which is a good compromise between fast forgetting (=strong locality of models) and low forgetting (=weak locality of models). The results are demonstrated in Table 1. Here, we also demonstrate the improvement of the predictive accuracy over analytical models by both, static and evolving fuzzy models. Three different types of errors are reported: the mean absolute error over all on-line samples (note that first a prediction is done and then the model updated with the same samples and based on feedback), the number of mean absolute error greater than 20, the maximal mean absolute error over all samples where the prediction was too low and the maximal mean absolute error over all samples where the prediction was too high. The latter value is the most important one as harming the steel plate is more dangerous as in case of predicting too low values. The results (Table 1) demonstrate the impact of the gradual forgetting.

Another interesting aspect is that the error on the single measurements starts to *drift* over time when gradual forgetting is not applied. This is underlined in the left image of Figure 7 which shows the single errors over the 6600 on-line samples: note the *drift* of the main error area away from the zero line at the end of the data stream.

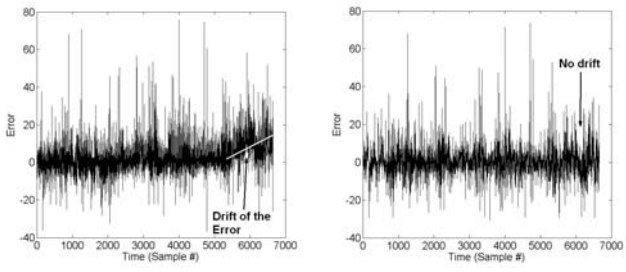


Figure 7: Left: The error curve for the 6600 on-line samples when no forgetting is applied: at the end the error starts drifting; right: no *drift*.

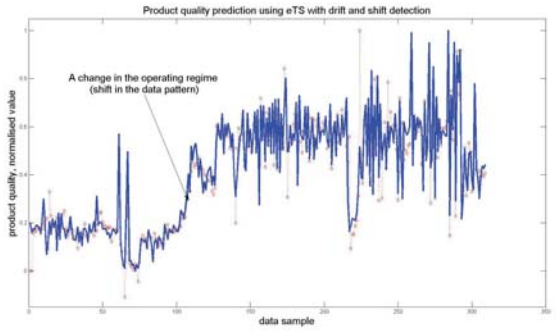


Figure 8: A comparison of the predicted and real data of product quality. Solid line - predictions by *eTS* with *shift* and *drift* detection; diamonds - real data.

5.2 *Another application example*

A case study based on real data (courtesy of Dr. Arthur Kordon, The Dow Chemical Co.) from the chemical industry is used as an illustration of the detection and reaction to *drift* and *shift* in *eTS* [6]. The *eTS* has been applied for prediction of the properties of a chemical composition by modelling the product composition in a distillation tower. The data set includes a change of the operating regime of the process which brings a challenge to the structure of the model (fuzzy rule based system). The data set includes also a number of other challenges, such as noise in the data, a large number of initial variables, etc. These problems cover a wide range of real issues in the industry. The process data is retrieved from physical ('hard') sensors used as inputs to the *eTS* applying hourly averages for every eight hours. The product composition (real output) is estimated by a laboratory analysis for comparison (it is given with diamonds in the Figure 8).

The estimation of the product composition contains noise due to the nature of the analysis. A significant operating condition change takes place after sample 127 (please see the Figure 8). The *eTS* was able to efficiently detect and react to this *shift* as well as to the *drifts* as depicted in Figures 3 and 4. The overall prediction is very good and compares favorably with the conventional models as detailed in the Table 2. Note, that the non-dimensional error index (NDEI) is defined as the ratio of the root mean square error over the standard deviation of the target data and should be ideally 0 while the variance accounted for (VAF) is defined as the ratio between the variance of the real data and the model output and is given out of a maximum of 100 (when the predictions coincide with the real data).

6 Conclusion

In this paper, we propose novel strategies and techniques for addressing concept *drift* and *shift* in on-line data streams. Therefore, two EFS approaches (*eTS* and *FLEXFIS*) are exploited as on-line modelling methodologies. These are extended by mechanisms which are 1.) able to detect *drifts* and *shifts* with fuzzy rule *age* and *util-*

ISBN: 978-989-95079-6-8

Table 2: Error measures when applying *eTS* with (Column #3) and without (Column #2) drift detection and reaction in chemical composition modelling

Measure	Without	With	Best (theor.) value
NDEI	0.3559	0.33465	0
VAF, %	87.319	88.807	100
correlation	0.9357	0.94285	1

ity, and 2.) to react on such occurrences appropriately. The latter is applied 1.) to the rule antecedent parts directly in the cluster space for reacting on *drifts* and *shifts* in the input space and 2.) to rule consequent parameters for reacting on *drifts* and *shifts* in the output variable, applicable to any EFS technique exploiting Takagi-Sugeno type fuzzy systems. Evaluation on real-world data sets shows that the novel techniques are able to improve the accuracy and stability of the fuzzy models, whenever the occurrence of *drifts* and *shifts* is present.

References

[1] P. Angelov and D. Filev. Simpl_eTS: A simplified method for learning evolving Takagi-Sugeno fuzzy models. In *Proceedings of FUZZ-IEEE 2005*, pages 1068–1073, Reno, Nevada, U.S.A., 2005.

[2] P. Angelov and X. Zhou. Evolving fuzzy-rule-based classifiers from data streams. *IEEE Transactions on Fuzzy Systems*, 16(6):1462–1475, 2008.

[3] P.P. Angelov. Evolving Takagi-Sugeno fuzzy systems from streaming data, eTS+. In P. Angelov, D. Filev, and N. Kasabov, editors, *Evolving Intelligent Systems: Methodology and Applications*, page to appear. John Wiley & Sons, New York, 2008.

[4] P.P. Angelov and D. Filev. An approach to online identification of Takagi-Sugeno fuzzy models. *IEEE Trans. on Systems, Man and Cybernetics, part B*, 34(1):484–498, 2004.

[5] P.P. Angelov, E. Lughofer, and X. Zhou. Evolving fuzzy classifiers using different model architectures. *Fuzzy Sets and Systems*, 159(23):3160–3182, 2008.

[6] P.P. Angelov and X.-W. Zhou. Evolving fuzzy systems from data streams in real-time. In *2006 International Symposium on Evolving Fuzzy Systems*, pages 29–35, 2006.

[7] J. Beringer and E. Hüllermeier. Efficient instance-based learning on data streams. *Intelligent Data Analysis*, 11(6):627–650, 2007.

[8] F. Hopner and F. Klawonn. Obtaining interpretable fuzzy models from fuzzy clustering and fuzzy regression. *Proc. 4th Intern. Conf. on Knowledge-based Intell. Eng. Syst. (KES)*, Brighton, UK, pages pp.162–165, 2000.

[9] P. Angelov. *Machine Learning (Collaborative Systems)*, patent (WO2008053161, priority date: 1 November 2006).

[10] P. Angelov and X.-W. Zhou. On Line Learning Fuzzy Rule-based System Structure from Data Streams. In *2008 IEEE International Conference on Fuzzy Systems within the IEEE World Congress on Computational Intelligence*, Hong Kong, June 1-6, 2008, pages 915–922, 2008.

[11] P.P. Angelov *Evolving Rule-based Models: A Tool for Flexible Systems Design*, Springer Physica Verlag, Heidelberg, Germany, February, 2002.

[12] N. Sundararajan H.-J. Rong, G.-B. Huang, and P. Saratchandran. Sequential adaptive fuzzy inference system (SAFIS) for nonlinear system identification and prediction. *Fuzzy Sets and Syst.*, 157(9):1260–1275, 2006.

[13] N. K. Kasabov and Q. Song. DENFIS: Dynamic evolving neural-fuzzy inference system and its application for time-series prediction. *IEEE Trans. on Fuzzy Systems*, 10(2):144–154, 2002.

[14] R. Klinkenberg. Learning drifting concepts: Example selection vs. example weighting. *Intelligent Data Analysis*, vol. 8 (3), pages 281–300, 2004.

[15] G. Leng, T.M McGinnity, and G. Prasad. An approach for on-line extraction of fuzzy rules using a self-organising fuzzy neural network. *Fuzzy Sets and Syst.*, 150(2):211–243, 2005.

[16] E. Lima, F. Gomide, and R. Ballini. Participatory Evolving Fuzzy Modeling. In P. Angelov, D. Filev, and N. Kasabov, editors, *2006 International Symposium on Evolving Fuzzy Systems*, IEEE Press, pages 36–41, 2006.

[17] L. Ljung. *System Identification: Theory for the User*. Prentice Hall PTR, Prentic Hall Inc., Upper Saddle River, New Jersey 07458, 1999.

[18] E. Lughofer. *Evolving Fuzzy Models — Incremental Learning, Interpretability and Stability Issues, Applications*. VDM Verlag Dr. Müller, Saarbrücken, 2008.

[19] E. Lughofer. Extensions of vector quantization for incremental clustering. *Pattern Recognition*, 41(3):995–1011, 2008.

[20] E. Lughofer. FLEXFIS: A robust incremental learning approach for evolving TS fuzzy models. *IEEE Trans. on Fuzzy Systems (sp. issue on EFS)*, 16(6):1393–1410, 2008.

[21] E. Lughofer. Towards robust evolving fuzzy systems. In P. Angelov, D. Filev, and N. Kasabov, editors, *Evolving Intelligent Systems: Methodology and Applications*, to appear. John Wiley & Sons, New York, 2009.

[22] E. Lughofer, P. Angelov, and X. Zhou. Evolving single- and multi-model fuzzy classifiers with FLEXFIS-Class. In *Proc. FUZZ-IEEE 2007*, pages 363–368, London, UK, 2007.

[23] E. Lughofer, J. E. Smith, M. A. Tahir, P. Caleb-Solly, C. Eitzinger, D. Sannen, and M. Nuttin. Human-Machine Interaction Issues in Quality Control Based on On-Line Image Classification. *IEEE Trans. on Systems, Man and Cybernetics part A: Systems and Humans*, 2009, to appear.

[24] S. Ramamurthy and R. Bhatnagar. Tracking recurrent concept drift in streaming data using ensemble classifiers. In *Proc. 6th Intern. Conf. on Machine Learning and Applic. (ICMLA), 2007*, pages 404–409, 2007.

[25] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans. on Systems, Man and Cybernetics*, 15(1):116–132, 1985.

[26] A. Tsymbal. The problem of concept drift: definitions and related work. Technical Report TCD-CS-2004-15, Dept of Comp Science, Trinity College Dublin, Ireland, 2004.

[27] L.X. Wang. Fuzzy systems are universal approximators. In *Proc. 1st IEEE Conf. Fuzzy Systems*, pages 1163–1169, San Diego, CA, 1992.

[28] S. Wu, M.J. Er, and Y. Gao. A fast approach for automatic generation of fuzzy rules by generalized dynamic fuzzy neural networks. *IEEE Trans. on Fuzzy Systems*, 9(4):578–594, 2001.

ISBN: 978-989-95079-6-8

How to learn fuzzy user preferences with variable objectives

Alan Eckhardt[1],[2] Peter Vojtáš[1],[2]

1.Department of Software Engineering, Charles University
Prague, Czech Republic
2.Institute of Computer Science, Czech Academy of Science
Prague, Czech Republic
Email: {eckhardt , vojtas}@ksi.mff.cuni.cz

***Abstract*— *This paper studies a possibility to learn a complex user preference model, based on CP-nets, from user ratings. This work is motivated by the need of user modelling in decision making support, for example in e-commerce. We extend our user model based on fuzzy logic to capture variation of preference objectives. The proposed method 2CP-regression is described and tested. 2CP-regression uses CP-nets idea behind and can be considered as learning of a simple CP-net from user ratings.*

***Keywords*—** user preferences, data mining, ceteris paribus

1 Introduction

The problem studied in this paper is based on the idea of helping user in making decisions. The main motivation lies in e-commerce, where user might benefit a recommendation of objects that might interest her without needing to process all objects manually. This recommendation should be rather automatic and transparent, because user typically does not want to fill in any complicated forms or invest much time in search.

When a user is buying a notebook, for example, she is considering the attributes such as price, manufacturer, size of RAM, harddisk, display. In traditional e-shop environment, it is possible to make some restrictions on these attributes and lower the number of notebooks the user has to process manually this way. But these restrictions are inflexible - when the user selects price from 1000$ to 2000$, a notebook with price 999$ would not appear in the result set. This is the reason for a flexible search, based on fuzzy user preferences.

In this paper we propose a method for handling more complicated user preferences. In [1] a phenomenon called ceteris paribus was described. Preference ceteris paribus means preference "all else being equal". When a user is comparing two notebooks, we can say that the size of RAM of 2GB is preferred to 1GB ceteris paribus, meaning that we assume that the remaining attributes of these notebooks are the same and their values are not important. A user model based on this idea was later proposed in [2] - a CP-network is a graph that captures preferences ceteris paribus or, in different view, it represents attributes, on which the preferences over the attribute RAM depend. And yet, as far as we know, there are no study of learning of CP-nets from user ratings. Our paper is a first subtle contribution to this field. We work with our user model based on fuzzy sets rather than CP-nets, still the proposal of learning the relations between attributes is strongly related to CP-nets.

In Section 2 our user model is described. Related work is studied in Section 3. In Section 4 is described in depth the proposed method for learning the relation between attributes. Then this approach is tested in Section 5 and we end with conclusion and future work in Section 6.

2 Two step user model

2.1 Notation

We will work with a set of objects X. Overall rating of an object is a fuzzy subset of X, i.e. a function $r(o) : X \to [0,1]$, where 0 means least preferred and 1 means most preferred. Every object has attributes $A_1, ..., A_N$ with domains $D_{A_1}, ..., D_{A_N}$, with one special attribute that serves as identification of object (ID). Let $X \subseteq \prod_{i=1}^{N} D_{A_i}$. We will use $X(a)$ when denoting a set of objects that have the attribute value a.

User model, in our view, is a method for representing user decision when considering user's preference of an object $o \in X$. Our user model consists of two steps. At the first step, attributes of o are normalised using fuzzy sets $f_i : D_{A_i} \to [0,1]$ (and here again, 0 is least preferred and 1 is most preferred value), so that the transformed object is in $[0,1]^N$. These fuzzy sets are also called objectives or preferences over attributes. At the second step, preference over attributes are aggregated into the rating of the whole object via a fuzzy aggregation function $@ : [0,1]^N \to [0,1]$. Aggregation function is also often called utility function.

We concentrate on learning of user models. User is expected to rate a small sample $S \subseteq X$ of objects ($r : S \to \{1,2,3,4,5\}$), where $|S| \ll |X|$. The size of training set is expected to be very small. We assume that these ratings were created by the user using some fuzzy sets $f_1, ..., f_N$ and an aggregation function $@$. Of course the user does not compute the result of a function, but her decision corresponds to this two step process.

Our model consist of $f_1, ..., f_N$ and of $@$ so we expect to construct these functions from user ratings of S. The learnt functions will be denoted as $\widehat{f_1}, ..., \widehat{f_N}$ and $\widehat{@}$. Learning of aggregation function is described in Section 2.2. This paper concentrates on learning of local preferences, which is described in Section 4.

2.2 Aggregation function learning

Aggregation function serves to combine preferences over attributes into a single rating that represents the preference of the object as a whole. There are many ways to aggregate local preferences, we are currently using a weighted average with

ISBN: 978-989-95079-6-8

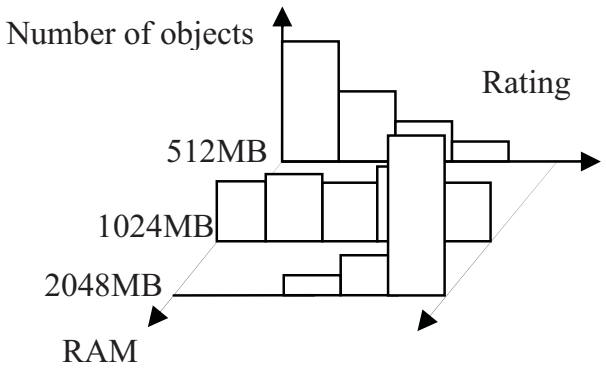


Figure 1: Uniform and expressive distributions of ratings.

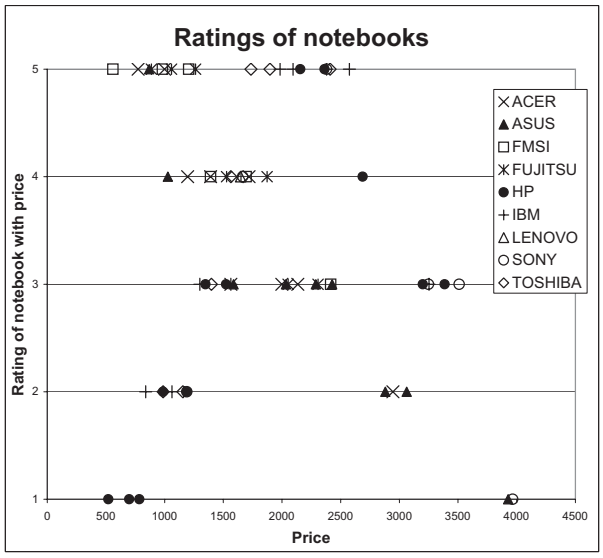


Figure 2: Distribution of ratings of notebooks with regard to price.

weights learnt from user ratings. Method called "Statistical" was described in [3], [4]. The distribution of ratings for attribute values $a_i \in A_i$ are considered when determining the weight of attribute A_i. In Figure 1 is an example of distribution of ratings across a domain of attribute RAM for a user that prefers larger sizes of RAM. If ratings are distributed near one point as in Figure 1 for 2048GB RAM, the attribute value is considered as decisive when doing the overall rating. However, if the distribution is rather uniform such as in Figure 1 for 1024MB RAM, then attribute value a_i does not play an important role in the overall rating. The measure of importance of an attribute value is computed with formula

$$imp(a_i) = 1/\sum_{o \in X(a_i)} |r(o) - avg_{o \in X(a_i)} r(o)|/|X(a_i)|$$

when taking only those objects o that have attribute value a_i. Then the importance of attribute A_i is computed as $W(A_i) = 1/(\sum_{a_i \in A_i} imp(a_i/|A_i|)$.

Our approach is not the only one - any data mining technique can be used here. Reader can simply imagine training a multilayer perceptron with normalised attribute values.

In Figure 2 is attribute price with ratings of notebooks. When a standard linear regression (described in Section 4.1) is used, attribute price would get a small weight. However, when notebooks of only a single manufacturer are used, then the error would be much lower and importance of price would be considerably higher. This may be visible for human, after a short inspection of the graph. Another visualisation of this situation is in Figure 3 with discretised attribute price. Ratings of all attribute values are distributed uniformly, so the importance of this attribute will be considered small.

3 Related work

User preferences are a wide field so there are plenty of research areas. For our case, user preference learning is the most important one. There are two totally different approaches to recommendation : content based and collaborative. For collaborative filtering, we may cite for example [5] but there are many other interesting works. Collaborative filtering is based on the idea that similar users tend to have similar preferences of an object. However, we do not consider this approach in this paper.

We concentrate here on content based, that means that the preferences are based on attributes of objects. For example, user considers the size of harddisk and other features when buying a notebook. This approach in our view corresponds more to how users decide in the real world. The difficulty here is that users are often inconsistent or take into account attributes that are not known or impossible to quantify, such as the design of notebook.

In this area, the main focus was on search of documents in the past [6, 7]. This is a specific area, because features of documents are of little structure. We are interested more in recommendation of more complicated objects like notebooks, with different types of attributes. Following works :[8, 9, 10, 11] deal with general, more structured, objects.

One of our main inspirations was CP-nets [2]. This user preference model captures complex user preferences in a graph representation, where a preference over one attribute may depend on the values of other attributes. There are many papers dealing with CP-nets, but there is none which describes a way to construct a CP-net automatically, by learning from ratings. Because preferences over attribute A_1 may depend on other attribute/s, e.g. A_2, user has to specify her preferences not only for all values of A_1 but she has to do it for every possible value of A_2. This means very much work and insight for the user, work that probably few people will undergo.

There is also another concept for representing user preferences other than ratings as in our case. These are preference relations, or specially fuzzy preference relations [12]. We do not consider this model, we rather follow approach identified by R.Fagin in [13] and consider user ratings as fuzzy sets.

4 Local preferences learning

As mentioned in Section 2, our user model is divided into two steps.

The first step is used for normalisation of every attribute to interval $[0, 1]$, so that the best object would have coordinates $[1, ..., 1]$. This can be viewed also as monotonisation of data - if for object o_1, the normalised value of its attribute value is lower in all attributes than object o_2, then o_2 is surely preferred to o_1 (if the normalisation is determined correctly).

ISBN: 978-989-95079-6-8

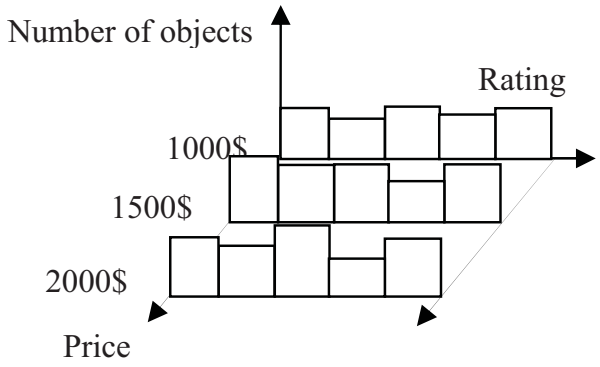


Figure 3: Uniform distributions of ratings across all domain.

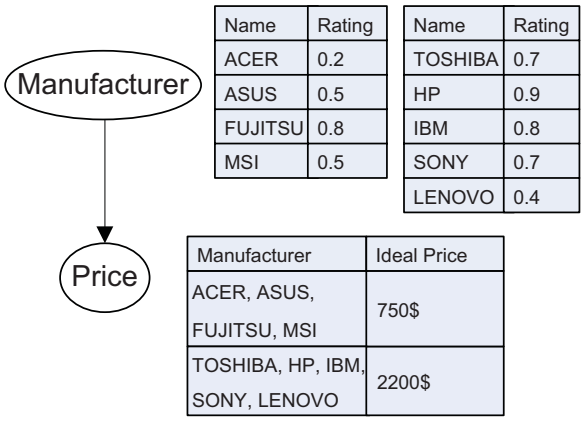


Figure 4: Example of a CP-net representing data about notebooks.

4.1 Linear regression

Linear regression is very useful method that serves to find a relation in a set of data in form of a linear function. It can be used for finding the preference of attribute values for numerical attributes, e.g. price. When given a set of notebooks' prices and ratings, linear regression will create a linear function $r(price) = \alpha * price + \beta$. Then we can normalise values of price using this function.

Traditional linear regression works on method of minimizing least squares of errors. This is not always useful, because it is dependent on the distribution of data points. For this reason, we proposed a method in [14] to accommodate better to the whole interval we study.

As a future work, we would like to implement a method for finding also more complicated functions, e.g. triangular fuzzy functions. The motivation is clear - linear functions are good for preferences of extreme values, e.g. lowest price, largest LCD etc. But if user prefers some value in the middle, like LCD of size 15", linear function would not capture this preference correctly.

4.2 2CP-regression

The main contribution of this paper is the proposal of a new method for acquisition of local preferences for numerical attributes, 2CP-regression. This method is motivated by preferences ceteris paribus [2] and CP-nets. Basic idea that represents ceteris paribus is that there are relations between attributes. For example, the manufacturer of a notebook influences the preferences of the price of that notebook. In our example, we present a user that prefers the price of 2200$ for manufacturers HP, IBM, Lenovo, Toshiba and Sony, and the price of 750$ for manufacturers Fujitsu-Siemens, Acer, Asus and MSI. A simple CP-net representing this scenario is in Figure 4, with ratings of manufacturers and ideal prices. Ideal price depends on the value of manufacturer, not on its preference. There are other examples and often the decisive attribute is a nominal one with rather small domain.

When constructing a fuzzy set for a numerical attribute A_1 (price in our case), we look at the values of other nominal attributes. For simplicity, let us consider only one, A_2. A_2 will be referred to as a "dividing attribute". The construction of the set is restricted to objects with value $a_2 \in A_2$; let us note this set $X(a_2) \subset X$. If dividing attribute A_2 really influences attribute A_1, then the values of A_1 should be "better" distributed in $X(a_2)$ than in whole X and the fuzzy set should better match the real preferences. The fuzzy set trained on $X(a_2)$ will be denoted as $\widehat{f_{a_2}^{A_1}}$. However, the size of training set may be radically decreased in this way. For that reason a general fuzzy set $\widehat{f^{A_1}}$ is constructed as in 4.1, which is used for values that did not occur during the construction of local preferences in the training set.

A method for construction of several fuzzy sets and a general one, all represented by linear functions, is described in the following pseudo code. Method `getObjectsWithA2(`a_2`)` returns the set $X(a_2)$, i.e. only those objects with attribute value a_2. Method `leaveOnlyA1AndRating` leaves only couples (A1, Rating) in the given set, which makes it suitable for linear regression that follows. Finally, the attribute value a_2 and the corresponding linear function $\widehat{f_{a_2}^{A_1}}$ are stored.

```
leaveOnlyA1AndRating(X);
f^{A_1} = buildClassifier(X);
for (∀a_2 ∈ A_2) {
    X(a_2) = getObjectsWithA2(a_2);
    if (|X(a_2)|< 2)
        continue;
    leaveOnlyA1AndRating(X(a_2));
    f_{a_2}^{A_1} = doLinearRegressionOn(X(a_2));
    storeCouple(a_2, f_{a_2}^{A_1});
}
```

Let us have an object o with unknown rating, with attribute values $a_1, ..., a_n$. When we want to normalise value a_1, we use the following method:

```
if(existsFuzzySetFor(a_2)){
    f_{a_2}^{A_1} = getFuzzySetFor(a_2);
    return f_{a_2}^{A_1}(a_1) ;
}
else
    return  f^{A_1}(a_1);
```

ISBN: 978-989-95079-6-8

It is possible that we did not encounter value a_2 in the training phase at all or only once. In such case, we use general function $\widehat{f^{A_1}}$ for normalisation of a_1.

4.3 A method for clustering results CP-regressions

In the previous section, we mentioned a problem that when using 2CP-regression, the number of training examples for linear regression decreases rapidly when the number of attribute values in the dividing attribute is high. In the following, we present a method for overcoming this problem.

Let us have a set of fuzzy sets $\widehat{f_{a_i}^{A_1}}$ for normalisation of attribute A_1 that depends on values of attribute A_2. Each fuzzy set is of the form $\widehat{f_{a_i}^{A_1}}(x) = \alpha * x + \beta$.

Now we start with fuzzy set $\widehat{f_{a_1}^{A_1}} = \alpha * x + \beta$ and match together those fuzzy sets, that have similar shape, ie. the values of α and β. After finding similar fuzzy sets, without the loss of generality let it be $\widehat{f_{a_2}^{A_1}}, ..., \widehat{f_{a_k}^{A_1}}$. Values $a_1, ..., a_k$ are grouped together and we construct a new linear function; we will use all objects $o \in X(a_1) \cup ... \cup X(a_k)$ for its construction. This new linear function should accommodate better to testing data, because it is constructed using a larger training set. When using this set, we treat it as $\widehat{f_{a_1,...,a_k}^{A_1}}$ so it is used for every object with attribute value in $a_1, ..., a_k$.

This method of using clusters for CP-regression was not fully implemented yet, but plan to implement it and test it in near future. We expect better performance especially for smaller training sets. In this case, we can do a bottom-up clustering that stops at a specified threshold of minimal training set size.

5 Experiments

5.1 Experiment settings

Our experiment was done on a set of 200 notebooks crawled from a web shop. There are five attributes: harddisk, display, price, producer and ram. We have created an artificial user U with preferences on the scale {1,2,3,4,5} for every notebook. Preferences were generated using a set of fuzzy sets f_i and an aggregation function @.

The user preferences on price were dependent on the actual value of the producer of the notebook. As it was mentioned in Section 4.2, for producers HP, IBM, Lenovo, Toshiba and Sony the best price was set to 2200$ and for Fujitsu-Siemens, Acer, Asus and MSI the best price was set to 750$. Hence the same price would produce different degree of preference for different manufacturers. Distribution of ratings of notebooks with regard to their price is in Figure 2. We can clearly see that the ratings are distributed more or less evenly across all the domain. However, a peak at about 2500$ appears when only the triangles are taken into account and another peak at about 1000$ is for crosses. Peaks do not correspond exactly to the artificial preferences; this is because the overall rating is influenced by other attributes, too. The overall trend is clearly visible for human - with increasing price crosses tend to go down and on the opposite triangles go up.

Testing was performed with a traditional cross-validation method. Because we are dealing with user ratings, it can not be expected that the user will rate many objects. That is why we limited the size of the training set (TSS in following text) to 75 ratings at maximum. Methods were tested on the rest of the set. When the user model is constructed from the first 10 notebooks, then it is tested on the remaining 190 notebooks. Then the next 10 notebooks are taken as training set and again the model is tested on the remaining 190. For TSS = 75 the model is tested on remaining 125 notebooks. Resulting errors are averaged for every TSS so the results should be reliable.

There are five methods tested: Mean, Statistical with Linear regression, Statistical with 2CP-regression, Support Vector Machine and Multilayer perceptron. Mean is a baseline. This method always returns the average rating from the training set. Multilayer perceptron and Support Vector Machine are traditional data mining methods. Implementation from Weka [15] was used.

Statistical method as aggregation function was used with two different normalisations. The first one is traditional linear regression; the second is our proposed method 2CP-regression.

No tuning of any method was performed - these methods are supposed to work on arbitrary data automatically without the need of adjusting any parameters.

5.2 Experiment results

We studied three types of error measures. First, most commonly used, is RMSE - root mean squared error. This captures the average error across the whole testing set. Second, Weighted RMSE, adds weight to each contribution to error. The weight associated is the same as the real rating, because good objects are of more concern than objects that are not preferred. In other words, it is worse if a good notebook does not appear on the first page of results than if a bad notebook does. Finally, we used tau coefficient to study the methods. Tau coefficient is a measure of correspondence between two ordered lists. Objects from testing set ordered by real user preferences are in the first list. The second list consists of the same objects, ordered by preferences proposed by the method. In this way, we can estimate the similarity of the ordering of objects by the method and by the real user. Tau coefficient ranges from -1 to 1; 1 means absolute correlation, e.g. the same lists, and -1 means the opposite correlation, e.g. reversed lists. Tau coefficient of 0 means that there is no relation between the two lists.

In Figure 5 are represented counts of unpredicted rating for each method. At TSS = 2 there were 40 notebooks that every method failed to predict and SVM failed even for 90 notebooks. At TSS = 5, there were 5 notebooks for all methods and 25 notebooks for SVM without prediction. So SVM performed worse than the rest of methods for small TSS.

In Figure 6 are results for RMSE. 2CP regression is an overall winner, only at TSS = 40 it was outperformed by Multilayer perceptron and at TSS = 2 by SVM, but SVM failed to predict far more notebooks.

The results are about the same for Weighted RMSE, as we can see in Figure 7.

Finally, results for Tau coefficient are in Figure 8. 2CP-regression was outperformed by Mean for the three smallest TSS, otherwise it was again the best method.

In all cases, 2CP-regression was better than simple Linear regression and Statistical with 2CP-regression outperformed even SVM and Multilayer perceptron.

ISBN: 978-989-95079-6-8

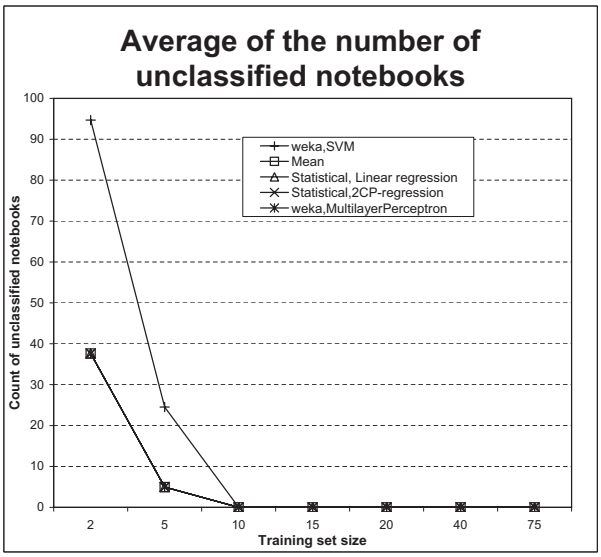


Figure 5: Count of unpredicted ratings of notebooks.

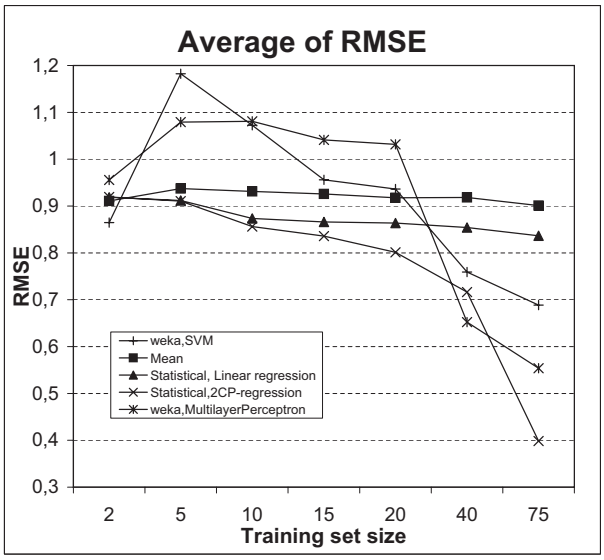


Figure 6: RMSE.

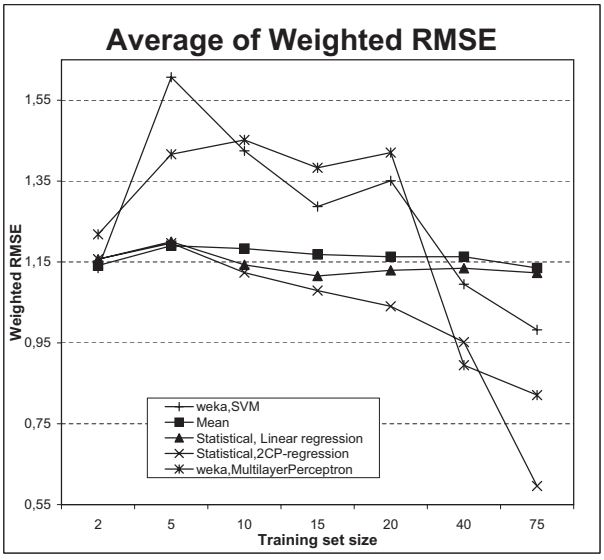


Figure 7: Weighted RMSE.

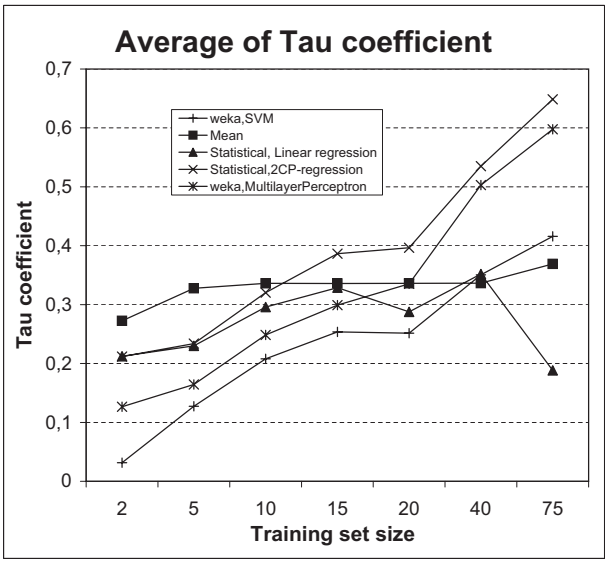


Figure 8: Tau coefficient.

6 Conclusions

We have proposed a method for finding preference dependence between two attributes, which was demonstrated on the example of a user for whom the ideal price depends on the manufacturer of the notebook. As far as we know this is the first contribution for learning ceteris paribus-network like structure from user ratings. We are aware that our model does not correspond to CP-nets completely, but the proposed part of our fuzzy model can be considered as a simple CP-net.

In our experiments, 2CP-regression performed very well, outperforming traditional data-mining techniques and also the normalisation with linear regression. Three different error measures were taken into account and in every one our proposed method was the best.

For future work, we would like to extend this approach for finding relations between more than two attributes, making a more general nCP-regression. We would also like to test it on real user preferences data, for example on NetFlix data [16].

Acknowledgment

The work on this paper was supported by Czech projects MSM 0021620838, 1ET 100300517 and GACR 201/09/H057.

References

[1] G. H. Wright. The logic of preference reconsidered. In *Theory and Decision*, volume 3, pages 140–169, 1972.

[2] Craig Boutilier, Ronen I. Brafman, Holger H. Hoos, and David Poole. Cp-nets: A tool for representing and reasoning with conditional ceteris paribus preference statements. *Journal of Artificial Intelligence Research*, 21:2004, 2004.

[3] Alan Eckhardt. Inductive models of user preferences for semantic web. In Jaroslav Pokorný, Václav Snášel, and Karel Richta, editors, *DATESO 2007*, volume 235 of *CEUR Workshop Proceedings*, pages 108–119. Matfyz Press, Praha, 2007.

[4] Alan Eckhardt, T. Horváth, D. Maruščák, R. Novotný, and Peter Vojtáš. Uncertainty issues in automating process connecting web and user. In P. C. G. da Costa, editor, *URSW '07 Uncertainty Reasoning for the Semantic Web - Volume 3*, pages 97–108. The 6th International Semantic Web Conference, 2007.

ISBN: 978-989-95079-6-8

[5] Al Mamunur Rashid, Istvan Albert, Dan Cosley, Shyong K. Lam, Sean M. McNee, Joseph A. Konstan, and John Riedl. Getting to know you: learning new user preferences in recommender systems. In *IUI '02: Proceedings of the 7th international conference on Intelligent user interfaces*, pages 127–134, New York, NY, USA, 2002. ACM.

[6] M. Ackerman, D. Billsus, S. Gaffney, S. Hettich, G. Khoo, D. Kim, R. Klefstad, A. Ludeman, J. Muramatsu, K. Omori, M. Pazzani, D. Semler, B. Starr, and P. Yap. Learning probabilistic user profiles: Applications to finding interesting web sites, notifying users of relevant changes to web pages, and locating grant opportunities. *AI Magazine*, 18:47–56, 1997.

[7] Thorsten Joachims. Optimizing search engines using clickthrough data. In *KDD '02: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 133–142, New York, NY, USA, 2002. ACM Press.

[8] Kim Cao-Van. *Supervised Ranking, from semantics to algorithms*. Ph.D. dissertation, Ghent University, 2003.

[9] Bruno Apolloni, Giacomo Zamponi, and Anna Maria Zanaboni. Learning fuzzy decision trees. *Neural Networks*, 11(5):885–895, 1998.

[10] Stefan Holland, Martin Ester, and Werner Kiessling. Preference mining: A novel approach on mining user preferences for personalized applications. In *Knowledge Discovery in Databases: PKDD 2003*, pages 204–216. Springer Berlin / Heidelberg, 2003.

[11] Sung Young Jung, Jeong-Hee Hong, and Taek-Soo Kim. A statistical model for user preference. *Knowledge and Data Engineering, IEEE Transactions on*, 17(6):834– 843, June 2005.

[12] Bernard de Baets and Jnos C. Fodor. Twenty years of fuzzy preference structures (19781997). *Decisions in Economics and Finance*, 20:45–66, 1997.

[13] R. Fagin, A. Lotem, and M. Naor. Optimal aggregation algorithms for middleware. In *In proceedings of Twentieth ACM Symposium on Principles of Database Systems, 2001 (PODS 2001)*, pages 102–113. ACM, 2001.

[14] Alan Eckhardt, T. Horváth, and Peter Vojtáš. Learning different user profile annotated rules for fuzzy preference top-k querying. In H. Prade and V. Subrahmanian, editors, *International Conference on Scalable Uncertainty Management*, volume 4772 of *Lecture Notes In Computer Science*, pages 116–130, Washington DC, USA, 2007. Springer Berlin / Heidelberg.

[15] Ian H. Witten and Eibe Frank. *Data Mining: Practical machine learning tools and techniques, 2nd Edition*. Morgan Kaufmann, San Francisco, 2005.

[16] Netflix dataset. *http://www.netflixprize.com.*

ISBN: 978-989-95079-6-8

Multi-Dimensional Scaling applied to Hierarchical Rule Systems

Thomas R. Gabriel Kilian Thiel Michael R. Berthold

Nycomed Chair for Bioinformatics and Information Mining
Department of Computer and Information Science
University of Konstanz, Box 712, 78457 Konstanz, Germany
Email: {gabriel,thiel,berthold}@inf.uni-konstanz.de

Abstract *— This paper presents an approach for visualizing high-dimensional fuzzy rules arranged in a hierarchy together with the training patterns they cover. A standard multi-dimensional scaling method is used to map the rule centers of the top hierarchy level to one coherent picture. Rules of the underlying levels are projected relatively to their parent level(s). In addition to the rules, all patterns are mapped onto the two-dimensional projection in relation to the positions of the corresponding rule centers. Visualization is further extended by showing hierarchical relationships between overlapping rules of different levels, which are generated by a hierarchical rule learner. This delivers interesting insights into the rule hierarchy and offers better explorative properties. Additionally, rules can be highlighted interactively emphasizing the subsequent rules at all underlying levels together with the patterns they cover. We demonstrate that this technique allows investigation of interesting rules at different levels of granularity, which makes this approach applicable even for a large number of rules. The proposed technique is illustrated and discussed based on a number of hierarchical rule model visualizations generated from well-known benchmark data sets.*

Keywords— Multi-Dimensional Scaling, Fuzzy Rule Induction, Rule Hierarchy, Rule Visualization.

1 Introduction

Rule learning algorithms are widely used in data mining to automatically extract rules from data. In [7, 15] and [20], algorithms are described that construct hyperrectangles in feature space. The resulting set of rules encapsulates regions in feature space that contain patterns of the same class. Other approaches, which construct fuzzy rules instead of crisp rules, are presented, for example, in [1, 12, 18] and [19]. All of these approaches have in common that they tend to build very complex rule systems for large data sets originating from a complicated underlying system. In addition, high-dimensional feature spaces result in complex rules relying on many attributes further increasing the number of required rules to cover the solution space. An approach that aims to reduce the number of constraints on each rule individually is presented in [3]. The generated fuzzy rules only constrain a few of the available attributes and hence remain readable even in the case of high-dimensional spaces. The fuzzy rules generated by this method have been visualized by parallel coordinates [4, 10].

In [8], we described a method that attempts to tackle this inherent problem of interpretability of large rule models. We achieve this by constructing rule models with varying degrees of complexity. The method builds a rule hierarchy for a given data set. The rules are arranged in a hierarchy of different levels of precision. Lower levels of the model describe regions in input space with low evidence in the given data, whereas rules at higher levels describe more strongly supported concepts of the underlying data. The method is based on the fuzzy rule learning algorithm described in [3, 9], which builds a single layer of rules autonomously. We use the resulting rule system recursively to determine rules of low relevance, which are then used as a filter for the next training phase. The result is a hierarchy of rule systems with the desired properties of simplicity and interpretability on each level. Experimental results demonstrated that fuzzy models at higher hierarchical levels show a dramatic decrease in number of rules while still achieving a better or similar generalization performance than the fuzzy rule system generated by the original, non-hierarchical algorithm.

In this paper we present an approach that enables the visualization of hierarchically structured rules and data in one coherent plot. A standard multi-dimensional scaling (MDS) method is applied to project rules of a certain level from the original, high-dimensional space, into a lower, usually two-dimensional space. Rules of the underlying levels are projected relatively to their parent level(s), likewise all data points are projected in relation to the positions of the corresponding rule centers. Furthermore, the visualization shows hierarchical relationships between overlapping rules of different levels as generated by a hierarchical rule learner. Due to the hierarchical nature of the induced rule system, interactive exploration becomes possible across the entire rule model and provides interesting insights into the underlying concept.

The paper is organized as follows: In the next section we briefly describe the used hierarchical rule learning method, followed by a short introduction to multi-dimensional scaling methods, and how data and hierarchical rules can be visualized by applying an extended multi-dimensional scaling method. In the following section, we describe how rule hierarchies, together with their original data, can be explored within this visualization. To illustrate the proposed method, hierarchical rule systems are generated based on the well-known iris data and on the vehicle silhouettes data.

2 Learning Hierarchical Rule Systems

The rule induction algorithm used here is based on a method described in [3], which is based on an iterative algorithm. During each learning epoch, i. e. presentation of all training patterns, new fuzzy rules are induced when necessary and existing ones are adjusted whenever a conflict occurs. For each pattern three main steps are executed. First, if a new training pattern lies inside the support-region of an existing fuzzy rule of the correct class, its core-region is extended in order to cover the new pattern. Second, if the new pattern is not yet

ISBN: 978-989-95079-6-8

covered, a new fuzzy rule of the correct class is introduced. The new example is assigned to its core, whereas the support-region is initialized by infinity, that is, the new fuzzy rule covers the entire domain. Finally, if a new pattern is incorrectly covered by an existing fuzzy rule, the fuzzy rules' support-region is reduced to avoid the conflict. This heuristic for conflict avoidance aims to minimize the loss in volume. In [9], three different heuristics are compared to determine the loss in volume. As discussed in [3], the algorithm terminates after only a few iterations over the set of example patterns.

The final set of fuzzy rules can be described as an n-dimensional IF clause as antecedence and one assigned $class$ in the rule's conclusion:

$$\begin{array}{lllll}
\mathcal{R}_1^1 & \texttt{IF} & x_1 \texttt{ IS } \mu_{1,1}^1 & \wedge\cdots\wedge & x_n \texttt{ IS } \mu_{n,1}^1 \quad \texttt{THEN class } 1 \\
 & & \vdots & \vdots & \vdots \\
\mathcal{R}_{r_1}^1 & \texttt{IF} & x_1 \texttt{ IS } \mu_{1,r_1}^1 & \wedge\cdots\wedge & x_n \texttt{ IS } \mu_{n,r_1}^1 \quad \texttt{THEN class } 1 \\
 & & \vdots & \vdots & \vdots \\
\mathcal{R}_j^k & \texttt{IF} & x_1 \texttt{ IS } \mu_{1,j}^k & \wedge\cdots\wedge & x_n \texttt{ IS } \mu_{n,j}^k \quad \texttt{THEN class } k \\
 & & \vdots & \vdots & \vdots \\
\mathcal{R}_{r_c}^c & \texttt{IF} & x_1 \texttt{ IS } \mu_{1,r_c}^c & \wedge\cdots\wedge & x_n \texttt{ IS } \mu_{n,r_c}^c \quad \texttt{THEN class } c
\end{array}$$

where $\mathcal{R}_j^k$ represents rule j for class k. The rule base contains rules for c classes and r_k indicates the number of rules for class k ($1 \leq j \leq r_k$ and $1 \leq k \leq c$). The fuzzy sets $\mu_{i,j}^k : \mathbb{R} \mapsto [0,1]$ are defined for every feature i ($1 \leq i \leq n$), but in cases of unconstrained features the membership degree constantly remains 1. The overall degree of fulfillment of a specific rule for an input pattern $\vec{x} = (x_1, \ldots, x_n)$ can be computed using the fuzzy set operator for conjunction, called T-norm ($\top$):

$$\mu_j^k(\vec{x}) = \top_{i=1,\cdots,n} \left\{ \mu_{i,j}^k(x_i) \right\}.$$

The combined degree of membership for all rules of class k can be calculated using the fuzzy set operator for disjunction, called S-norm ($\perp$):

$$\mu^k(\vec{x}) = \perp_{j=1,\cdots,r_k} \left\{ \mu_j^k(\vec{x}) \right\}.$$

From these membership values the predicted class k_{best} for an input pattern $\vec{x}$ is derived as:

$$k_{\text{best}}(\vec{x}) = \arg\max_{k=1,\ldots,c} \left\{ \mu^k(\vec{x}) \right\}.$$

The algorithm uses trapezoidal membership functions, which can be described by four parameters $\langle a_i, b_i, c_i, d_i \rangle$, where a_i and d_i define the fuzzy rule's support-, and b_i and c_i its core-region for each attribute i of the input dimension. The core-region is defined as a rectangular area with an activation of 1, whereby the support-region decreases linearly to its boundaries with a degree of fulfillment of 0.

The resulting set of fuzzy rules can then be used to classify new patterns by computing the overall degree of membership for each class. The accumulated membership degrees over all input dimensions and across multiple rules are calculated by a fuzzy norm. For the purpose of this paper, we concentrate on the rules' core-regions only, that is, the part of each rule where the degree of membership is equal to 1 – resulting in crisp rules – is considered.

An extension of this rule learning algorithm is proposed in [8], which allows the generation of a hierarchy of rules. Therefore, the classical fuzzy rule algorithm is used to determine rules of low relevance recursively, which are then used as a filter for the next training phase. Training examples that cause the creation of small, less important rules are therefore excluded from the training phase of the next layer, resulting in a more general rule system, ignoring the withheld, small details in the training data. An example of a two-stage rule induction process is illustrated in Fig 1.

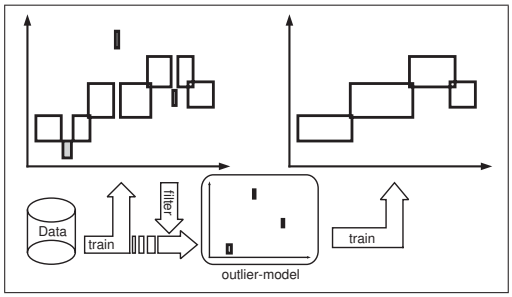


Figure 1: Two-stage outlier filtering as described in [2].

This process can be repeated to generate the desired hierarchy of rule systems with an increasing generality towards higher levels. The rule layers are arranged in a hierarchy of different levels of precision. Lower levels of the model describe regions in input space with low evidence in the training data, whereas rules at higher levels describe more strongly supported aspects of the underlying data.

3 Multi-Dimensional Scaling

To visualize objects of high-dimensional feature spaces, multi-dimensional scaling methods [6] can be applied to map them onto lower-dimensional spaces with usually two or three dimensions. In order to avoid a loss of proximity information, MDS methods try to preserve the pairwise distances between objects during the mapping process by minimizing an appropriate error function. The reduction of dimensions allows the visualization of high-dimensional objects in a lower-dimensional space, using traditional methods such as scatter plots or scatter matrices.

One technique to apply this kind of mapping by minimizing a particular error function is the well known Sammon algorithm [17]. For each object $\vec{X}_i$ in high-dimensional space $\mathbb{R}^H$, a spatial representation $\vec{x_i} \in \mathbb{R}^L$ in low-dimensional space (usually $L = 2$) has to be computed ($1 \leq i \leq N$ with N as the number of objects).

To approximate the distance information, the position of each object in the low-dimensional space has to be adjusted such that the pairwise distances between each two objects $(\vec{x_i}, \vec{x_j})$ in the low-dimensional space, $d_{ij} = d(\vec{x_i}, \vec{x_j})$, approximate the distances of the two corresponding objects $(\vec{X}_i, \vec{X}_j)$ in the high-dimensional space, $D_{ij} = D\left(\vec{X}_i, \vec{X}_j\right)$:

$$\forall_{i \neq j} : D_{ij} \approx d_{ij}, 1 \leq i, j \leq N.$$

Usually the Euclidean metric is used to measure these distances in the low-dimensional space, and often in the high-

ISBN: 978-989-95079-6-8

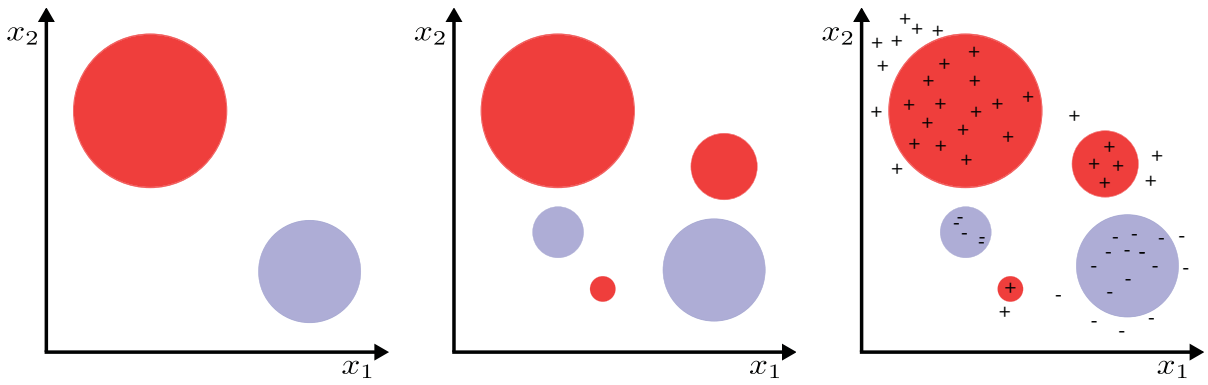


Figure 2: Shows a two-level rule hierarchy. (a) Two top-level rules assigned to two different classes, + (red) and − (blue), (b) three additional rules on the next level, (c) all rules together with their underlying data points.

dimensional space as well:

$$d_{ij}^2 = \sum_{l=1}^{L} (x_{i,l} - x_{j,l})^2.$$

Approximation of the pairwise distances in high- and low-dimensional space is formulated by Sammon as a minimization problem of a cost function, which aggregates the weighted squared differences of the distances in high- and low-dimensional space:

$$E = \sum_{i=1}^{N} \sum_{j>i}^{N} \omega_{ij} (d_{ij} - D_{ij})^2.$$

where the factors ω_{ij} are introduced to weight the distances individually and normalize the stress function E in order to be independent from the absolute values D_{ij}. The steepest gradient method is applied to minimize iteratively a cost function E for each object $\vec{x_i}$ at each step. Usually several iterations are needed by the algorithm to converge to a local cost minimum.

3.1 *Projective Multi-Dimensional Scaling*

As mentioned above, the original Sammon algorithm tries to preserve the pairwise distances of all objects. This means that the position of each object in the low-dimensional space is adjusted iteratively according to the position of the other objects in the high- and low-dimensional space. Subsequently this leads to a change of the positions of all objects in the low-dimensional space. With our projective multi-dimensional scaling approach, the positions of the objects in the low-dimensional space are adjusted according to a set of fixed projection objects.

First a set of objects in the high-dimensional space is selected. For each of these objects $\vec{X_p} \in \mathbb{R}^H$ a spatial representation $\vec{x_p} \in \mathbb{R}^L$ ($1 \leq p \leq M$ with $M \ll N$ as the number of fixed projection objects $\vec{x_p}$) has to be found by means of the standard MDS algorithm described above. Once these objects have been mapped onto the low-dimensional space, they are used as fixed objects and are not changed anymore. Furthermore, each of the other objects $\vec{X_i} \in \mathbb{R}^H$ are mapped onto the low-dimensional space according to the fixed projection objects. Therefore, the distance between a regular object $\vec{x_i} \in \mathbb{R}^L$ and a fixed projection object $\vec{x_p}$ in the low-dimensional space $d_{ip} = d(\vec{x_i}, \vec{x_p})$ has to be approximated to the distance between the two corresponding objects $(\vec{X_i}, \vec{X_p})$ in the high-dimensional space $D_{ip} = D\left(\vec{X_i}, \vec{X_p}\right)$, which translates to:

$$\forall_{i \neq p} : D_{ip} \approx d_{ip}, 1 \leq i \leq N, 1 \leq p \leq M.$$

Again, a cost function E_P is defined that aggregates the weighted squared differences of the distances in the high- and low-dimensional space:

$$E_P = \sum_{i=1}^{N} \sum_{p=1}^{M} \omega_{ip} (d_{ip} - D_{ip})^2.$$

The projective MDS method is useful if a set of objects has to be mapped according to an already existing set of mapped objects without modifying the mapping of the latter.

In the case of high-dimensional hierarchical data, projective MDS can be used to determine a spatial representation in several iterations. First the standard MDS method is applied to the top-level data. Once this data is mapped, projective MDS can be applied stepwise to the data of the lower levels using the data of higher levels as fixed projection objects. The low-dimensional representation of the already mapped data does not change anymore when applying projective MDS to the remaining levels.

3.2 *Hierarchical Rules in Multi-Dimensional Scaling*

It is more complex to visualize rule models on lower dimensions than simple data points. The main challenge is the mapping of the rules' antecedences, which are usually hyperrectangles in the original high-dimensional feature space. In order to visualize high-dimensional fuzzy rules by means of MDS, the center points of a rules' core-regions are mapped onto a low-dimensional space, as described in [11, 16]. For hierarchical fuzzy rule systems, the center points of the top-level rules are mapped onto the low-dimensional space using standard MDS. Subsequently, the rules of the underlying levels are mapped level by level via the projective MDS with the parent rule centers as fixed projection points. Finally, the data points are mapped according to all projected rules, whereas the neighborhood of rules and data points is approximated by the proposed MDS. The overlapping relation between rules of different levels may get lost in the visualization, but is again

ISBN: 978-989-95079-6-8

taken into account when rule systems are visually explored by interactive highlighting, see Sec. 4.2 for an example.

In addition to the rules' center points, the spread of a fuzzy rule has to be mapped onto the low-dimensional space in order to visualize the rules' sizes, possible overlaps, and the coverage according to the data points the rules are based on. In [11], we visualize the overlap of rules of flat rule systems. Dealing with hierarchical fuzzy rules, we only focus on the visualization of the spread and the number of covered data points of a rule as well as the rules' level and not on overlap with other rules.

The spread is visualized by a sphere around the mapped rule center with a radius $r_i = \rho(\lambda\omega_i + (1-\lambda)\sigma_i^2)$ according to the variance σ_i^2 and the number of the points ω_i covered by a rule $\mathcal{R}_i$, with λ as weighting coefficient $\lambda \in [0,1]$ and $\rho(\cdot)$ as scaling function. The number of vertices $v = 2^H$ of a hyperplane grows exponentially with the number of dimensions H whereas all vertices have the same distance to the center point. All vertices of a high-dimensional hyperplane are mapped by MDS, representing vertices of a low-dimensional polygon; all with the same distance to the polygons center point. With increasing dimensionality, the polygon spanned by the projected vertices converges to a sphere.

In Fig. 2 (a), two top-level rules are shown, covering data points of two different classes. The red rule covers data points of class $+$, while the blue covers data points of class $-$. Fig. 2 (b) illustrates rules of the primary and the secondary level, which consists of three smaller rules, two rules covering data points of class $+$ and two of class $-$. Fig. 2 (c) shows rules of all levels as well as the data points. It can be seen that the spheres representing the rules do not necessarily cover all the data points in the low-dimensional space as they do in the high-dimensional space. This is because the radius is based on the number of covered data instances of a rule as well as the variance of the covered data points in the high-dimensional space, but is not based on the position of the data points in the low-dimensional space.

4 Visualization Hierarchical Rules and Data

The following section illustrates the proposed approach on the well-known iris dataset before looking at a larger dataset, the vehicle silhouettes data. We compare the classical, non-hierarchical rule model to the hierarchical structured rules by visualizing all the rules together with their originating data instances in one coherent picture.

4.1 Iris Dataset

The first example shows a small two-level hierarchy trained on the iris data, which consists of 150 four-dimensional patterns assigned to three classes. Fig. 3 (a) shows the top-rule level containing three rules – one for each class, (b) the top and bottom-rule levels with additional 11 rules (some cover only a single or a few instances and therefore appear as small points), and (c) all data points and all rules together. The top-level rules are projected into the low-dimensional space based on each center point. The second level is then projected based on the first rule level. In the last step, data points are projected according to both hierarchical levels as shown in Fig. 3. The bottom picture shows rules and data instances of all three classes, which are almost perfectly separated from each other by three large rules. Smaller rules in-between are needed to cover artifacts and details in the data.

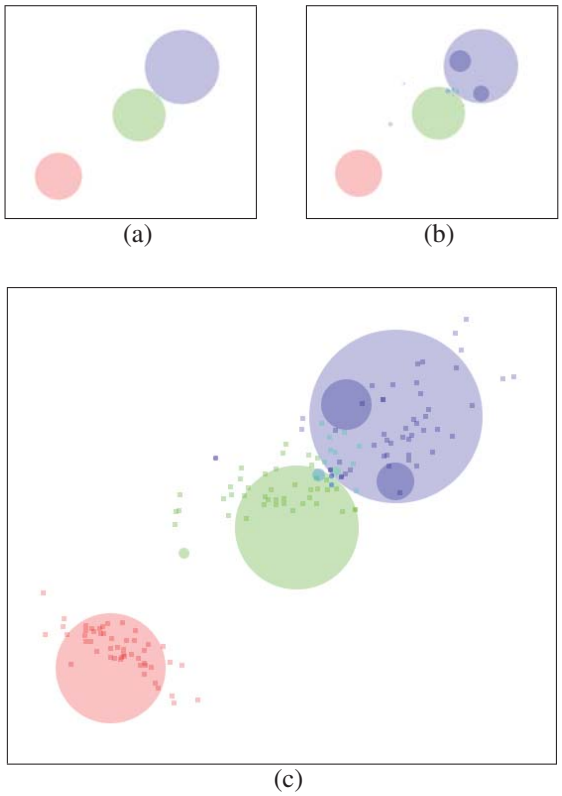


Figure 3: Two-level rule hierarchy trained on iris data: (a) three top-most rules, (b) additional 11 rules from the bottom-level, and all data points in (c).

4.2 Vehicle Silhouettes Dataset

The vehicle silhouettes dataset consists of 846 samples belonging to four car classes – Opel, Saab, bus, and van – represented in an 18-dimensional feature space. To demonstrate the usefulness of our proposed method, we trained a three-level fuzzy rule hierarchy on this data from the European StatLog–Project [14]. In a first experiment, a classical rule model with 222 rules without hierarchy information is trained on the vehicle data. As can be seen in Fig. 4, exploration is hardly pos-

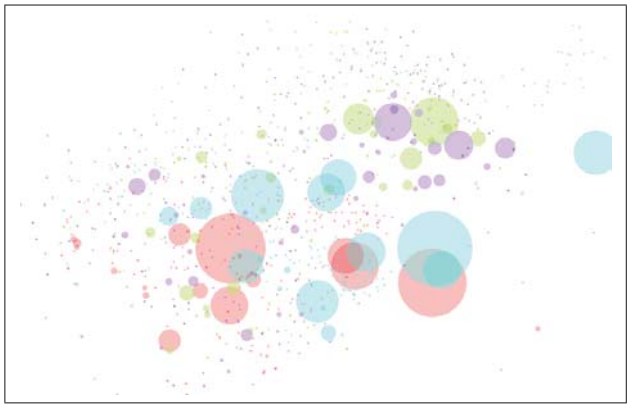


Figure 4: Classical, flat rule model without hierarchy information trained on the vehicle silhouettes dataset showing 222 rule for four classes.

ISBN: 978-989-95079-6-8

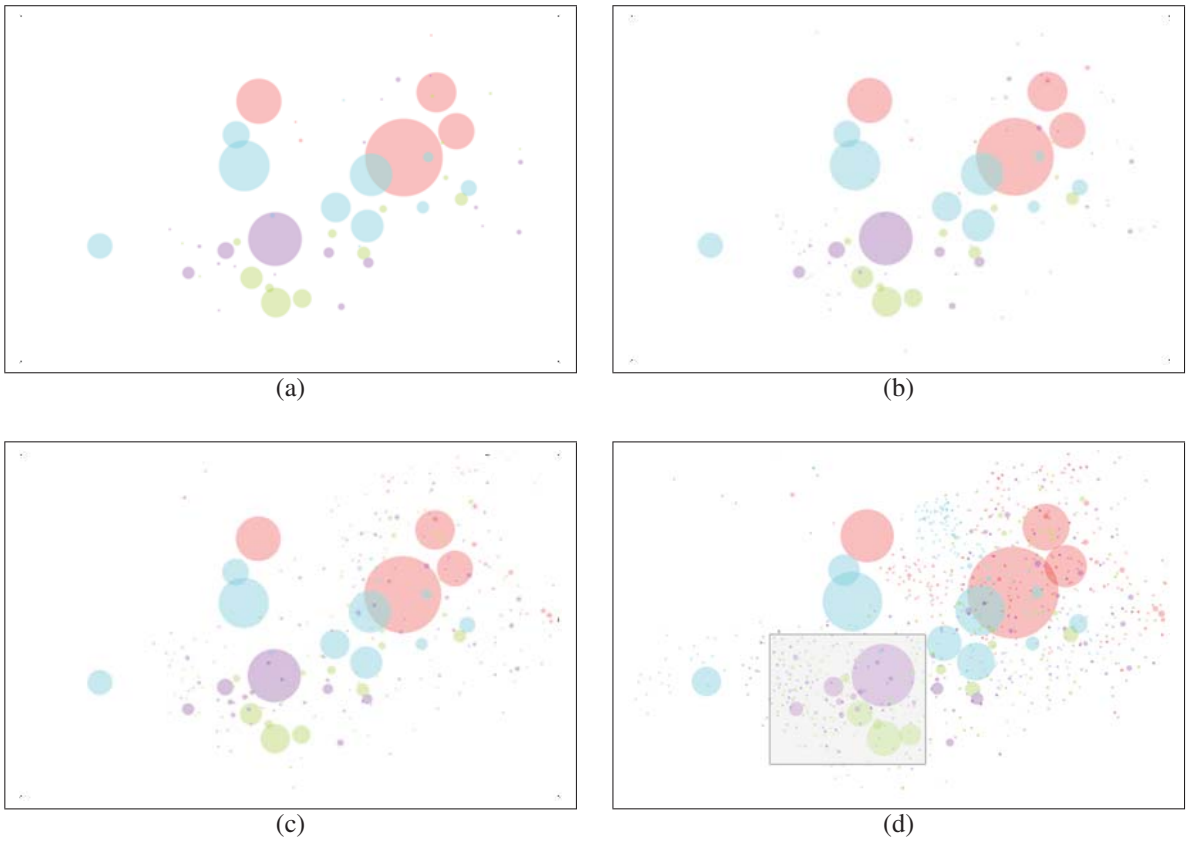


Figure 5: Stepwise rule hierarchy projection on vehicle data starting from top- to bottom-level (a)-(c), and (d) rules of all levels and data points.

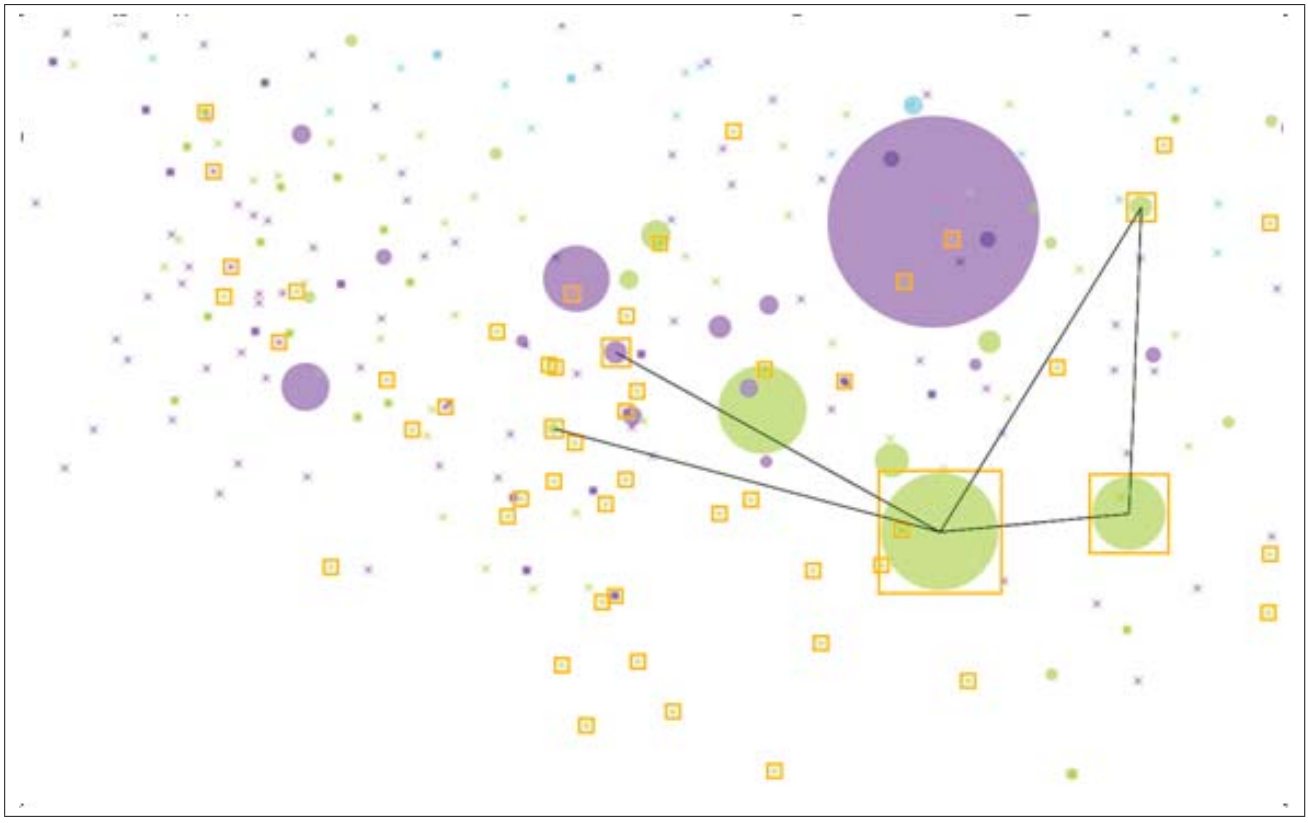


Figure 6: Exploration by zooming and highlighting (orange square) interesting rules and data points in the vehicle hierarchy.

ISBN: 978-989-95079-6-8

sible since overlapping information is only available between rules of the same single level [1].

The hierarchical rule induction algorithm is applied to generate a hierarchy of rule models in a second test. The three levels of the fuzzy rule hierarchy contain 55 rules in the top, 40 in the middle, and 174 at the bottom-most level. Fig. 5 shows each level together with its parent levels starting from (a) to (c), and (d) rules of all levels and data points. All levels are subsequently projected into the two-dimensional space always with respect to their parent level(s), as well as all data points that are projected relatively to all rule levels. Fig. 6 shows the surrounded area from Fig. 5 (d) enlarged. Selecting one rule highlights all overlapping rules in the levels above along with the data points they cover. All highlighted rules (orange square) that cover the same set of patterns are connected by a line to visually identify overlapping rules between different levels of the hierarchy. The figure shows five connected rules where a violet rule from a lower level completely overlaps with a large, green rule of a higher level. This is typically an indication for outliers and artifacts in the data expressed by smaller, more specific rules at lower levels generated by the rule learning algorithm at first place; whereas rules generated at higher levels of the hierarchy explain more general aspects of the data and are usually covered by larger rules. This visual line-up of rules of different levels allows further exploration by highlighting interesting data instances.

5 Conclusions

Hierarchically structured rules induced by a classical rule learning algorithm lead to a well-defined hierarchy of rules where levels further up explain more general aspects and rule models at lower levels concentrate on artifacts or outliers of the underlying concept. Combining this information with a mapping mechanism to visualize both this type of hierarchy and the data points enables interactive exploration across rule levels by focusing on overlapping regions. On the other hand, it highlights data points covered throughout the rule hierarchy. We demonstrated the explorative power of the proposed projective multi-dimensional scaling method based on the vehicle silhouettes dataset, delivering interesting insights into the underlying concept. Our approach is well suited for projecting other hierarchical structured data and rules driving interactive explorative data analysis.

References

[1] S. Abe, and M.S. Lan. A method for fuzzy rules extraction directly from numerical data and its application to pattern classification. *IEEE Transactions on Fuzzy Systems*. 3:18–28, 1995.

[2] M.R. Berthold. Learning fuzzy models and potential outliers. *In Computational Intelligence in Data Mining.* Springer, Berlin, 111–126, 2000.

[3] M.R. Berthold. Mixed Fuzzy Rule Formation. *International Journal of Approximate Reasoning (IJAR)*. Elsevier, 32:67–84, 2003.

[4] M.R. Berthold, and L.O. Hall. Visualizing Fuzzy Points in Parallel Coordinates. *IEEE Transactions on Fuzzy Systems*. IEEE Press, 11(3):369–374, 2003.

[5] M.R. Berthold, and R. Holve. Visualizing high dimensional fuzzy rules. *Proceedings of NAFIPS*. IEEE Press, 64–68, 2000.

[6] T.F. Cox, and M.A. Cox. Multidimensional Scaling. *Monographs on Statistics and Applied Probability*. Chapman and Hall, 1994.

[7] W. Duch, R. Setiono, and J.M. Zurada. Computational Intelligence Methods for Rule-based Data Understanding. *Proceedings of the IEEE*. IEEE Press, 92(5):771–805, 2004.

[8] Th.R. Gabriel, and M.R. Berthold. Constructing Hierarchical Rule Systems. In: M.R. Berthold, H.-J. Lenz, E. Bradley, R. Kruse, and C. Borgelt (Eds.): *Proc.* 5^{th} *International Symposium on Intelligent Data Analysis (IDA)*. Springer, Berlin, 76–87, 2003.

[9] Th.R. Gabriel, and M.R. Berthold. Influence of fuzzy norms and other heuristics on "Mixed Fuzzy Rule Formation". *International Journal of Approximate Reasoning (IJAR)*. Elsevier, 35:195–202, 2004.

[10] Th.R. Gabriel, A.S. Pintilie, and M.R. Berthold. Exploring Hierarchical Rule Systems in Parallel Coordinates. In: Famili et al. (Eds.): *Proc.* 6^{th} *International Symposium on Intelligent Data Analysis (IDA)*. Springer, Madrid, 97–108, 2005.

[11] Th.R. Gabriel, K. Thiel, and M.R. Berthold. Rule Visualization based on Multi-Dimensional Scaling. *IEEE International Conference on Fuzzy Systems*. IEEE Press, Vancouver, 66–71, 2006.

[12] C.M. Higgins, R.M. Goodman. Learning fuzzy rule-based neural networks for control. *Advances in Neural Information Processing Systems*. California, Morgan Kaufmann, 5:350–357, 1993.

[13] J. Meulman. A distance approach to nonlinear multivariate analysis. DSWO Press, Leiden, The Netherlands, 1986.

[14] D. Michie, D.J. Spiegelhalter, and C.C. Taylor, editors. *Machine Learning, Neural and Statistical Classification*. Ellis Horwood Limited, 1994.

[15] S. Salzberg. A nearest hyperrectangle learning method. *Machine Learning*. 6:251–276, 1991.

[16] F. Rehm, F. Klawonn, and R. Kruse. Visualization of Fuzzy Classifiers. *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems*. World Scientic Publishing Company, 15(5):615–624, 2007.

[17] J.W. Sammon, Jr. A non-linear mapping for data structure analysis. *IEEE Transactions on Computers*. C18(5):401–409, May 1969.

[18] P.K. Simpson. Fuzzy min-max neural networks – part 1: Classification. *IEEE Transactions on Neural Networks*. 3:776–786, 1992.

[19] L.X. Wang, and J.M. Mendel. Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man, and Cybernetics*. 22:1313–1427, 1992.

[20] D. Wettschereck. A hybrid nearest-neighbour and nearest-hyperrectangle learning algorithm. *Proceedings of the European Conference on Machine Learning*. 323–335, 1994.

[1] In some rare cases, neighboring data points of different classes generate rules with overlapping core-regions.

ISBN: 978-989-95079-6-8

Rule Weight Optimization and Feature Selection in Fuzzy Systems with Sparsity Constraints *

Edwin Lughofer[1] Stefan Kindermann[2]

1. Department of Knowledge-Based Mathematical Systems, Johannes Kepler University
Altenbergerstrasse 69, A-4040 Linz, Austria
2. Department of RICAM, Johannes Kepler University
Altenbergerstrasse 69, A-4040 Linz, Austria
Email: edwin.lughofer@jku.at, stefan.kindermann@indmath.uni-linz.ac.at

Abstract *— In this paper, we are dealing with a novel data-driven learning method (*SparseFIS*) for Takagi-Sugeno fuzzy systems, extended by including rule weights. Our learning method consists of three phases: the first phase conducts a clustering process in the input/output feature space with iterative vector quantization. Hereby, the number of clusters = rules is pre-defined and denotes a kind of upper bound on a reasonable granularity. The second phase optimize the rule weights in the fuzzy systems with respect to least squares error measure by applying a sparsity-constrained steepest descent optimization procedure. This is done in a coherent optimization procedure together with elicitation of consequent parameters. Depending on the sparsity threshold, more or less rules weights can be forced towards 0, switching off some rules. In this sense, a rule selection is achieved. The third phase estimates the linear consequent parameters by a regularized sparsity constrained optimization procedure for each rule separately (local learning approach). Sparsity constraints are applied here in order to force linear parameters to be 0, triggering a feature selection mechanism per rule. In some cases, this may also yield a global feature selection, whenever the linear parameters of some features in each rule are near 0. The method is evaluated based on high-dimensional data from industrial processes and based on benchmark data sets from the internet and compared to well-known batch training methods in terms of accuracy and complexity of the fuzzy systems.*

Keywords— Takagi-Sugeno fuzzy systems, iterative vector quantization, rule weight optimization, sparsity constraints, rule selection, embedded feature selection

1 Introduction

1.1 Motivation and State of the Art

Nowadays data-driven learning of fuzzy systems enjoy a great attraction in many industrial applications, as, opposed to expert-based fuzzy systems, they can be fully automatically generated from process data such as measurements, images (features) or signal streams and hence save a lot of time and money for the development of system models. Their big advantage over other methods such as neural networks or genetic programming are that they are universal approximators [22] (i.e. being able to model a given problem to any degree of accuracy) and at the same time they still allow some insights in form of linguistically [5] and visually interpretable rules [19]. Due to these benefits, data-driven fuzzy systems are applied in many application areas such as fault detection, image classification or identification models in control systems and hence serve as important components in these systems.

During the recent 15 to 20 years a lot of data-driven fuzzy systems approaches were developed: some methods such as popular *genfis2* [23] (as implemented in MATLAB's fuzzy logic toolbox) and its successor *genfis3*, *FMCLUST* [2], exploit various clustering techniques (e.g. Gustafsson-Kessel [14] or variations of mountain and subtractive clustering [6]) for extraction clusters out of the data which are directly associated with the rules. A big family of design methods use genetic algorithms [9] for 1.) eliciting the optimal number of rules by achieving a reasonable tradeoff between accuracy and model complexity and 2.) for (fine-)tuning the parameters: these are also called genetic fuzzy systems [10] and enjoy a great attraction in the fuzzy community since more than 10 years [8].

Other methods use numerical optimization procedures for linear (consequent) and non-linear (antecedent) parameters in the fuzzy systems for minimizing the least squares errors (and variants) between estimated and predicted target values: a well-known method is the *ANFIS* approach [15] exploiting a neuro-fuzzy system architecture similar to a Takagi-Sugeno type system; another method demonstrated in [3] uses Levenberg-Marquardt algorithm for optimizing non-linear antecedent parameters in triangular and Gaussian fuzzy sets. *RENO* [4] exploits a generalized Gauss-Newton like method as a second-order approximation method with (regularized) smoothing constraints.

None of these methods perform an implicit feature selection approach, neither globally nor locally, which may be important in case of high-dimensional spaces for decreasing curse of dimensionality effect. Moreover, most of these methods do not have an integrated approach for a deterministic optimization of rule structure (number of rules) and consequent parts in a coherent procedure at the same time. Indeed, genetic fuzzy systems provide this aspect, however are non-deterministic and often lack efficiency in terms of computational performance.

*This work was supported by the Upper Austrian Technology and Research Promotion. This publication reflects only the author's view.

ISBN: 978-989-95079-6-8

1.2 Our Approach

In this paper, we present a novel technique for data-driven learning of Takagi-Sugeno fuzzy systems called *SparseFIS*, which is short for *Sparse Fuzzy Inference Systems*. This approach is characterized by mainly three issues (more details in Section 2):

1. it applies a top-down approach where the upper bound on the number of rules are allowed to be predefined (e.g. by an expert) and the non-linear parameters in the fuzzy sets of the antecedent parts are pre-estimated by an iterative vector quantization approach (Section 3).

2. it performs a rule selection strategy by optimizing rule weights together with the linear consequent parameters within one non-linear optimization procedure (Section 4). The application of a threshold operator therein (for forcing as much as possible rule weights towards zero) allows optimality in the least squares sense and readability at the same time.

3. it includes an embedded local (=for each rule separately) feature selection technique by applying a semi-smooth Newton method for a sparsity regularized estimation of the linear consequent parameters (Section 5). This may decrease the curse of dimensionality effect and improves interpretability of the achieved fuzzy systems.

Combining these three concepts together will guide us to the *SparseFIS* algorithm in Section 6, which is able to generate accurate and sparse fuzzy systems at the same time. Section 7 compares the novel method in terms of predictive accuracy and complexity with other well-known (and in MATLAB achievable) data-driven fuzzy system methods based on four data sets from the UCI respository and on noisy engine test bench data.

2 Problem Statement

Let us first define the Takagi-Sugeno fuzzy model with rule weights and by applying product t-norm with Gaussian fuzzy sets (in case of C rules):

$$y_{fuz} = \sum_{i=1}^{C} l_i(x)\Psi_i(x,\rho), \quad l_i(x) = \sum_{k=1}^{p} w_{i,k}x_k \qquad (1)$$

with

$$\mu_i(x) = \Pi_{j=1}^{p}\mu_{i,j}(x_j), \quad \mu_{i,j}(x_j) = e^{-\frac{(x_j - c_{i,j})^2}{2\sigma_{i,j}^2}}$$

and

$$\Psi_i(x,\rho) = \frac{\rho_i\mu_i(x)}{\sum_{j=1}^{C}\rho_j\mu_j(x)} \qquad (2)$$

Here $\rho = (\rho_1, \ldots \rho_C)$ are positive weights, which more or less steer the number of rules (as weights smaller than a low number $\epsilon > 0$ can be seen as not useful and hence omitted). For optimization purposes we use the least squares error measure, which, when applying TS fuzzy systems as above and in dependency of the unknown parameters, can be written as

$$J(w,c,\sigma,\rho) = \sum_{k=1}^{N}(y_k - \sum_{i=1}^{C} l_i(\vec{x}_k)\Psi_i(\vec{x}_k))^2 \qquad (3)$$

The training goals (=steps in the algorithm) in this paper can be summarized as follows:

1. Estimating the non-linear antecedent parameters in a preliminary phase due to a pre-defined upper bound on the number of rules.

2. Minimizing the least squares error by keeping the model complexity (in terms of the number of rules) as low as possible (rule selection from the initial number). This is done in a coherent procedure with optimization of the consequent parameter synchronously.

3. Minimizing the least squares error by keeping the number of significant linear consequent parameters in each rule as low as possible. This has the effect of a kind of local feature selection (which can turn into a global one if certain parameters belonging to the same input variable are not significant in any rule) and hence serves as an additional interpretability aspect as well as regularization property.

The last two steps guarantee a reasonable tradeoff between accuracy and complexity resp. readability of the achieved models, as we optimize the rules resp. consequent parameters according to an optimization criterion by applying synchronously constraints on the number of significant rules resp. significant features in the consequents. This also means that this technique does not need any computational intensive post-processing techniques (such as e.g. [20]) for reducing complexity. In the next three sections we describe how to achieve these three goals.

3 Estimating Antecedent Parts of the Rules

The first step is to yield a proper estimation of the non-linear premise parameters c and σ in the Gaussian fuzzy sets of the rules antecedent parts. For doing so, we apply a clustering approach, which is called iterative vector quantization as iterating over the whole data set a multiple times, and associate each cluster to one rule. The fuzzy sets of the rules antecedent parts are obtained by projection onto the one-dimensional axis: the centers of the fuzzy sets are the center components of a cluster in each dimension, the widths are achieved by calculating the variance in each dimension. In principle, any clustering method requiring a fixed number of cluster as input and delivering cluster prototypes = centers of fuzzy sets and rules as well as the range of influence of the clusters in each dimension can be applied here. In fact, the projection concept is also carried out in many other approaches such as genfis2 [6] or FMCLUST [2] and hence not new. The difference in this paper is that we are applying vector quantization [12] where the

ISBN: 978-989-95079-6-8

a-forementioned methods apply other clustering algorithms (such as subtractive clustering and Gustafson-Kessel).

First, we estimate an initial number of rules (=clusters) and send these to the iterative vector quantization algorithm for fine tuning the non-linear parameters. The number of initial rules are

- Either pre-defined according to user knowledge, interpretability or computational aspects (usually $<$ 100 rules).
- Or automatically extracted from the data by using an evolving clustering approach (*eVQ* [16]) which starts with 0 clusters and evolve new ones according to the distribution and characteristics of new data.

Then, we apply the iterative vector quantization (VQ) algorithm (please refer to [12]) plus estimate the ranges of influence of each cluster in each dimension after the end of VQ.

4 Rule Selection by Optimizing the Rule Weights

Once having the centers and widths of the predefined number (upper bound) of rules estimated, it is a big challenge to reduce the rules in order to enhance interpretability by synchronously not loosing significant accuracy. In this sense, we propose to optimize the rule weights as defined in (2) within a iterative optimization process using sparsity constraints. In a sense, we try to find a minimal least squares error solution subject to the constraint that a significant number of rule weights should be zero. In an optimization formulation a goal would be to minimize the least squares error (3) subject to the constraint that the number of weights ρ_i should be bounded:

$$\min J(w,c,\sigma,\rho) \quad \text{such that} \#\{\rho_i \ / \tag{4}$$

where $\#$ denotes the cardinality of a set and K is an appropriate constant. A minimizer of such a problem is expected to achieve a small error together with a small number of nonzero rule weights. However, there is an intrinsic difficulty with this formulation, as this is a combinatorial optimization procedure (note that we only bound the number of nonzero elements, there is no restriction, where the nonzero elements are). A method that solves such a problem would have to test all subsets of $\{1,\dots N\}$ as candidates for the nonzero entry index of ρ and hence requires exponential complexity in N. However, recently the surprising fact came up, that an almost optimal solution to (4) can be computed by l^1-minimization [7]: That is, instead of (4) the problem

$$\min J(w,c,\sigma,\rho) \quad \text{such that} \sum_{m=1}^{N} |\rho_i| \leq K, \tag{5}$$

is solved. This is a much more accessible approach, because the constraints are convex and there are efficient methods to solve it.

Our algorithm is a nonlinear version of the well-known iteration proposed by [11]. The iteration is basically a projected gradient descent algorithms. The gradient of least squares functional with respect to the jth rule weight ρ can be calculated to:

$$\frac{\partial y_{fuz}}{\partial \rho_j} = \sum_{n=1}^{C}\sum_{k=1}^{N} e_k l_n(x_k) \frac{\partial \Psi_n(x_k)}{\partial \rho_j} \tag{6}$$

with $l_n(x_k)$ the linear rule consequent functions in sample x_k, e_k the residual in sample x_k defined by:

$$e_k(c,\sigma,\rho) := y_k - \sum_{i=1}^{C} l_i(\vec{x}_k)\Psi_i(c,\sigma,\rho)(\vec{x}_k)) \tag{7}$$

and $\frac{\partial \Psi_n(x_k)}{\partial \rho_j}$ the partial derivative of the nth basis function with respect to ρ_j:

$$\frac{\partial \Psi_i}{\partial \rho_k} = \frac{\mu_i}{\sum_{j=1}^{C} \rho_j \mu_j(x)} \left(\delta_{i,k} - \frac{\rho_i \mu_k}{\sum_{j=1}^{C} \rho_j \mu_j(x)} \right) \tag{8}$$

where

$$\delta_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

The idea of the projected gradient method is to apply a threshold operator on the rule weight update in each gradient descent iteration step T_α. Additionally, we also have to optimize the consequent parameters in each update step as needed in $l_n(x_k)$ and e_k. This is done by the local learning approach, estimating the parameters for each rule separately and independently. This has some robustness [18], computational, flexibility [1] and interpretability [24] advantages over global learning approach. In this sense, the whole rule weight optimization procedure becomes as in algorithm 1.

Algorithm 1 Rule Weight Optimization with Sparsity Constraints

1. Initialization of w^0 is done by using local learning with (11) on the initial number of rules (see Step 5 below); ρ^0 is set to a vector of ones (all rules contribute equally at the beginning).
2. w^k, ρ^k are given in iteration k.
3. Calculate $\frac{\partial \Psi_n}{\partial \rho_j}$ as given in (8).
4. Update the rule weights $j = 1,...,C$ componentwise separately by:

$$\rho_j^{k+1} = T_\alpha \left(\rho_j^k - \tau \sum_{n=1}^{C}\sum_{k=1}^{N} e_k l_n(x_k) \frac{\partial \Psi_n(x_k)}{\partial \rho_j} \right) \tag{9}$$

with τ a small constant defining the step size per iteration (usual setting $\tau = 0.1$) and T_α the soft-thresholding operator defined by:

$$T_\alpha = \max(\rho_j^{k+1}, \alpha) - \alpha \tag{10}$$

with α set to a small constant (usual guess is 0.1).

ISBN: 978-989-95079-6-8

5. Calculate $\Psi(c, \sigma, \rho^{k+1})$ for the new $\rho_j^{k+1}, j = 1, ..., C$ and generate the regression matrix $\sqrt{Q_i}R_i, i = 1, ..., C$ for each rule separately with R_i containing the original variables and a column of ones for the intercept and the weighting matrix Q_i defined by $diag(\Psi_i(\vec{x}(k))), k = 1, ..., N$.

6. Perform the weighted least squares approach for the linear consequent parameters whose analytical solution in closed form is:

$$\hat{\vec{w}_i} = (R_i^T Q_i R_i)^{-1} R_i^T Q_i \vec{y} \tag{11}$$

   Here, we also apply Tikhonov regularization [21] if required, i.e. when the condition of the matrix $R_i^T Q_i R_i$ is very high, indicating a matrix close to singular and hence an unstable solution in (11).

7. If the difference between two iterations $\|\rho^{k+1} - \rho^k\|$ is low enough or a pre-defined maximal number of iterations is reached, then stop, otherwise perform next iteration (goto Step 2).

After this optimization procedure, the rules with weights close to 0 are eliminated and the remaining rules C_{rem} are kept for further processing through the last step (described in the subsequent section). Note the in view of our derivation of this iteration, the output of this method should be on the one hand close to the data (this is done by minimizing the least squares functional), and have many rule weights as zero, (which can be discarded afterwards).

In order to visualize the update progress of the rule weights through Algorithm 1, Figure 1 shows the rule weights from three iterations (1, 7 and 13). The weights are sorted in order to get an idea how they are reduced through the optimization procedure by applying T_α. Clearly, after the first iteration almost all rule weights are bigger than zero, whereas after the 13th iterations more than 60 rules can be deleted.

5 Feature Selection by Learning of Consequent Parameters with Sparsity Constraints

Now the final step in our algorithm is to perform a final estimation of the linear consequent parameters $\vec{w}_i, i = 1, ...C_{rem}$ for all rules separately. Indeed, the outcome of Algorithm 1 contains not only the rule weights but also estimates the consequent parameters: these are optimized with standard weighted least squares method (and eventually plus a regularization term) in order to minimize least squares error. Here, we want to go a step further and make the consequent parameters also sparse, i.e. forcing as many of these as possible towards zero, again without loosing (significant) accuracy as optimized in an iterative optimization procedure called Semi-smooth Newton method. This evoke a local feature selection, as variables with low weights may be ignored. Furthermore, if through this selection scenario it turns out that certain variables have very low weights in all rules, it can be concluded that these ones are not necessary for the whole global fuzzy model at all.

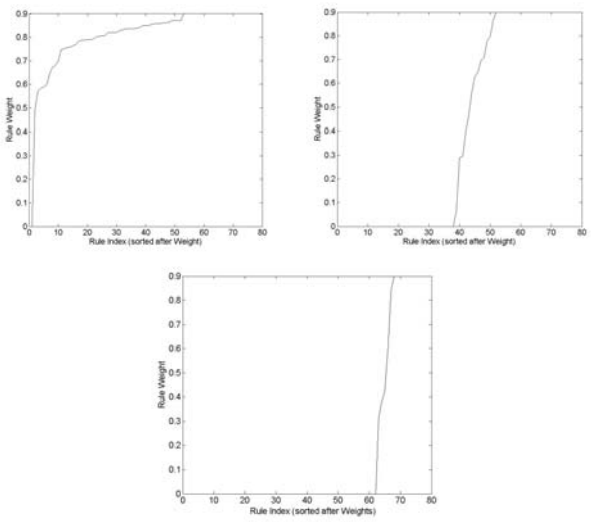


Figure 1: From left to right: sorted rule weights after iteration step 1, 7 and 13; note how the rule base is sparsed out from iteration to iteration (more and more rule weights become 0)

The consequent parameter $\vec{w}_i$ appear linearly in the Function F, in contrast to the rule weights. Although adding sparsity constraints make the optimization a nonlinear problem, because of the linear structure it seems favorable to use a Newton-type method for this problem. Despite the fact that a sparsity constraint problem is usually not differentiable, a generalized version of derivative (the slant derivative) can be defined and subsequently a Newton-type iteration (the Semi-smooth Newton method, [13]) can be applied. The main steps are a soft shrinkage operation on the residual and a Newton step, which makes use of active and inactive sets, see Algorithm 2. For the sake of simplicity, we define this algorithm for the ith rule. Local convergence for this iteration has been shown e.g. in [13].

Algorithm 2 Semi-Smooth Newton with Sparsity Constraints
Let $\sigma > 0$ and parameter k be an iteration index, and M_i the following covariance matrix for local learning

$$M_i = R_i^T Q_i R_i$$

then we apply the following steps:

1. Initialization is done by simply taking the latest estimated $\vec{w}$ (in the last iteration of Algorithm 1).
2. For the kth step we have already $\vec{w}_i^k$ evaluated
3. Step $k \to k + 1$:
   Calculate

$$\vec{u}_{ik} := \vec{w}_i^k - \sigma(R_i^T Q_i (R_i \vec{w}_i^k - \vec{y}))$$

4. Elicit in $\vec{u}_{ik}$ those indices which are greater or smaller than σ (with σ a small value, usually set to 0.1 and not sensitive):

$$\mathcal{A} := \{j : |(\vec{u}_i^k)_j| > \sigma\} \quad \mathcal{I} := \{j : |(\vec{u}_i^k)_j| \leq \sigma\}$$

ISBN: 978-989-95079-6-8

5. Calculate the residual by applying the sparsity constraint operator T_α as defined in (10):

$$\vec{r}_i^{\,k} = T_\alpha(\vec{w}_i^{\,k} - \sigma(R_i^T Q_i(R_i \vec{w}_i^{\,k} - \vec{y})))$$

6. Calculate the update: for all indices on the inactive set:

$$(update_I)_j = -(\vec{r}_i^{\,k})_j \quad \text{für all } j \in \mathcal{I}$$

7. Divide matrix M_i in active and inactive sets:

$$M_{\mathcal{A},\mathcal{A}} := (M_i)_{k,l}, \quad k,l \in calA$$

$$M_{\mathcal{A},\mathcal{I}} := (M_i)_{k,l}, \quad k \in calA, l \in calI$$

8. Calculate update on the inactive set:

$$(update_A)_j = (M_{\mathcal{A},\mathcal{A}})^{-1}\left(\frac{1}{\sigma}(-\vec{r}_i^{\,k})) - M_{\mathcal{A},\mathcal{I}} update_I\right)$$

9. Assign the consequent vector the current update vector one times for the active, one times for the inactive set:

$$\vec{w}_i^{\,k+1} = update$$

10. **If** the residual $\|\vec{r}_i^{\,k}\|$ is large and the iteration smaller than the maximal # of iterations, goto Step 2

An example of the impact of this algorithm on the development of the sparseness of the consequent parameters and hence local/global feature selection will be demonstrated in Section 7.

6 SparseFIS - The Algorithm

The whole *SpareFIS* method is simply obtained by combining the three algorithms described in the previous three sections to one, so applying iterative vector quantization first, then rule weight optimization and finally semi-smooth Newton method for sparse learning of consequent parameters. The explicit formulation is not made here due to space restrictions.

7 Evaluation

7.1 Experimental Setup

The evaluation of the novel method *SparseFIS* includes a comparison with other famous fuzzy batch learning algorithms such as *genfis2* [6] [23], *ANFIS* [15], *FMCLUST* [2] and the batch off-line version of the evolving fuzzy system approach *FLEXFIS* [17], which are all available as implementations in MATLAB. The evaluation will be based on five different high-dimensional data sets for regression: 1.) four data sets from the UCI repository [1] (clean, noise-free): auto-mpg, housing, forrest fires and concrete and 2.) one data set from an engine test bench, where measurement data were collected on-line and is affected by some white noise and disturbances.

The engine test bench data set includes various measurement channels together with their time delay (up to 10), the task was to build a high-performance k-step ahead prediction model for the emission channel NOX for two purposes: 1.) in order to save expenses for the sensor channel and 2.) to perform early recognition of faults (pipe leakage, sensor overheating, interface defects etc.) when using this model in a multi-channel fault detection approach.

A summary of the characteristics of all the data sets used in this evaluation is shown in Table 1. In all data set cases, the evaluation will be based on the accuracy of the models obtained by applying 10-fold cross-validation with a coupled best parameter grid search scenario. The later means that a parameter grid for the most essential parameter(s) in each method is defined consisting of a lower bound, an upper bound and a step-size and for each grid point cross-validation is performed. The best solution with respect to the CV-error (averaged mean absolute error over all folds or averaged maximal error over all folds) will be reported. For *SparseFIS* we will observe the impact of the sparsity constraints on local variable selection (per rule separately). The parameter grids were defined in the following way:

- *genfis2*: radius cluster is varied from 0.1 to 0.9 in steps of 0.05
- *ANFIS*: an initial model is constructed with genfis1 by specifying the number of fuzzy sets per input dimension varied from 2 to 5 in steps of 1
- *FMCLUST*: the number of rules is varied from 1 to 50 in steps of 1
- *FLEXFIS*: the vigilance parameter responsible for the number of clusters = rules is varied from 0.1 to 0.9 in steps of 0.05

For *SparseFIS*, we set the number of maximal allowed (initial) rules to 50 and τ to 0.05 for all experiments, requiring no parameter grid at all. All other parameters are set as mentioned in above algorithms.

7.2 Results

The obtained results are reported in Table 2. The rows represent the different methods and the columns the different data sets: the first number is the CV-error, i.e. the minimal averaged mean absolute error (averaged over the CV-folds, minimal over all parameter settings) plus-minus the standard deviation over these folds. The second number denotes the maximal error over all samples for the minimal averaged mean absolute error (so worst-case prediction error). The third number reports the complexity of the fuzzy models in terms of the number of rules (belonging to the model with minimal CV-error). The results on *SparseFIS* are reported two times, one time by using the full algorithm, one time without applying the sparsity constraint threshold T_α, so without forcing rule weights and consequent parameters towards zero (denoted as *SparseFIS uncon.*). From this table it can be realized that the full *SparseFIS* approach (with constraints) can outperform all other methods in three data sets (forrest fires, housing and NOX) in terms of mean absolute error (MAE), in the case of forest fires

[1] http://archive.ics.uci.edu/ml/

ISBN: 978-989-95079-6-8

Table 1: Some data sets from the UCI repository and their characteristics

	# Training Samples	# Input Variables	Source	Noise Level
Auto-MPG	398	8	UCI	None
Concrete	1030	8	UCI	None
Forrest Fires	517	12	UCI	None
Housing	506	13	UCI	None
NOX	667	181	Engine Test Bench	Medium to high

Table 2: Comparison of fuzzy modelling variants on four data sets from UCI repository and (noisy) NOX data from an engine test bench

Method	**Auto-MPG**			**Concrete**			**Forest Fires**		
	MAE±STD	Max	#rules	MAE±STD	Max	# rules	MAE±STD	Max	# rules
genfis2	2.23± 0.85	3.88	6	8.37± 1.70	11.52	3	19.64± 15.50	58.36	3
ANFIS	2.41± 0.84	4.07	16	11.25± 9.98	39.13	8	20.27± 15.60	58.38	2
FMCLUST	2.35± 0.91	3.99	20	7.75± 2.06	12.11	3	19.83± 15.52	58.93	2
FLEXFIS	**2.17± 0.73**	3.59	11	**7.73± 1.97**	11.96	8	19.69± 15.32	58.35	3
SparseFIS	2.20± 0.30	2.53	8	8.91± 4.55	21.01	6	**15.13± 6.76**	30.77	5
SparseFIS uncon.	2.48± 0.80	4.19	21	11.02± 5.95	25.55	14	20.13± 15.50	59.53	9

Method	**Housing**			**NOX**		
	MAE±STD	Max	#rules	MAE±STD	Max	# rules
genfis2	3.14± 1.31	6.53	4	12.97± 1.08	15.98	5
ANFIS	3.59± 1.39	6.56	4	13.72± 1.01	14.72	8
FMCLUST	2.84± 1.08	5.38	6	13.99± 1.05	15.55	6
FLEXFIS	2.98± 1.27	6.09	6	12.96± 0.98	14.31	8
SparseFIS	**2.84± 0.98**	5.65	6	**12.95± 0.97**	15.22	7
SparseFIS uncon.	3.24± 1.25	6.37	20	13.59± 0.83	14.76	27

even significantly. In the case of the auto-mpg data set, the new method performs similar than the other approaches, whereas for the concrete data set the *SparseFIS* is worse than the others. A similar interpretation of the results can be reported when inspecting the maximal error rate. The standard deviations shows us how sensitive the method is with respect to variations in the training data samples. Lower values represent less variations and hence more robust methods. The best performing methods for each data set with respect to the mean absolute error are shown in bold face. It is also worth mentioning that *SparseFIS* in its unconstrained version performs (mostly) weaker as the full approach; an explanation of this occurrence is that the short version tends to over-fit, as it does not reduce any rules or consequent parameters (and hence complexity).

Another important aspect when training fuzzy systems from data is a reasonable complexity of the obtained models in order to guide the models to more interpretable power [5]. Otherwise, the benefit of fuzzy models over other modelling techniques (such as neural networks, genetic programming etc.) is diminished. We measure the complexity of the final fuzzy models by the number of generated rules by the learning procedure applied onto the whole data set with the optimal parameter setting as obtained in the best parameter grid search and CV procedure. This number is reported in the third columns of each single part in Table 2. Here, we can realize that *SparseFIS* can compete with the other methods, although *genfis2* is mostly the best performer with respect to the number of generated rules. It is also remarkable that the numerical rule weight optimization concept as described in Algorithm 1 is able to reduce the number of rules from 50 to below 10 for all data sets (and this with a reasonable accuracy), confirming Figure 1.

Finally, in Figure 2 we inspect the impact of the sparsity constraints onto the linear consequent parameters as applied in Algorithm 2 (Step 4). Therefore, we show a plot of consequent parameters for all extracted rules (=rows) for the second (upper row) and the third features (lower row) in case of auto-mpg and using all input variables in *SparseFIS* method. While for the second feature four rules are sparsed out (compare left and right image in upper row), for the third feature all consequent parameters are forced to a small value around 0 which is not significant (note the range of this feature is 184). This means that feature #3 is not significant for modelling the target in the auto-mpg data set.

ISBN: 978-989-95079-6-8

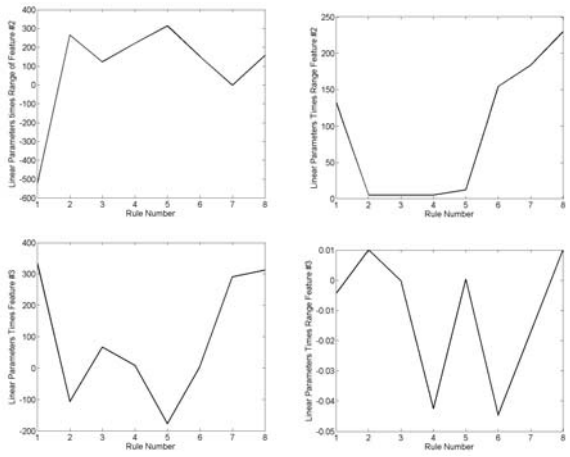


Figure 2: Left: Linear parameters times the ranges for features #2 (upper) and #3 (lower) when not applying sparsity constraints in consequent learning; right: the same as in left, but applying sparsity constraints

8 Conclusion

In this paper, we presented a novel data-driven learning technique for Takagi-Sugeno type fuzzy systems, called *SparseFIS*, which is able 1.) to perform rule selection together with consequent parameter estimation in a coherent optimization procedure and 2.) to gain more insight into the fuzzy models by yielding a local feature selection through sparsing out unimportant features in each rule. As a top-down method, it is possible to elicit an initial upper bound of number of (allowed) rules by expert-knowledge or by initial guess from the data. Tests on various multi-dimensional data sets from UCI repository and real-world applications show that the novel method can compete with other well-known learning methods in terms of both, accuracy and complexity and synchronously is able to put in more interpretation in from of local importance of the features.

References

[1] P.P. Angelov, E. Lughofer, and X. Zhou. Evolving fuzzy classifiers using different model architectures. *Fuzzy Sets and Systems*, 159(23):3160–3182, 2008.

[2] R. Babuska. *Fuzzy Modeling for Control*. Kluwer Academic Publishers, Boston, 1998.

[3] J. Botzheim, E. Lughofer, E.P. Klement, L. Kóczy, and T.D. Gedeon. Separated antecedent and consequent learning for takagi-sugeno fuzzy systems. In *Proceedings of FUZZ-IEEE 2006*, pages 2263–2269, Vancouver, Canada, 2006.

[4] M. Burger, J. Haslinger, U. Bodenhofer, and H. W. Engl. Regularized data-driven construction of fuzzy controllers. *J. Inverse Ill-Posed Probl.*, 10(4):319–344, 2002.

[5] J. Casillas, O. Cordon, F. Herrera, and L. Magdalena. *Interpretability Issues in Fuzzy Modeling*. Springer Verlag, Berlin Heidelberg, 2003.

[6] S. Chiu. Fuzzy model identification based on cluster estimation. *Journal of Intelligent and Fuzzy Systems*, 2(3):267–278, 1994.

[7] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *J. Amer. Math. Soc.*, 22(1):211–231, 2009.

[8] 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.

[9] O. Cordon and F. Herrera. Identification of linguistic fuzzy models by means of genetic algorithms. *Fuzzy Model Identification. Selected Approaches*, pages 215–250, 2003.

[10] O. Cordon, F. Herrera, F. Hoffmann, and L. Magdalena. *Genetic Fuzzy Systems — Evolutionary Tuning and Learning of Fuzzy Knowledge Bases*. World Scientific, 2001.

[11] Ingrid Daubechies, Michel Defrise, and Christine De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. Pure Appl. Math.*, 57(11):1413–1457, 2004.

[12] R.M. Gray. Vector quantization. *IEEE ASSP Magazine*, pages 4–29, 1984.

[13] R. Griesse and D.A. Lorenz. A semismooth Newton method for Tikhonov functionals with sparsity constraints. *Inverse Problems*, 24(3):19 p., 2008.

[14] D. Gustafson and W. Kessel. Fuzzy clustering with a fuzzy covariance matrix. In *Proc. IEEE CDC*, pages 761–766, San Diego, CA, USA, 1979.

[15] J.-S.R. Jang. ANFIS: Adaptive-network-based fuzzy inference systems. *IEEE Trans. Syst. Man Cybern.*, 23:665–685, 1993.

[16] E. Lughofer. Extensions of vector quantization for incremental clustering. *Pattern Recognition*, 41(3):995–1011, 2008.

[17] E. Lughofer. FLEXFIS: A robust incremental learning approach for evolving TS fuzzy models. *IEEE Trans. on Fuzzy Systems (special issue on evolving fuzzy systems)*, 16(6):1393–1410, 2008.

[18] E. Lughofer. Towards robust evolving fuzzy systems. In P. Angelov, D. Filev, and N. Kasabov, editors, *Evolving Intelligent Systems: Methodology and Applications*, page to appear. John Wiley & Sons, New York, 2009.

[19] D. Nauck and R. Kruse. Nefclass-x – a soft computing tool to build readable fuzzy classifiers. *BT Technology Journal*, 16(3):180–190, 1998.

[20] M. Setnes. Simplification and reduction of fuzzy rules. In J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, editors, *Interpretability Issues in Fuzzy Modeling*, volume 128 of *Studies in Fuzziness and Soft Computing*, pages 278–302. Springer, Berlin, 2003.

[21] A.N. Tikhonov and V.Y. Arsenin. *Solutions of ill-posed problems*. Winston & Sonst, Washington D.C., 1977.

[22] L.X. Wang. Fuzzy systems are universal approximators. In *Proc. 1st IEEE Conf. Fuzzy Systems*, pages 1163–1169, San Diego, CA, 1992.

[23] R.R. Yager and D.P. Filev. Learning of fuzzy rules by mountain clustering. In *Proc. of SPIE Conf. on Application of Fuzzy Logic Technology*, pages 246–254. Boston, MA, U.S.A., 1993.

[24] J. Yen, L. Wang, and C.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.

ISBN: 978-989-95079-6-8

Increasing the Performance of Fuzzy Retrieval Using Impact Ordering

Carlos D. Barranco[1] Sven Helmer[2]

[1]Division of Computer Science, School of Engineering, Pablo de Olavide University
Seville, Spain
[2]School of Computer Science and Information Systems, Birkbeck, University of London
London, United Kingdom
Email: cbarranco@upo.es, sven@dcs.bbk.ac.uk

Abstract— *We propose an approach for indexing fuzzy data based on inverted files that speeds up retrieval considerably by stopping the traversal of postings lists early. This is possible because the entries in the postings lists are organized in a way that guarantees that there are no matching items beyond a certain point in a list. Consequently, we can reduce the number of false positives significantly, leading to an increase in retrieval performance. We have implemented our approach and evaluated it experimentally, comparing it to an approach that has previously been shown to be superior to other methods.*

Keywords— fuzzy databases, access methods, inverted files, physical design

1 Introduction

Being able to handle imprecise or uncertain data becomes ever more important in today's world. There are numerous applications that have to manage imperfect data in areas such as knowledge discovery, mediator systems, information retrieval, multimedia, and profiling (vague) users' preferences. One way to model imprecision and uncertainty is to employ concepts from fuzzy set theory. Fuzzy sets are making their way into the database world, evidenced by the fact that several proposals for fuzzy database models and systems exist [1, 2, 3, 4, 5] (for an overview see [6]).

We are focusing on a well-known model for fuzzy databases, namely the possibilistic database model [5]. This approach uses the concepts of possibility and necessity measures for flexible querying. While this is an elegant way to formulate queries, indexing possibility distributions is not straightforward. Bosc and Galibourg introduced an indexing principle based on α-cuts (i.e. the elements of a fuzzy set with a membership degree of at least α) [7]. The parameter α is tied directly to a threshold value determined by a user defining a query, which makes it unrealistic to materialize dozens or even hundreds of α-cuts in different indexes to be able to answer various queries efficiently. For that reason Bosc and Galibourg developed a filtering mechanism based on the supports and cores of the indexed possibility distributions[1]. However, this can introduce a large number of false positives, as the filter only acts as a "quick-and-dirty" test. All the candidates that pass the filter have to be checked for eligibility in a second step.

We propose an access method for indexing possibility distributions that cuts down the number of false positives significantly. Many different α-cuts are stored compactly in a single index structure by arranging the references to possibility distributions in a clever way. The idea is based on impact ordering of inverted files used in information retrieval (IR). In IR systems, document references are sorted in descending order according to the term frequency value to make sure that important documents for a query are processed first [8]. In our case references to possibility distributions are sorted according to the certainty degrees of their elements, making it possible to stop scanning lists of references at an early stage (because we know that no eligible data items can follow). By applying customized gap compression schemes to the inverted files, we can store the lists very compactly.

The remainder of this paper is organized as follows. In the section immediately following this introduction, we cover basic definitions of possibilistic database systems and how to query them. In Section 3 we describe our access method in more detail and explain how to answer queries based on the possibility measure and those based on the necessity measure. Section 4 includes a brief description of the approach that we compared ours to. This is followed by a section on the experimental evaluation, specifying the environment in which this evaluation was run and presenting the results of the evaluation. We wrap up the paper with a conclusion and outlook in Section 6.

2 Preliminaries

Before going into details about indexing, let us briefly define possibility distributions and their application in flexible querying as defined in [5].

2.1 Possibilistic Databases

We assume (without loss of generality) n data items $(o_1, o_2, \ldots, o_n)$ in our database, all of which have an attribute A with a discrete domain Ω. The value of attribute A of data item o_i is described by a possibility distribution function $\pi_{A(o_i)}$ on Ω.

The possibility distribution $\pi_{A(o_i)}$ is a set of possible values for the attribute A of the data item o_i (together with the certainty of each value) and is defined as:

$$\pi_{A(o_i)} : \Omega \mapsto [0, 1] \tag{1}$$

$\pi_{A(o_i)}(\omega_j) = 0$ (for $\omega_j \in \Omega$) means that it is *impossible* that the attribute A can take on the value ω_j. $\pi_{A(o_i)}(\omega_j) = 1$, on the other hand, means that it is *completely* possible that A can take on the value ω_j. If we want to express that it might

[1]Supports and cores are special α-cuts.

ISBN: 978-989-95079-6-8

ω_1	$\rightarrow$	$\alpha_{1,1}$	$r(o_1^{1,1}), r(o_2^{1,1}), r(o_3^{1,1}), \ldots$	$\alpha_{1,2}$	$r(o_1^{1,2}), r(o_2^{1,2}), r(o_3^{1,2}), \ldots$	$\ldots$	$\ldots$
ω_2	$\rightarrow$	$\alpha_{2,1}$	$r(o_1^{2,1}), r(o_2^{2,1}), r(o_3^{2,1}), \ldots$	$\alpha_{2,2}$	$r(o_1^{2,2}), r(o_2^{2,2}), r(o_3^{2,2}), \ldots$	$\ldots$	$\ldots$
$\vdots$							
ω_m	$\rightarrow$	$\alpha_{m,1}$	$r(o_1^{m,1}), r(o_2^{m,1}), r(o_3^{m,1}), \ldots$	$\alpha_{m,2}$	$r(o_1^{m,2}), r(o_2^{m,2}), r(o_3^{m,2}), \ldots$	$\ldots$	$\ldots$

Figure 1: Inverted file index for fuzzy retrieval

be possible for A to take on the value ω_j, then we can assign a value from the interval $(0, 1)$ to $\pi_{A(o_i)}(\omega_j)$, depending on how plausible we think it is that A takes on this value. Not that the possible values for A are mutually exclusive, i.e. A takes on a single value, we are just uncertain about which one is the correct value. For reasons of consistency, $\pi_{A(o_i)}$ should be normalized, i.e. $\exists\omega \in \Omega : \pi_{A(o_i)}(\omega) = 1$.

As already mentioned, we assume that Ω is discrete, since we are interested in indexing scalar data. In the case of applications using continuous numerical data, we either have to discretize the possibility distribution or use a different kind of index structure [9]).

2.2 Flexible Querying

We not only allow imprecision in the data, but also flexible querying, i.e. the query condition c determining the acceptable values of an attribute is described by a normalized fuzzy set $\mu_c : \Omega \mapsto [0, 1]$.

In order to be able to check whether a data item satisfies a query condition, we need to introduce the concept of a fuzzy measure. Let X be an element of the power set of Ω ($X \in \mathcal{P}(\Omega)A$). The *possibility* of X, $\Pi(X)$, is measured by looking at the elements of X:

$$\Pi(X) = \max_{\omega \in X} \pi(\omega) \tag{2}$$

Possibility theory uses two concepts to measure the likeliness of X: the possibility measure, as already described above, $\Pi(X)$ and the necessity measure $N(X)$. The *necessity* of X, $N(X)$, is defined by the unlikeliness of the complement of X:

$$N(X) = 1 - \Pi(\overline{X}) \tag{3}$$

What does this mean for an attribute value satisfying a query condition c? The possibility of doing so is measured with the help of the possibility measure:

$$\Pi(c|A(o_i)) = \max_{\omega \in \Omega} \min(\mu_c(\omega), \pi_{A(o_i)}(\omega)) \tag{4}$$

Informally speaking, $\Pi(c|A(o_i))$ returns the highest degree to which $A(o_i)$ can possibly satisfy c.

The necessity of an attribute value satisfying query condition c is defined as:

$$N(c|A(o_i)) = \min_{\omega \in \Omega} \max(\mu_c(\omega), 1 - \pi_{A(o_i)}(\omega)) \tag{5}$$

$N(c|A(o_i))$, on the other hand, returns the degree to which $A(o_i)$ certainly satisfies c.

Usually users are interested in a small subset of data items in the database. Selective queries can be formulated by providing an acceptance threshold α (e.g. return all data items whose attribute values possibly satisfy c to at least a degree of 0.8: $\{o_i|\Pi(c|A(o_i)) \geq 0.8\}$). Furthermore, using the necessity measure also leads to more selective queries than using the possibility measure.

3 Indexing Possibility Distributions

As mentioned in Section 1 we use an inverted file index to index possibility distributions. An inverted file consists of a *directory* containing all distinct values $\omega_1, \omega_2, \ldots, \omega_m$ of the domain Ω and a *list* (also called a *postings list*) for each value.[2] Within each postings list of the inverted file, we sort the references to data items o_j, denoted by $r(o_j)$, in descending order of the degree of membership of w_i in $\pi_{A(o_j)}$. Since we want to apply gap compression (more on this later in Section 3.3), we arrange the references in blocks, each block preceded by a value $\alpha_{i,k}$, meaning that this is the k-th block of the posting list associated with ω_i. All attributes of the data items referenced in a block satisfy $\pi_{A(o_j)}(\omega_i) \geq \alpha_{i,k}$. Figure 1 illustrates the layout described above; note that the data items have superscripts indicating which block they belong to. The threshold values $\alpha_{i,k}$ divide up the interval [0,1] into equally-sized partitions (i.e. each block covers an equally-sized interval). Within each block all references to data items (in the form of IDs) are sorted in ascending order.

Depending on the query type, we access the inverted file index slightly differently. We will first look at queries based on the possibility measure and then turn to those based on the necessity measure.

3.1 Possibility Measure Queries

If we are searching for all data items o_j whose attribute A satisfies condition c possibly to a degree of at least α, we only have to look up the values ω_i in the inverted file for which $\mu_c(\omega_i) \geq \alpha$ (see Equation (4)). In addition to that, only data items for which $\pi_{A(o_j)}(\omega_i) \geq \alpha$ will qualify. During a search, we scan a postings list until we reach the first block whose value $\alpha_{i,l}$ is smaller than the query threshold value α.[3] If α is between $\alpha_{i,l-1}$ and $\alpha_{i,l}$, then we still have to check the data items referenced in this block whether they are false positives or not (as some of them could satisfy the query predicate). However, only the data items in this last block can be false

[2] For an overview of traditional inverted files see [10].

[3] Obviously, we also stop when we reach the end of a postings list.

ISBN: 978-989-95079-6-8

positives and have to be checked, all other data items referenced in earlier blocks will satisfy the query predicate and can be retrieved without checking for false positives. If α is equal to $\alpha_{i,l-1}$ we do not have to check the items of block $\alpha_{i,l}$. Actually, in this case we do not have to check for false positives at all.

3.2 Necessity Measure Queries

Processing queries based on the necessity measure can be done in two different ways, depending on the cardinality of the domain Ω. For large domains we can simply use possibility measure queries as a filter (we will call this method 'simplified' in the following); for small domains we can use the index described above to determine the answer set (without accessing the data items themselves).

The inequality $N(X) \leq \Pi(X)$ always holds for necessity and possibility measures, i.e. $N(c|A(o_i)) \geq \alpha \Rightarrow \Pi(c|A(o_i)) \geq \alpha$. So instead of directly searching for the data items that satisfy the query condition c necessarily to the degree α, we search for data items satisfying c possibly. We retrieve these data items and check if they also satisfy c necessarily. A drawback of this technique is that we introduce false positives, due to the data items that satisfy c possibly but not necessarily.

Accessing all the candidates returned by the possibility measure query in the simplified processing will cause a lot of random I/O. However, we can use the inverted file index to help us in sorting out false positives returned by a possibility measure search. If we can find an ω_i for a data item o_j such that $\max(\mu_c(\omega_i), 1 - \pi_{A(o_j)}(\omega_i)) < \alpha$, then we know that $A(o_j)$ cannot satisfy c necessarily (more on this in just a moment, see also Equation 5). Having a value for $\mu_c(\omega_i)$ that is greater or equal to α will never result in $\max(\mu_c(\omega_i), 1-\pi_{A(o_j)}(\omega_i)) < \alpha$. So in order to sort out false positives, we have to access the postings lists of the values ω_i for which $\mu_c(\omega_i) < \alpha$. Unfortunately, this means we have to look up the complement of the values ω_i used for the possibility measure query and scan the corresponding postings lists.[4] (Consequently, this will only be efficient for relatively small domains.) During query processing, we keep scanning such a postings list in a blockwise fashion as long as $\alpha_{i,l} > 1 - \alpha$ holds and we stop as soon as this condition is not satisfied anymore.[5] Any references to data items found during this scan can be safely discarded from the candidates determined by the possibility measure query. The reasoning is the following: while $\alpha_{i,l} > 1 - \alpha$, we know that for all data items referenced in this block $\pi_{A(o_j)}(\omega_i) > 1 - \alpha$ ($\Leftrightarrow 1 - \pi_{A(o_j)}(\omega_i) < \alpha$) holds. Together with $\mu_c(\omega_i) < \alpha$, this means that $A(o_j)$ cannot satisfy c necessarily. Once $\alpha_{i,l}$ drops below $1 - \alpha$, we do not have enough information to determine whether the data items referenced in this block (and the following blocks) satisfy c necessarily or not and we stop scanning the list further. All the data items remaining in the candidate set after scanning the lists of the ω_i for which $\mu_c(\omega_i) < \alpha$ have to be fetched and checked for false positives.

[4] This is why we will also call this approach complement-based.

[5] Again, we also stop when reaching the end of a list.

3.3 Index Compression

The reason for arranging the references to data items in blocks is that we can use gap encoding on the ascending sequences of data item IDs. For example, instead of storing the sequence 103,110,114,116,121 we store 103,7,4,2,5. Compared to absolute ID numbers, relative gaps can be stored using less storage space. However, from time to time even gaps can become quite large, i.e. a large gap could need almost as much storage space as a regular ID. So that we do not have to allocate the same space for each gap, we use variable length encoding [11]. The main idea is to use few bits for small gaps and many bits for large gaps.

In order to keep the codewords byte-aligned for faster processing, we use Variable-Byte (VB) code [12]. Each byte contains a so-called continuation bit, which signals if the end of the current codeword has been reached or if we have to continue decoding. Usually the high bit is dedicated to this purpose, while the lower 7 bits of each byte encode gap information. If a gap fits into 7 bits, then we encode this gap in the lower 7 bits and set the continuation bit to 1. Otherwise, we encode the highest bits of the gap and set the continuation bit to 0. The remaining bits are encoded in the same manner. VB-encoding is a good compromise between compression ratio and processing speed. For example, encoding 5 in VB-code would result in **1**0000101, while encoding 824 would result in **0**0000110 **1**0111000 (where the bits in boldface are the continuation bits).

Instead of using floating-point numbers for the threshold levels $\alpha_{i,k}$ directly in the index, we store IDs of threshold levels using a one-byte integer, looking up the floating-point values in a table mapping the IDs to these values. Due to the fixed threshold levels we use throughout the whole index, we can do this. As we will see later in the evaluation section, 256 different threshold levels are more than enough. If more than 256 threshold levels should be needed, we can apply VB-encoding to the threshold levels as well.

4 Comparison to Existing Approach

We use the approach by Bosc and Galibourg [7] as a reference for comparison. There are two special α-cuts, the *core* $L_1(\mu_F)$ and the *support* $L_{>0}(\mu_F)$ of a fuzzy set F. Bosc and Galibourg have shown that the following implications hold (for all $\alpha > 0$):

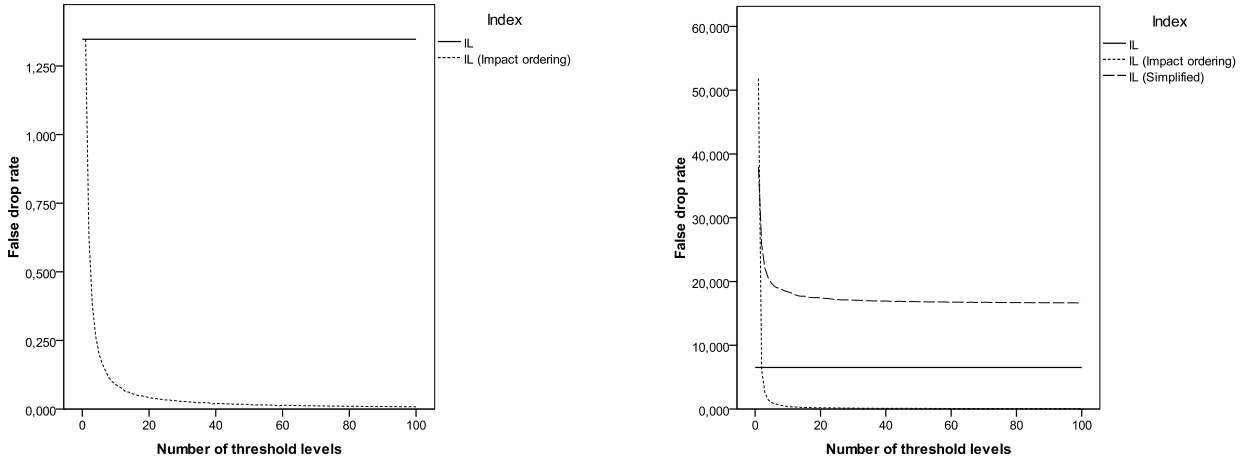
$$\Pi(c|A(o_i)) \geq \alpha \quad \Rightarrow \quad L_{>0}(\pi_{A(o_i)}) \cap L_\alpha(\mu_c) \neq \emptyset \qquad (6)$$

$$N(c|A(o_i)) \geq \alpha \quad \Rightarrow \quad L_1(\pi_{A(o_i)}) \subseteq L_\alpha(\mu_c) \qquad (7)$$

Given a query condition c we have to determine whether the α-cut of μ_c ($L_\alpha(\mu_c)$ intersects with the support of a data item's attribute value (for possibility queries) or if the core of the attribute value is a subset of $L_\alpha(\mu_c)$ (for necessity queries). If this is not the case, then we can safely discard a data item, as it will not satisfy the original query condition.

We implemented this technique using the same framework as for our approach, i.e. using compressed inverted files for indexing the cores and supports of possibility distributions. The index containing the cores stores the cardinality of each core together with the reference to the possibility distribution. In

ISBN: 978-989-95079-6-8



(a) Possibility measure

(b) Necessity measure

Figure 2: False drop rates

this way, superset queries (as described in Equation (7)) can be answered using an inverted file by filtering out all references whose cores have a higher cardinality than the query set. For a more detailed description on how to optimize this kind of data structure, see [13] (comparisons to other data structures can also be found in this paper; the inverted file approach, however, was the data structure with the best performance among all those that were tested).

5 Experimental Evaluation

5.1 Benchmark Environment

The benchmarks were run on a Pentium 4 (3.2 GHz) PC with 1GByte main memory running Windows XP. The index structures and a simulator for the storage system were implemented in Java 1.6. We decided to implement the simulator in Java, as this makes us platform-independent. The simulator allows us to avoid unwanted effects due to caching. We set the block size of the simulator to a page size of 4K, which is the page size of the underlying operating system.

We generated possibility distributions for the data items using the following parameters: the domain size ($|\Omega|$) is equal to 25 and the database size is equal to 100K data items (each data item is associated with a possibility distribution). The elements of a particular possibility distribution are taken randomly from the domain, using a uniform distribution. This means that we first determine the cardinality of the support of a possibility distribution, which is uniformly distributed in [1,22] (22 elements cover 90% of the domain). All the elements in the domain have the same probability of making it into the support. The membership degrees of the chosen elements are uniformly distributed in (0,1.0]. If a possibility distribution is not normalized, the highest degree is converted to 1.0.

In this paper we wanted to focus on typical fuzzy scalar domains, therefore we decided to keep the domain cardinality rather small. This is motivated by the fact that the elements of fuzzy domains usually represent categories. Normally these categories are determined by experts based on the criteria (1) that they should be easy to use by ordinary users and (2) that fuzzy relations (allowing flexible comparisons) can be defined on them. It is much easier to satisfy both criteria with a small domain cardinality; that is why typical fuzzy scalar domains are rather small.

Another important parameter is the clustering of the data items on disk pages. For the results shown in this paper we assume a "worst-case scenario" in terms of indexing, meaning that all the candidate items returned by the index are stored in consecutive blocks with 64.5 data items sharing a block on average. This will lessen the effect that false drops have on the retrieval performance, as we do not need a separate page access to fetch each of the false drops.

5.2 Results

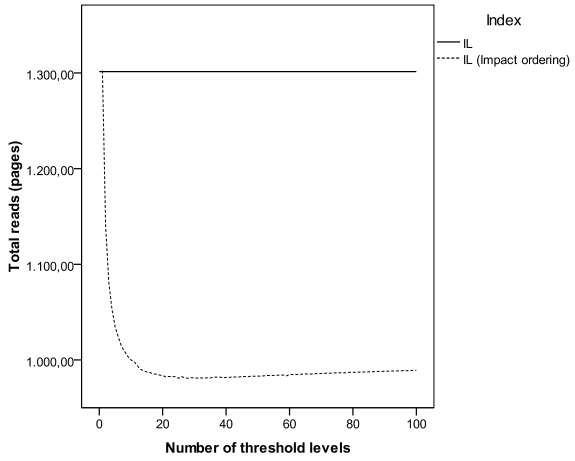
We were interested in several criteria, most importantly in the false drop ratio of the two different approaches and its impact on the query performance. To a lesser extent we also wanted to make sure that the overhead in terms of the index size is not prohibitive. Furthermore, we investigate the impact of the query threshold α on the performance of the indexes. In Sections 5.2.1 to 5.2.3 our competitor is included as a reference; the parameter we vary (number of threshold levels) only applies to our approach in these cases.

5.2.1 False Drop Rate

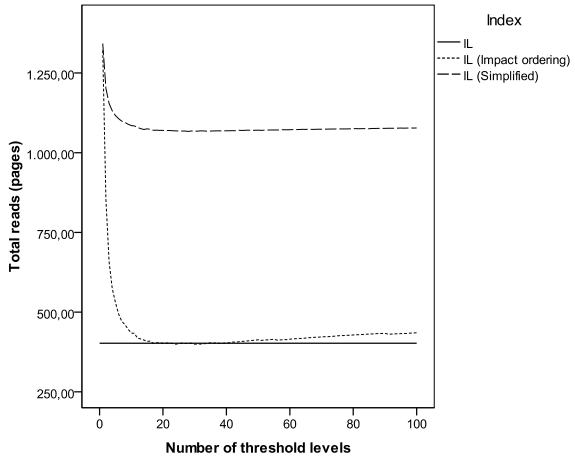
Figure 2 shows the results we obtained for measuring the false drop rate for varying the number of threshold levels in our index (Figure 2(a) depicts the results for possibility measure queries, while Figure 2(b) depicts those for necessity measure queries[6]). Each threshold level covers an equally-sized interval in [0,1]. The false drop rate is measured in the following way: (total number of IDs returned by index - number of true positives)/number of true positives. As can be clearly seen, adding more threshold levels to the data structure reduces the false drop ratio. However, the law of diminishing returns applies to this, i.e. with each added threshold level the improvement gain decreases. We can also notice that using a

[6]The method labeled 'simplified' applies possibility measures as a filter, while the other curve shows the results for the complement-based approach.

ISBN: 978-989-95079-6-8



(a) Possibility measure



(b) Necessity measure

Figure 3: Total number of page accesses

possibility measure as a filter for a necessity measure is not competitive in this scenario.

5.2.2 Retrieval Performance

While the false drop rate suggests to use a very high number of threshold levels, it is not as simple as that to determine the optimal number of threshold levels. By increasing the number of threshold levels we introduce an overhead in terms of index size and retrieval performance (larger indexes result in a longer query processing time). Let us first look at the retrieval performance of the access methods. Figure 3 shows the retrieval performance in total number of page accesses (part (a) for possibility measure queries and part (b) for necessity measure queries). The total number of page accesses includes the page accesses needed to navigate the index and the pages accessed for retrieving all candidate data items (false and true positives).

There is a sharp decrease in the number of page accesses when going from one threshold level to about twenty threshold levels. This is caused by the rapidly dropping false drop ratio. Then the curve peters out and slowly goes up again. In this part of the curve, the drop in performance, due to the aforementioned overhead, gains the upper hand over the slowly decreasing false drop rate. In our benchmark scenario we have identified 25 to be the optimal number of threshold levels for possibility measure queries and 28 to be the optimum for necessity measure queries. Again, the approach using the possibility measure as a filter for necessity measure queries is not competitive, so we will drop it from the graphs in the following sections.

5.2.3 Index Size

In Figure 4 we present the size of the index structures in terms of disk pages. The overhead added for introducing a larger number of threshold levels can be clearly seen. This is caused by a deterioration of the compression ratio. A few larger threshold blocks with smaller gaps between the IDs can be compressed better than a lot of smaller blocks with larger gaps. The stair-like appearance is caused by postings lists expanding by one page at roughly the same time (due to the uniform distribution of domain elements in the possibility distributions). At first glance it seems that our index structure needs more storage space than our competitor. However, as we only need about 20 to 30 threshold levels, this increase is only moderate and, more importantly, we can answer both, possibility measure and necessity measure, queries with the same index. For the other scheme separate indexes are needed (one storing the supports and one storing the cores of the possibility distributions).

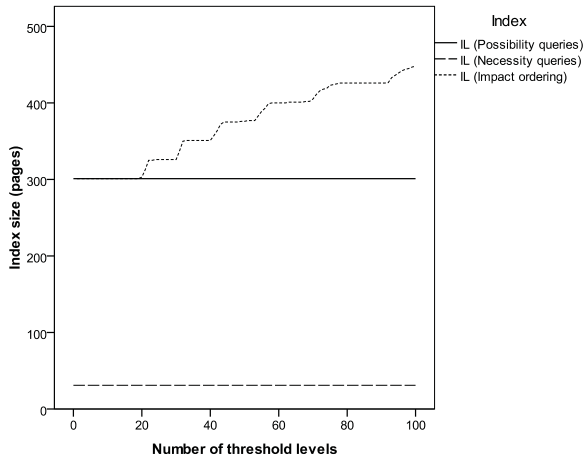
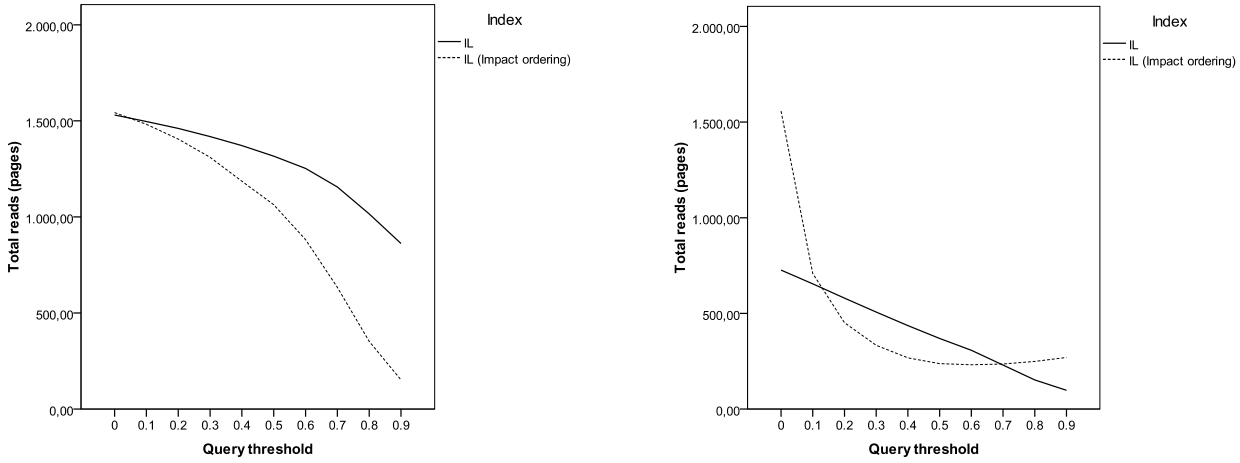


Figure 4: Index size

5.2.4 Impact of Query Threshold

Figure 5 shows the impact of the query threshold (α) on the performance. Generally, as the query threshold grows for possibility measure queries (Figure 5(a)), we have to check smaller and smaller parts of the query set (only elements that have a membership degree larger than or equal to α have to be considered). This in turn decreases the number of postings lists we have to traverse, leading to less page accesses. In addition to this, our approach can stop scanning postings lists earlier for higher query thresholds, leading to further improve-

ISBN: 978-989-95079-6-8



(a) Possibility measure

(b) Necessity measure

Figure 5: Impact of query threshold

ment in terms of the number of page accesses.

For necessity measure queries ((Figure 5(b)), things are more complicated. For the core-filtering approach by Bosc and Galibourg the number of postings lists we have to scan decreases with increasing threshold α. For our approach there are two effects at work. With increasing α the false drop rate drops, which leads to fewer page accesses due to false positives. However, with increasing threshold values the number of page accesses in the index increases. Although the number and lengths of postings list traversals for the possibility measure filter step goes down, the number and lengths of postings list scans for the subsequent complementary check on the necessity measure goes up at quite a fast rate. At some point this effect catches up with us and the number of accessed blocks increases again.

6 Conclusion and Outlook

We have shown how to build an index that stores threshold levels of possibility distributions in a compact way. Being able to represent possibility distributions using a much finer granularity than existing approaches helps us in reducing the number of false positives significantly. Even in an unfavorable environment, our index far outperforms existing approaches for possibility measure queries and was on a par for necessity measure queries. This kind of performance can already be achieved by introducing only a moderate number of threshold levels, meaning that there is practically no overhead, since we only need to maintain one index (while our main competitor needs one for possibility measure queries and one for necessity measure queries).

For future work we would like to analyze how to optimize the number and distribution of threshold levels given a fixed database. Even more interesting would be to optimize and update the index in a dynamic way, given a database with statistics on query workloads.

Acknowledgments

This work has been partially supported by the "Ministerio de Ciencia y Tecnología (MCYT)" (Spain) under grant TIN2007-68084-CO2-01, and the "Consejería de Innovación Ciencia y Empresa de Andalucía" (Spain) under research projects P06-TIC-01570 and P07-TIC-02611. We also thank the anonymous referees for improving the readability of the paper.

References

[1] B. Buckles and F. Petry. A fuzzy model for relational databases. *Fuzzy Sets and Systems*, 7(3):213–226, 1982.

[2] R. Caluwe de, editor. *Fuzzy and Uncertain Object-oriented Databases (Concepts and Models)*, volume 13 of *Advances in Fuzzy Systems - Applications and Theory*. World Scientific, 1997.

[3] R. George, A. Yazici, F.E. Petry, and B.P. Buckles. Uncertainty modeling in object-oriented geographical information systems. In *Third Int. Conf. for Database and Expert System Application (DEXA)*, Valencia, 1992.

[4] F. E. Petry and P. Bosc. *Fuzzy databases: principles and applications*. International Series in Intelligent Technologies. Kluwer Academic Publishers, 1996.

[5] H. Prade and C. Testemale. Generalizing database relational algebra for the treatment of incomplete or uncertain information and vague queries. *Information Sciences*, 34:115–143, 1984.

[6] J. Galindo, A. Urrutia, and M. Piattini. *Fuzzy Databases: Modeling, Design and Implementation*. Idea Group Publishing, Hershey, PA, USA, 2006.

[7] P. Bosc and M. Galibourg. Indexing principles for a fuzzy database. *Information Systems*, 14(6):493–499, 1989.

[8] C.D. Manning, P. Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[9] C. D. Barranco, J. R. Campaña, and J. M. Medina. A B+-tree based indexing technique for fuzzy numerical data. *Fuzzy Sets and Systems*, 159(12):1431–1449, 2008.

[10] R. Sacks-Davis and J. Zobel. *Indexing Techniques for Advanced Database Systems*. Kluwer Academic Publishers, 1997.

[11] I.H. Witten, A. Moffat, and T.C. Bell. *Managing Gigabytes*. Morgan Kaufmann, San Francisco, 1999.

[12] A. Trotman. Compressing inverted files. *Information Retrieval*, 6(1):5–19, 2003.

[13] S. Helmer. Evaluating different approaches for indexing fuzzy sets. *Fuzzy Sets and Systems*, 140(1):167–182, 2003.

ISBN: 978-989-95079-6-8

Colour Image Segmentation using A-IFSs

Pedro Couto, Pedro Melo-Pinto[1] Humberto Bustince, Edurne Barrenechea, Miguel Pagola[2]

1.CITAB Research Centre, UTAD University
Vila Real, Portugal
2.Departamento de Automática y Computación, UPNA University
Pamplona, Spain
Email: {pcouto,pmelo}@utad.pt, {bustince,edurne.barrenechea,miguel.pagola}@unavarra.es

***Abstract*— *The problem of segmentation is an important research field and often a critical and decisive preprocessing step for many image processing applications. During the past years, the use of colour images instead of gray images is growing fast in many research areas. Moreover, the use of segmentation over colour images is advised when additional colour information conveyed is necessary to further discriminate the image regions. Recently, fuzzy logic theory, and Atanassov's intuitionistic fuzzy sets (A-IFSs), have been successfully applied to image segmentation and, in this paper we propose a general methodology for RGB colour image segmentation based on A-IFSs. The proposed methodology is based on a multilevel colour thresholding framework that uses Atanassov's intuitionistic index values for representing the uncertainty present in the assigning of the pixels to the different regions. This framework is to be applied to each RGB component separately to finally aggregate the results of the three components. The number of thresholds is auto established by the methodology based on the image entropy. Experimental results are presented.*

Keywords— Atanassov's Intuitionistic Fuzzy Sets, Color Image Segmentation, Fuzzy Logic, Multi-Thresholding.

1 Introduction

Color image segmentation deals with the subdivision of an image into disjoined regions according to the pixels colour with the goal of finding the image objects. As a result of the segmentation, the original set of pixels in the image is partitioned into a finite set of separated regions (objects).

Within the framework of fuzzy theory [1, 2, 3, 4], segmentation algorithms that use the concept of fuzzy entropy [1, 3, 5] are widely used in image processing applications. In previous works [6] we have proposed methodologies for image segmentation that use A-IFSs [7, 8, 9]. In those methodologies, Atanassov's intuitionistic fuzzy index values are used for representing the uncertainty/impresision on determining whether a pixel of the image belongs to an object of that image.

The general framework presented in [6] is the basis for colour segmentation framework proposed in this paper where it is extended to a multi-level thresholding framework since, often, images include several regions and, thus, it is necessary to compute more than one threshold in order to correctly segment the image.

In this methodology we use RGB images in such way that each RGB component is primarily segmented separately. After the three components are segmented we then aggregate the three segmented components into the final segmented image. For each component, the number of computed thresholds is automatically determined by the methodology on the basis of the images intuitionistic fuzzy entropy. This entropy is obtained trough the Atanassov's intuitionistic fuzzy index values.

The remainder of this paper is organized as follows: In Section 2 we describe the proposed methodology. Section 3 presents the experimental results obtained and, finally, in Section 4 some conclusions and some lines of future work are presented.

2 Proposed methodology

In this section we present the proposed methodology for colour image segmentation that besides its low computational cost is also able to auto determine the number of required thresholds based on the image intuitionistic fuzzy entropy.

Since this methodology deals with each RGB component separately, we will only describe the algorithm for one of the components. The exact same algorithm is to be applied to the other two components.

This methodology successively applies the bi-level thresholding algorithm proposed in [6], for the computation of one threshold, an unspecified number of times to an image, resulting in a set of thresholds for each RGB component. After the number of required thresholds needed for each RGB component are computed, the total number of thresholds is calculated and, on the basis of that number, the three component segmentations are combined in order to create a resulting gray segmented image. The combination of the components is made by adding all three intensities and afterwards normalizing them back to the original scale. Different ways of combining the three RGB thresholds can be explored.

Thus, in the following sub-sections, we will first describe the bi-level thresholding algorithm used and, in the last subsection the multi-level thresholding algorithm.

2.1 Bi-level Thresholding Algorithm

In this section we describe the general framework for image thresholding using A-IFSs and restricted dissimilarity functions proposed in [6]. The goal is to obtain the better threshold in order to separate the object from the background.

We will denote by (x, y) the coordinates of each pixel on the image Q, being $q(x, y)$ the gray level of the pixel (x, y) so that $0 \leq q(x, y) \leq L - 1$ for each $(x, y) \in Q$ where L is the image grayscale.

In this methodology Atanassov Intuitionistic Fuzzy Sets are used in the following way: In order to choose/construct the

ISBN: 978-989-95079-6-8

membership function of each pixel of the image to the associated fuzzy set, three numerical values are assigned to each one of them.

- A value for representing the expert knowledge of the membership of the pixel to the background. A membership function, constructed using dissimilarity functions, is used to obtain this value (see [10]).
- Dissimilarity functions are also used by the expert to construct a membership function to retrieve a value for representing the expert knowledge of the membership of the pixel to the object.
- The expert knowledge/ignorance, in determining the above mentioned membership functions, is represented by a third value obtained trough Atanassov's intuitionistic index.

In this sense, L fuzzy sets Q_{Bt} associated with the background and L fuzzy sets Q_{Ot} associated with the object are constructed. Each one of these fuzzy sets is associated with a gray level t of the grayscale L used. The membership functions of these sets are defined by means of restricted dissimilarity functions and, two of the possible expressions one might use are:

$$\mu_{Q_{Bt}}(q) = F\left(d\left(\frac{q}{L-1}, \frac{m_{\mathrm{B}}(t)}{L-1}\right)\right)$$

$$\mu_{Q_{Ot}}(q) = F\left(d\left(\frac{q}{L-1}, \frac{m_{\mathrm{O}}(t)}{L-1}\right)\right)$$

where

$$m_{\mathrm{B}}(t) = \frac{\sum_{q=0}^{t} qh(q)}{\sum_{q=0}^{t} h(q)} \quad (1)$$

$$m_{\mathrm{O}}(t) = \frac{\sum_{q=t+1}^{L-1} qh(q)}{\sum_{q=t+1}^{L-1} h(q)} \quad (2)$$

being $h(q)$ the number of pixels of the image with the gray level q, $F(x) = 1-0.5x$ and, the restricted dissimilarity function $d(x, y) = |x - y|$ (see [6, 11, 12]).

Due to these constructions, the membership functions constructed are always greater than or equal to 0.5 and, a pixel unequivocally belongs to one of the regions (object or background) if and only if its intensity q is equal to the mean of intensities of the that region ($m_O(t)$ or $m_B(t)$). When the difference between the pixel's intensity q and the mean of intensities of the region is maximal, then the value of its membership function to that region is minimal. Thus, the lesser the distance between a pixel's intensity q and the mean of intensities of the region, the greater the value of its membership to that region.

In this methodology, it is considered that the membership functions ($\mu_{Q_{Bt}}$ and $\mu_{Q_{Ot}}$) indicate the expert's degree of knowledge of the pixel belonging to the regions, meaning that, the Atanassov's intuitionistic fuzzy index (π) represents the ignorance/intuition present in the construction of these fuzzy sets.

Thus, the value represented by Atanassov's intuitionistic index indicates the knowledge/ignorance of the expert when assigning a pixel either to the background or the object. The Atanassov's intuitionistic index value associated with a pixel has the value of zero, when the expert is absolutely sure of that pixel belonging (either to the background or the object). The Atanassov's intuitionistic index value increases with respect to the ignorance/intuition of the expert as to whether the pixel belongs to the background or the object. If the expert doesn't know if a pixel belongs to the background or the object, its membership to both must be represented with the value 0.5 and, it is said that the expert used the greatest ignorance/intuition allowed in the construction of the membership functions, of the set associated with that pixel, to the background and the object resulting in a Atanassov's Intuitionistic Fuzzy Index maximum value. For this reason, A-IFSs (Atanassov's Intuitionistic Fuzzy Set [8, 13]) are used.

However, the ignorance/intuition should have the least possible influence on the choice of the membership degree. In this methodology the ignorance/intuition will have a maximum influence of 25 percent.

Under these conditions, within the set of all the possible expressions(see [6]), the following expression is used to calculate π:

$$\pi(q) = \wedge(1 - \mu_{Q_{Bt}}(q), 1 - \mu_{Q_{Ot}}(q)).$$

This π is now used to construct an A-IFS with each one of the fuzzy sets Q_{Bt} and Q_{Ot}.

$$\tilde{Q}_{Bt} = \{(q, \mu_{\tilde{Q}_{Bt}}(q), \nu_{\tilde{Q}_{Bt}}(q)) | q = 0, 1, \cdots, L-1\},$$

given by

$$\mu_{\tilde{Q}_{Bt}}(q) = \mu_{Q_{Bt}}(q)$$
$$\nu_{\tilde{Q}_{Bt}}(q) = 1 - \mu_{\tilde{Q}_{Bt}}(q) - \pi(q)$$

and

$$\tilde{Q}_{Ot} = \{(q, \mu_{\tilde{Q}_{Ot}}(q), \nu_{\tilde{Q}_{Ot}}(q)) | q = 0, 1, \cdots, L-1\},$$

given by

$$\mu_{\tilde{Q}_{Ot}}(q) = \mu_{Q_{Ot}}(q)$$
$$\nu_{\tilde{Q}_{Ot}}(q) = 1 - \mu_{\tilde{Q}_{Ot}}(q) - \pi(q)$$

At this stage, the Atanassov's intuitionistic fuzzy index is used to calculate the entropy IE of each one of the L Atanassov's intuitionistic fuzzy sets associated with the image.

The following expression is used to calculate the entropy IE, so that $0 \leq IE(\tilde{Q}_{Bt}) \leq 0.25$.

$$IE(\tilde{Q}_{Bt}) = \frac{1}{N \times M} \sum_{q=0}^{L-1} h(q) \cdot \pi(q) \quad (3)$$

where $N \times M$ are the image dimensions in pixels.

This way, the entropy on A-IFSs is interpreted as a measure of the degree of a A-IFS that a set has with respect to the fuzzyness of the said set (see [14]). Under these conditions, the entropy will be null when the set is a FSs and will be maximum when the set is totally intuitionistic.

ISBN: 978-989-95079-6-8

Finally, the gray level associated with the Atanassov's intuitionistic fuzzy set $\tilde{Q}_{Bt}$ of lowest entropy IE is selected as the best threshold.

2.2 *Multi-level Thresholding Algorithm*

In this approach, the goal is to obtain a set of thresholds in order to separate all the image objects. This method is based on a divide and conquer strategy and its main idea is to successively apply the algorithm proposed in [6] and presented in the last section, for the computation of one threshold, an unspecified number of times to an image. In Fig. 1 we illustrate the methodology computational process progress.

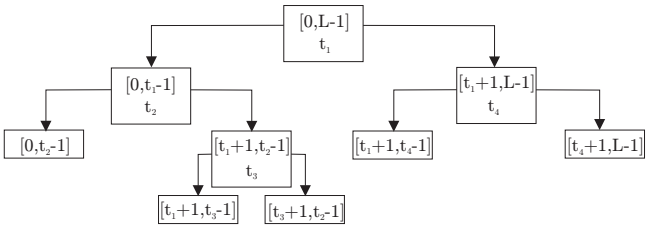


Figure 1: Computational process progress

First, the algorithm is applied to the original image resulting in the determination of the threshold value t_i. This threshold value is then used to create two sub-images: the sub-image with intensity values lower than t_i and the sub-image with intensity values greater than t_i. Then, the algorithm is to be applied to the sub-image which has the greater amplitude between gray-levels entropy values (difference between the intensity value of maximum entropy and the one of minimum entropy).

Hence, The proposed algorithm consecutively divides the resulting sub-images by means of the threshold value t_i obtained through the application of the algorithm for the computation of one threshold to each one of them.

The algorithms stopping criteria is based on the sub-images gray-levels entropy values in such way that the process stops when the difference between the entropy of the computed threshold value t_i and the maximum entropy of the sub-image that generated it, is inferior to the difference between the maximum and minimum entropies of the entire image divided by a constant K.

After experimental evaluation, the value of the constant K was set to 2.

The number of computed thresholds is closely related to the value of this constant K. This value is directly proportional to the number of computed thresholds since, the maximum difference of global entropy allowed for each sub-image to produce a new threshold is inversely proportional to the value of K. Thus, the value of the constant K can be chosen according to the purpose of the application where the algorithm is to be applied.

3 Experimental results

In order to illustrate the methodology we used the "Lena" image in Fig. 2 and Fig. 3.

In Fig. 2 the original "Lena" image and the segmentation result are presented. In Fig. 3, in the first column, we present



Figure 2: Original and segmented "Lena" image.

the image decomposed in its RGB components and, in the second column, each RGB component segmentation result is presented. For all the "Lena" segmented images the number of computed thresholds is also presented.

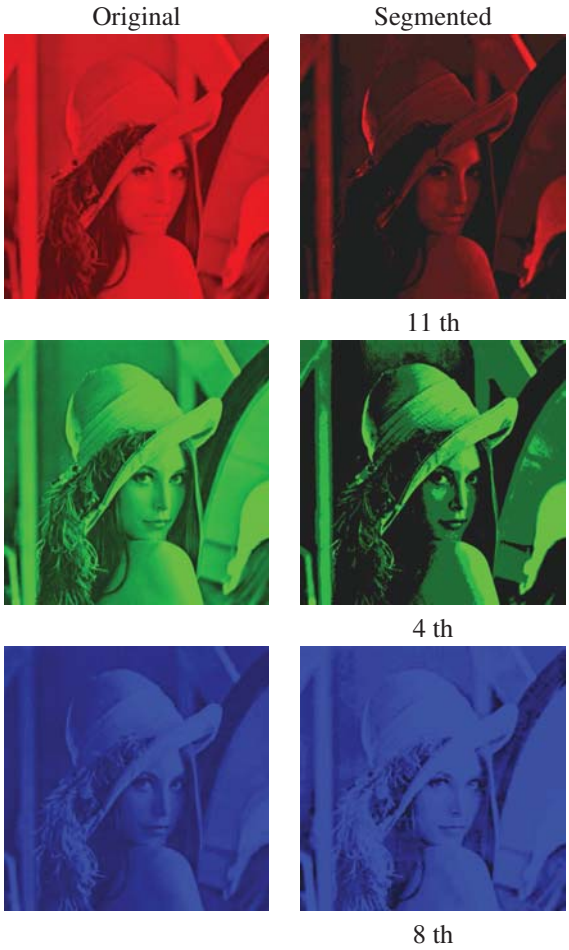


Figure 3: Original and segmented RGB components of "Lena" image.

Regarding the number of computed thresholds in each RGB component we can conclude that, as one can see looking at he original image, the red component is the one that has more information regarding the segmentation of the image and, therefore the red component was the one with more computed thresholds. We think that this is an indicator of the goodness of the methodology since it was able to accurately auto determine the required number of thresholds in each component

ISBN: 978-989-95079-6-8

separately.

Looking at Fig. 3, we can also see that the RGB components are highly related. We consider this to be the major drawback of segmenting images in the RGB colour space.

To test the performance of the proposed approach, more than a hundred images where segmented. However, in order to illustrate the obtained results, 5 images from the original set of images (taken form the Berkeley Segmentation Dataset: `http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/`) were selected and used as test images (see Fig. 4).



Figure 4: Original and segmented test images.

In Fig 4 we present the results from the application of the proposed methodology to complex images where, the value of the constant K was also set at 2 and, where we can see that the main features of the images are preserved.

4 Conclusions and Future Work

Since, an image usually contains more than one region, in this paper we presented a new extension of an existing method, that uses A-IFSs for multi-level colour image segmentation. The algorithm first segments each RGB component separately and finally combines them in order to obtain the final segmented image. The use of A-IFSs allowed us to use entropy in the same sense as fuzzy entropy is used in the fuzzy algorithms and, consequently, allowed us to endow the methodology with the capability of dealing with uncertainty.

The proposed methodology successively applies a bi-level thresholding algorithm an undetermined number of times to an image RGB component subdividing it into several disjoined regions. Moreover, the algorithm is able to auto determine the number of thresholds required to segment each RGB image component.

Considering the experimental results, we can say that, in general terms, the proposed algorithm provides suitable results and therefore can be considered when colour image segmentation applications are needed.

The proposed methodology is parameterizable through the setting of a constant K. The value of this constant is directly proportional to the optimal number of thresholds computed. Thus, the value of the constant K should always depend of the application needs. Future work is intended to study the incorporation of heuristics in the algorithm through the setting of the constant K value. Different combinations of the resulting thresholds of each one of the RGB components can be an object of future work in order to incorporate the perceptual differences among the three components.

Due to the high correlation between the three RGB components, further work is also intended to apply an extension of this algorithm to images in the HSV colour space. Being the HSV colour space a more intuitive color space, it's our belief that a more coherent segmentation will be achieved.

Acknowledgment

This work has been partially supported by *Acções integradas Luso-Espanholas* 2007, *Acciones Integradas Hispano-Portuguesas*, ref. HP2006-0128 (MEC), and POCI 2010 (*Programa Operacional Ciência e Investigação* 2010)

References

[1] J. Bezdek, J. Keller, R. Krisnapuram, and N. Pal. *Fuzzy Models and algorithms for pattern recognition and image processing, The Handbooks of Fuzzy Sets Series*. Kluwer Academic Publishers, Boston/London/Dordrecht, 1999.

[2] Z. Chi, H. Yan, and T. Pham. Optimal image thresholding. *Fuzzy algorithms: with application to image processing and pattern recognition*, pages 45–84, 1998.

[3] L. Huang and M. Wang. Image thresholding by minimizing the measure of fuzziness. *Pattern recognition*, 28(1):41–51, 1995.

[4] S. Zenzo, L. Cinque, and S. Levialdi. Image thresholding using fuzzy entropies. *IEEE Transactions on Systems, Man and Cybernetics*, 28(1):15–23, 1998.

[5] M. Forero. *Fuzzy Filters for Image Processing*, chapter Fuzzy thresholding and histogram analysis, pages 129–152. Springer, 2003.

[6] H. Bustince, M. Pagola, P. Melo-Pinto, E. Barrenechea, and P. Couto. *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models, in Studies in Fuzziness and Soft Computing*, chapter Image threshold computation by modelizing knowledge/unknowledge by means of A-IFSs, pages 627–644. Springer-Verlag, Berlin Heidelberg New York, 2007.

[7] K. Atanassov. Intuitionistic fuzzy sets. *VII ITKR's Session, Deposed in Central Sci. Techn. Library of Bulg. Acd. of Sci., Sofia*, pages 1684–1697, 1983.

ISBN: 978-989-95079-6-8

[8] K. Atanassov. Intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, 20:87–96, 1986.

[9] P. Burillo and H. Bustince. Construction theorems for intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, 84:271–281, 1996.

[10] E. Barrenechea. *Image Thresholding with Interval-valued Fuzzy Sets. Edge Detection. Contrast*. PhD thesis, Universidad Pública de Navarra, 2005.

[11] H. Bustince, E. Barrenechea, and M. Pagola. Relationship between restricted dissimilarity functions, restricted equivalence functions and e_n-functions: Image threshold invariant. *Pattern Recognition Letters*, In Press 2007.

[12] P. Couto. *Image segmentation using Atanassov intuitionistic fuzzy sets*. PhD thesis, Trás-os-Montes e Alto Douro University, Vila Real, Portugal, 2006.

[13] K. Atanassov. *Intuitionistic fuzzy sets. Theory and applications*. Springer-Verlag, Berlin Heidelberg New York, 1 edition, January 1999.

[14] P. Burillo and H. Bustince. Entropy on intuitionistic fuzzy sets and on intervalvalued fuzzy sets. *Fuzzy Sets and Systems*, 78:81–103, 1996.

ISBN: 978-989-95079-6-8

SaM: A Split and Merge Algorithm for Fuzzy Frequent Item Set Mining

Christian Borgelt[1] and Xiaomeng Wang[2]

1. European Center for Soft Computing, c/ Gonzalo Gutiérrez Quirós s/n, 33600 Mieres, Spain
2. Otto-von-Guericke-University of Magdeburg, Universitätsplatz 2, 39106 Magdeburg, Germany
Email: christian.borgelt@softcomputing.es, xwang@iws.cs.uni-magdeburg.de

***Abstract*—*This paper presents SaM, a split and merge algorithm for frequent item set mining. Its distinguishing qualities are an exceptionally simple algorithm and data structure, which not only render it easy to implement, but also convenient to execute on external storage. Furthermore, it can easily be extended to allow for "fuzzy" frequent item set mining in the sense that missing items can be inserted into transactions with a user-specified penalty. In order to demonstrate its performance, we report experiments comparing it with the "fuzzy" frequent item set mining version of RElim (an algorithm we suggested in an earlier paper [15] and improved in the meantime).*

Keywords— data mining, frequent item set mining, fuzzy frequent item set, fault tolerant data mining

1 Introduction

Although frequent item set mining and association rule induction has been a focus of research in data mining for a long time now, leading to well-known algorithms like Apriori [1], Eclat [11] and FP-growth [7], there is still room for improvement. Recent research lines include filtering the found frequent item sets and association rules [16, 17], identifying temporal changes in discovered patterns [3, 4], and mining fault-tolerant or "fuzzy" frequent item sets [6, 10, 15].

In this paper we follow the last of these lines by presenting SaM, a split and merge algorithm for frequent item set mining, which can easily be extended to allow for "fuzzy" mining in the sense that missing items can be inserted into transactions with a user-specified penalty. Other distinguishing qualities of our method are its exceptionally simple processing scheme and data structure, which not only render it very easy to implement, but also convenient to execute on external storage.

2 Frequent Item Set Mining

Frequent item set mining is the following task: we are given a set B of *items*, called the *item base*, and a database T of *transactions*. An item may, for example, represent a product, and the item base may then represent the set of all products offered by a supermarket. The term *item set* refers to any subset of the item base B. Each transaction is an item set and may represent, in the supermarket setting, a set of products that has been bought by a customer. Since several customers may have bought the exact same set of products, the total of all transactions must be represented as a vector or a multiset (or, alternatively, each transaction must be enhanced by a *transaction identifier (tid)*). Note that the item base B is usually not given explicitly, but only implicitly as the union of all transactions. The *support* $s_T(I)$ of an item set $I \subseteq B$ is the number of transactions in the database T it is contained in. Given a user-specified *minimum support* $s_{\min} \in \mathbb{N}$, an item set I is called *frequent* (in T) iff $s_T(I) \geq s_{\min}$. The goal of frequent item set mining is to find all item sets $I \subseteq B$ that are frequent in the database T and thus, in the supermarket setting, to identify all sets of products that are frequently bought together.

A standard approach to find all frequent item sets w.r.t. a given database T and a support threshold $s_{\min}$, is a *depth-first search* in the subset lattice of the item base B. This approach can be seen as a simple *divide-and-conquer* scheme. For a chosen item i, the problem to find all frequent item sets is split into two subproblems: (1) find all frequent item sets containing i and (2) find all frequent item sets *not* containing i. Each subproblem is then further divided based on another item j: find all frequent item sets containing (1.1) both i and j, (1.2) i, but not j, (2.1) j, but not i, (2.2) neither i nor j etc.

All subproblems occurring in this recursion can be defined by a *conditional transaction database* and a *prefix*. The prefix is a set of items that has to be added to all frequent item sets that are discovered in the conditional database. Formally, all subproblems are tuples $S = (C, P)$, where C is a conditional database and $P \subseteq B$ is a prefix. The initial problem, with which the recursion is started, is $S = (T, \emptyset)$, where T is the given transaction database and the prefix is empty.

A subproblem $S_0 = (C_0, P_0)$ is processed as follows: choose an item $i \in B_0$, where B_0 is the set of items occurring in C_0. This choice is arbitrary, but usually follows some predefined order of the items. If $s_{C_0}(i) \geq s_{\min}$, then report the item set $P_0 \cup \{i\}$ as frequent with the support $s_{C_0}(i)$, and form the subproblem $S_1 = (C_1, P_1)$ with $P_1 = P_0 \cup \{i\}$. The conditional database C_1 comprises all transactions in C_0 that contain i, but with this item removed. This implies that all transactions are removed that do not contain any no other item than i. If C_1 is not empty, process S_1 recursively. In any case (that is, regardless of whether $s_{C_0}(i) \geq s_{\min}$ or not), form the subproblem $S_2 = (C_2, P_2)$ with $P_2 = P_0$. The conditional database C_2 comprises all transactions in C_0 (including those that do not contain i), but again with the item i removed. If C_2 is not empty, process S_2 recursively.

This recursive scheme is adopted by Eclat, FP-growth, RElim and several other frequent item set mining algorithms. They differ in how conditional transaction databases are represented: in a *horizontal representation*, a database is stored as a list (or array) of transactions, each of which lists the items contained in it. In a *vertical representation*, the items are first referred to with a list (or array) and for each item the transactions containing it are listed. However, this distinction is not pure, since there are combinations of the two forms of representing a database. Our SaM algorithm is, as far as we know, the first algorithm that is based on the general scheme outlined above and uses a purely horizontal representation.

ISBN: 978-989-95079-6-8

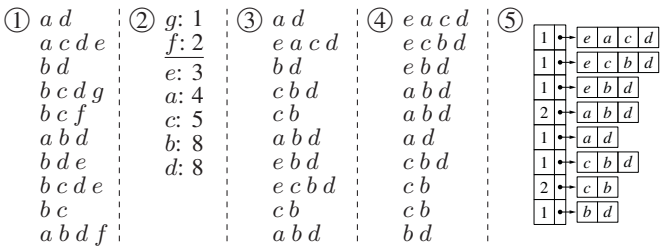


Figure 1: An example database: original form (1), item frequencies (2), transactions with sorted items (3), lexicographically sorted transactions (4), and the used data structure (5).

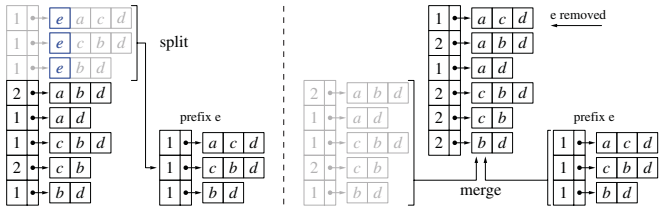


Figure 2: The basic operations: split (left) and merge (right).

The basic processing scheme can be improved with so-called *perfect extension pruning*: an item $i \notin I$ is called a *perfect extension* of an item set I, iff I and $I \cup \{i\}$ have the same support. Perfect extensions have the following properties: (1) if an item i is a perfect extension of an item set I, then it is also a perfect extension of any item set $J \supseteq I$ as long as $i \notin J$ and (2) if K is the set of all perfect extensions of an item set I, then all sets $I \cup J$ with $J \in 2^K$ (where 2^K denotes the power set of K) have the same support as I. These properties can be exploited by collecting in the recursion not only prefix items, but also, in a third element of a subproblem description, perfect extension items. They are also removed from the conditional databases and are only used to generate all supersets of the prefix that have the same support.

3 A Simple Split and Merge Algorithm

In this section we describe the basic form of our SaM (split and merge) algorithm. Preprocessing is very similar to many other frequent item set mining algorithms. The steps are illustrated in Figure 1 for a simple example transaction database: step 1 shows the transaction database in its original form. In step 2 the item frequencies are determined in order to discard infrequent items. With a minimum support of 3, items f and g are infrequent and thus eliminated. In step 3 the (frequent) items in each transaction are sorted according to their frequency, because processing the items in the order of increasing frequency usually leads to the shortest execution times. In step 4 the transactions are sorted lexicographically into descending order, with an item with higher frequency preceding an item with lower frequency. In step 5 the basic data structure is built by combining equal transactions and setting up an array, in which each element consists of two fields: an occurrence counter and a pointer to the sorted transaction.

The basic operations of the recursive processing, which follows the general divide-and-conquer scheme reviewed in Section 2, are illustrated in Figure 2: in the *split step* (left) the given array is split w.r.t. the leading item of the first transaction (item e in our example): all elements referring to transactions starting with this item are transferred to a new array. In this process the pointer (in)to the transaction is advanced by one item, so that the common leading item is "removed" from all transactions. Obviously, this new array represents the conditional database of the first subproblem (see Section 2), which is then processed recursively to find all frequent item sets containing the split item (provided this item is frequent).

The conditional database for frequent item sets *not* containing this item (second subproblem, see Section 2) is obtained with a simple *merge step* (right part of Figure 2). The new array and the rest of the original array are combined with a procedure that is almost identical to one phase of the well-known *mergesort* algorithm. Since both arrays are lexicographically sorted, one merging traversal suffices to create a lexicographically sorted merged array. The only difference to a *mergesort* phase is that equal transactions (or transaction suffixes) are combined: There is always only one instance of each transaction (suffix), while its number of occurrences is kept in a counter. In our example this results in the merged array having two elements less than the input arrays together: the transaction (suffixes) cbd and bd, which occur in both arrays, are combined and their occurrence counters are increased to 2.

Pseudo-code of SaM is shown in Figure 3: a single page of code suffices to describe the whole recursion in detail. The actual C code is even shorter, despite the fact that it contains additional functionality (like perfect extension pruning, Section 2), because certain operations can be written very concisely in C (especially when using pointer arithmetic).

4 Fuzzy Frequent Item Set Mining

There are many applications of frequent item set mining, in which the transactions do not contain all items that are actually present. However, standard algorithms are based on exact matching and therefore are not equipped to meet the needs arising in such applications. An example is the analysis of alarm sequences in telecommunication networks, where a core task is to find collections of alarms occurring frequently together, so-called *episodes*. One approach to accomplish this task is to slide a time window over the alarm sequence. Each window position then captures a specific slice of the alarm sequence [12]. The underlying idea is that in this way the problem of finding frequent episodes is reduced to that of finding frequent item sets in a database of transactions: each alarm can be seen as an item and the alarms in a time window as a transaction. The support of an episode is the number of window positions, so that the episode occurred in the window.

Unfortunately, alarms often get delayed, lost, or repeated due to noise, transmission errors, failing links etc. If alarms do not get through or are delayed, they are missing from the transaction (time window) its associated items (alarms) occur in. If we required exact containment of an item set in this case, the support of some item sets, which would be frequent if the items did not get lost, may be lower than the user-specified minimum. This leads to a possible loss of potentially interesting frequent item sets and to distorted support values.

To cope with such missing information, we rely on the notion of a "fuzzy" or approximate frequent item set. In contrast to research on fuzzy association rules (see, for example, [13]),

ISBN: 978-989-95079-6-8

```
function SaM (a: array of transactions,    (* conditional database *)
              p: set of items,             (* prefix of the cond. database a *)
              s_min: int) : int            (* min. support of an item set *)
var i: item;                               (* buffer for split item *)
    s: int;                                (* support of current split item *)
    n: int;                                (* number of frequent item sets *)
    b, c, d: array of transactions;        (* cond. and merged database *)
begin                                      (* – split and merge recursion – *)
  n := 0;                                  (* init. number of freq. item sets *)
  while a is not empty do                  (* while database is not empty *)
    b := empty; s := 0;                    (* init. split result, item support *)
    i := a[0].items[0];                    (* get the leading item *)
    while a is not empty                   (* of the first transaction and *)
    and   a[0].items[0] = i do             (* split database w.r.t. this item *)
      s := s + a[0].wgt;                   (* sum occurrences (support) *)
      remove i from a[0].items;            (* remove the split item *)
      if    a[0].items is not empty        (* if trans. is not empty *)
      then remove a[0] from a and append it to b;
      else remove a[0] from a; end;        (* move it to cond. db., *)
    end;                                   (* otherwise simply remove it *)
    c := b; d := empty;                    (* initialize the output array *)
    while a and b are both not empty do    (* merge step *)
      if    a[0].items > b[0].items        (* copy trans. from a *)
      then remove a[0] from a and append it to d;
      else if a[0].items < b[0].items      (* copy trans. from b *)
      then remove b[0] from b and append it to d;
      else b[0].wgt := b[0].wgt +a[0].wgt;
           remove b[0] from b and append it to d;
           remove a[0] from a;             (* combine weights and *)
      end;                                 (* move and remove trans.: *)
    end;                                   (* only one instance per trans. *)
    while a is not empty do                (* copy the rest of a *)
      remove a[0] from a and append it to d; end;
    while b is not empty do                (* copy the rest of b *)
      remove b[0] from b and append it to d; end;
    a := d;                                (* loop for second recursion *)
    if s ≥ s_min then                      (* if the split item is frequent: *)
      p := p ∪ {i};                        (* extend the prefix item set and *)
      report p with support s;             (* report the frequent item set *)
      n := n + 1 + SaM(c, p, s_min);
      p := p − {i};                        (* process cond. db. recursively, *)
    end;                                   (* sum the frequent item sets, *)
  end;                                     (* then restore the orig. prefix *)
  return n;                                (* return num. of freq. item sets *)
end;    (* function SaM() *)
```

Figure 3: Pseudo-code of the SaM algorithm.

where a fuzzy approach is used to handle quantitative items, we use the term "fuzzy" to refer to an item set that may not be present exactly in all supporting transactions, but only approximately. Related work in this direction suggested Apriori-like algorithms and mining with approximate matching was performed by counting the number of different items in the two item sets to be compared [6, 10]. However, here we adopt a more general scheme, based on an approximate matching approach exhibiting much greater flexibility. Our approach has two core ingredients: *edit costs* and *transaction weights* [15].

Edit costs: A convenient way of defining the distance between item sets is to consider the costs of a cheapest sequence of edit operations needed to transform one item set into the other [14]. Here we consider only insertions, since they are most easily implemented with our algorithm[1]. With the help of an *insertion cost* or *penalty* a flexible and general framework for approximately matching two item sets can be established. How one interprets such costs or penalties depends, of course, on the application. Note also that different items may be associated with different costs. For example, in telecommunication networks different alarms can have a different probability of getting lost: usually alarms raised in lower levels of the module hierarchy get lost more easily than alarms originating in higher levels. In such cases it is convenient to be able to associate the former with lower insertion costs than the latter. Insertions of a certain item may also be completely inhibited by assigning a very high insertion cost.

Transaction weights: Each transaction t is associated with a weight $w(t)$, the initial value of which is 1. If an item i is inserted into a transaction t, the transaction weight is "penalized" with a cost $c(i)$ associated with the item. Formally, this can be described as applying a combination function: the new weight of the transaction t after inserting an item $i \notin t$ is $w_{\{i\}}(t) = f(w(t), c(i))$, where f is a function that combines the weight $w(t)$ before editing and the insertion cost $c(i)$. The combination function f depends, of course, on the application and may be chosen from a wide range of possible functions. For example, any t-norm may be used. We choose multiplication here, that is, $w_{\{i\}}(t) = w(t) \cdot c(i)$, mainly for reasons of simplicity. Note, however, that with this choice lower values of $c(i)$ mean higher costs as they penalize the weight more, but that it has the advantage that it is easy to extend to inserting several items: $w_{\{i_1,\ldots,i_m\}}(t) = w(t) \cdot \prod_{k=1}^{m} c(i_k)$. It should be clear that it is $w_\emptyset(t) = 1$ due to the initial weighting $w(t) = 1$.

How many insertions into a transaction are allowed may be limited by a user-specified lower bound $w_{\min}$ for the transaction weight. If the weight of a transaction falls below this threshold, it is not considered in further mining steps. Of course, this weight may also be set to zero (unlimited insertions). As a consequence, the *fuzzy support* of an item set I w.r.t. a transaction database T can be defined as $s_T^{(\text{fuzzy})}(I) = \sum_{t \in T} \tau(w_{I-t}(t) \geq w_{\min}) \cdot w_{I-t}(t)$, where $\tau(\phi)$ is a kind of "truth function", which is 1 if ϕ is true and 0 otherwise.

Note that SaM is particularly well suited to handle item insertions, because its purely horizontal transaction representation makes it easy to incorporate transaction weights. With other algorithms, more effort is usually needed in order to extend them to approximate frequent item set mining.

For an implementation, it is beneficial to distinguish between unlimited item insertions ($w_{\min} = 0$) and limited item insertions ($w_{\min} > 0$). If $w_{\min} = 0$, it is possible to combine equal transactions (or transaction suffixes) without restriction: two equal transactions (or suffixes) t_1 and t_2 with weights w_1 and w_2, respectively, can be combined into one transaction (suffix) t with weight $w_1 + w_2$ even if $w_1 \neq w_2$. If another item i needs to be inserted into t_1 and t_2 in order to make them contain a given item set I, the distributive law (that is, $w_1 \cdot c(i) + w_2 \cdot c(i) = (w_1 + w_2) \cdot c(i)$) ensures that we still compute the correct support for the item set I. If, however, we have $w_{\min} > 0$ and, say, $w_1 > w_2$, then using $(w_1 + w_2) \cdot c(i)$ as the contribution of the combined transaction t to the support of the item set I may be wrong, because it may be that $w_1 \cdot c(i) \geq w_{\min}$, but $w_2 \cdot c(i) < w_{\min}$. Then the support contributed by the two transactions t_1 and t_2 would rather be $w_1 \cdot c(i)$. Effectively, transaction t_2 does not contribute, since its weight has fallen below the transaction weight threshold. Hence, with limited insertions, we may combine equal transactions (or suffixes) only if they have the same weight.

[1] Note that deletions are implicit in the mining process (as we search for *subsets* of the transactions). Only replacements are an additional case we do not consider here.

ISBN: 978-989-95079-6-8

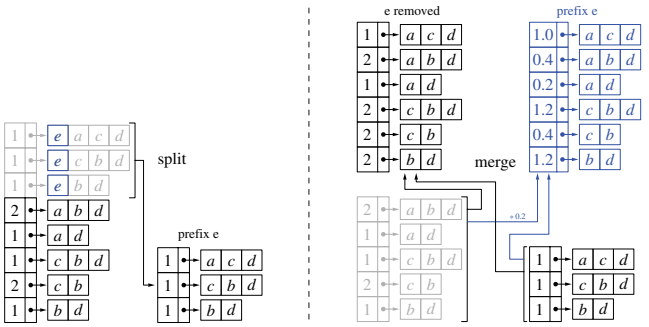


Figure 4: Unlimited item insertions, first recursion level.

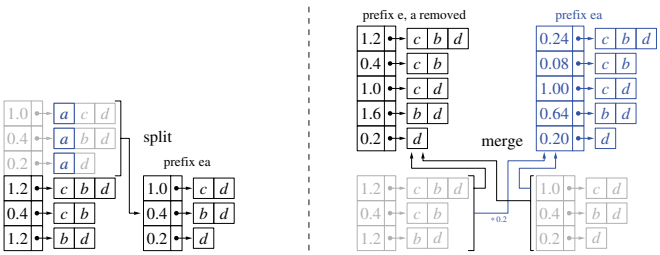


Figure 5: Unlimited item insertions, second recursion level.

5 Unlimited Item Insertions

If unlimited item insertions are possible ($w_{\min} = 0$), only a minor change of the data structure is needed: the integer occurrence counter for the transactions (or suffixes) has to be replaced by a real-valued transaction weight. In the processing, the split step stays the same (see Figure 4 on the left), but now it only yields an intermediate database with all transactions (or suffixes) that actually contain the split item under consideration (item e in the example). In order to form the full conditional database, we have to add those transactions that do not contain the split item, but can be made to contain it by inserting it. This is achieved in the merge step, in which two parallel merge operations are carried out now (see Figure 4 on the right). The first part (shown in black) is the merge that yields the conditional database for frequent item sets *not* containing the split item. The second part (shown in blue/grey) adds those transactions that do not contain the split item, weighted down with the insertion penalty, to the intermediate database created in the split step. Of course, this second part of the merge operation is only carried out, if $c(i) > 0$, where i is the split item, because otherwise no support would be contributed by the transactions not containing the item i and hence it would not be necessary to add them. In such a case the result of the split step would already yield the conditional database for frequent item sets containing the split item.

Note that in both merge operations equal transactions (or suffixes) can be combined regardless of their weights. As a consequence we have in Figure 4 entries like for the transaction (suffix) cbd, with a weight of 1.2, which stands for one occurrence with weight 1 and one occurrence with weight 0.2 (due to the penalty factor 0.2, needed due to the insertion of item e). As an additional illustration, Figure 5 shows the split and merge operations for the second recursion level.

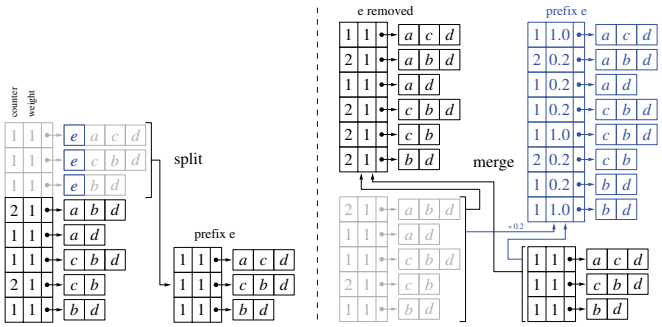


Figure 6: Limited item insertions, first recursion level.

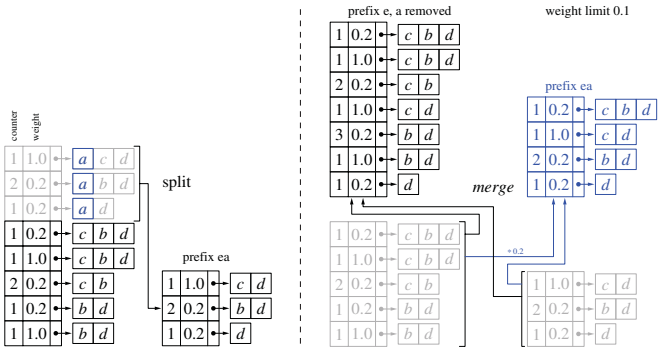


Figure 7: Limited item insertions, second recursion level.

6 Limited Item Insertions

If item insertions are limited by a transaction weight threshold ($w_{\min} > 0$), the transaction weight has to be represented explicitly and kept separate from the number of occurrences. Therefore the data structure must comprise, per transaction (suffix), (1) a pointer to the item array, (2) an integer occurrence counter, and (3) a real-valued transaction weight. The last field will be subject to a thresholding operation by $w_{\min}$, which eliminates all transactions with a weight less than $w_{\min}$. Hence there may now be array elements that refer to the same transaction (suffix), but differ in the transaction weight.

The processing scheme is illustrated in Figure 6. The split step is still essentially the same. However, the merge step differs due to the fact that equal transactions (or suffixes) can no longer be combined if their weight differs. As a consequence, there are now, in the result of the second merge operation (shown in blue) two array elements for cbd and two for bd, which carry different weights. This is necessary, because they may reach, due to item insertions, the transaction weight threshold at different times and thus cannot be combined.

That transactions are discarded due to the weight threshold rarely happens on the first level of the recursion. (This can occur only if the insertion penalty factor of the split item is smaller than the transaction weight threshold, which is equivalent to inhibiting insertions of this item altogether). Therefore, in order to illustrate how transactions are discarded, Figure 7 shows the second recursion level, where the conditional database with prefix e is processed. Here the second merge operation actually discards transactions if we set a transaction weight limit of 0.1: all transactions, which need two items (namely both e and a) to be inserted, are not copied.

ISBN: 978-989-95079-6-8

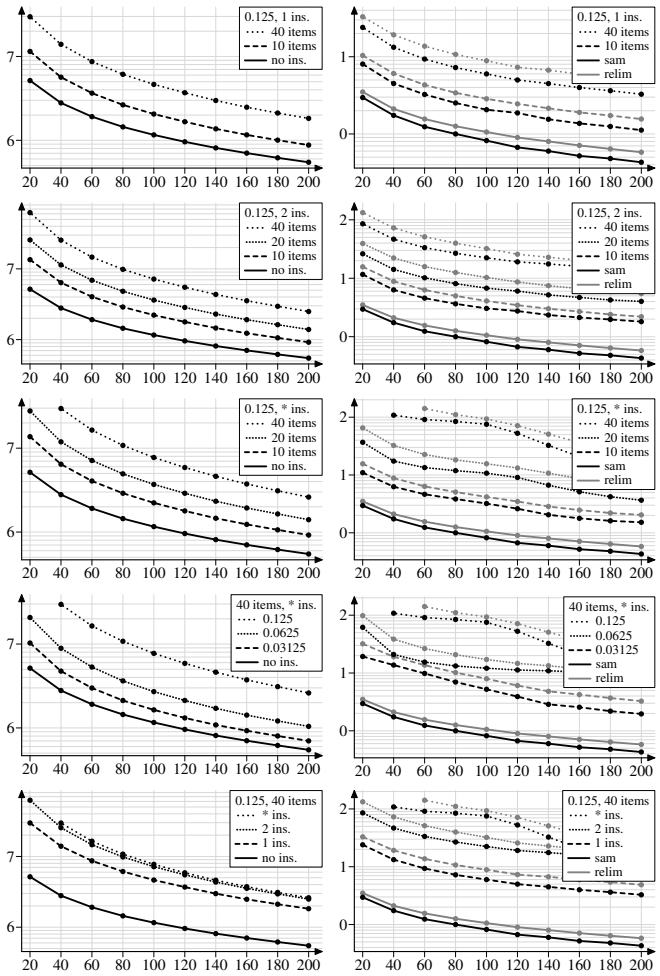


Figure 8: Experimental results on the Census (Adult) data; left: frequent item sets, right: execution times.

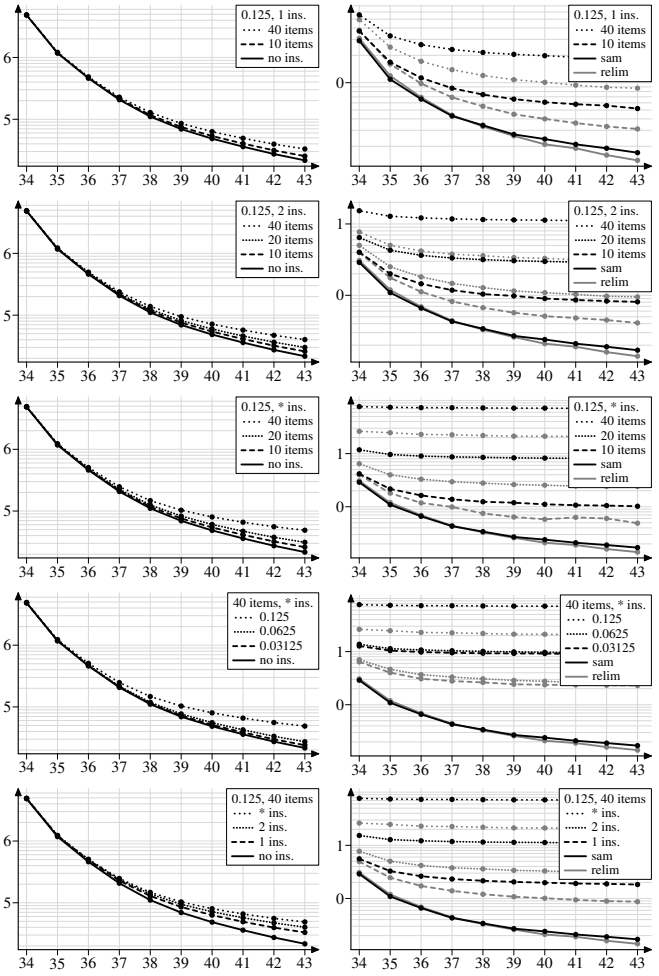


Figure 9: Experimental results on the BMS-Webview-1 data; left: frequent item sets, right: execution times.

7 Experiments

We ran experiments on several data sets, of which we chose two for this paper: Census (aka Adult, a data set derived from an extract of the US census bureau data of 1994, which was preprocessed by discretizing numeric attributes [2]) and BMS-Webview-1 (a web click stream from a leg-care company that no longer exists, which has been used in the KDD cup 2000 [8]). We chose these data sets, because Census is rather dense (a rather large fraction of all items occur in each transaction), while BMS-Webview-1 is rather sparse, and SaM and RElim [15] (the two algorithms of which we have implementations that can find approximate frequent item sets) exhibit a significantly different behavior on dense and sparse data sets.

The results are shown in Figure 8 for the census data set and in Figure 9 for the BMS-Webview-1 data set. In both figures the diagrams on the left show the decimal logarithm of the number of found frequent item sets, while the diagrams on the right show the decimal logarithm of the execution times (in seconds) for our implementations of SaM and RElim.[2] We tested insertion penalty factors of $\frac{1}{8} = 0.125$, $\frac{1}{16} = 0.0625$, and $\frac{1}{32} = 0.03125$, non-vanishing insertion penalty factors for 10, 20, and 40 items, and transaction weight thresholds that allowed for 1, 2 or an unlimited number of insertions.

As can be seen from the diagrams on the left of each figure, the two data sets react very differently to the possibility of inserting items into transactions. While the number of found frequent item sets rises steeply with all parameters for Census, it rises only very moderately for BMS-Webview-1, with the factor even leveling off for lower support values. As it seems, this effect is due, to a large degree, to the sparseness of BMS-Webview-1 (this still needs closer examination, though).

SaM fares considerably better on the dense data set (Census), beating RElim by basically the same margin (factor) in all parameter settings, while SaM is clearly outperformed by RElim on the sparse data set (BMS-Webview-1), even though the two algorithms are actually on par without item insertion (solid lines). On both data sets, the number of insertions that are allowed has the strongest influence: with two insertions execution times are about an order of magnitude larger than with only one insertion. However, the possibility to combine equal transactions with different weights still seems to keep the execution times for unlimited insertions within limits.

The number of items with a non-vanishing penalty factor and the value of the penalty factor itself seem to have a similar influence: doubling the number of items leads to roughly the same effect as keeping the number the same and doubling the

[2]Execution times were measured on an Intel Core 2 Quad Q9300 machine with 3 GB of main memory running openSuSE Linux 11.0 (32 bit) and gcc version 4.3.1.

ISBN: 978-989-95079-6-8

penalty factor. This is plausible, since there should not be much difference in having the possibility to insert twice the number of items or preserving twice the transaction weight per item insertion. Note, however, that doubling the penalty factor from from $\frac{1}{32}$ to $\frac{1}{16}$ has only a comparatively small effect on the BMS-Webview-1 data compared to doubling from $\frac{1}{16}$ to $\frac{1}{8}$. On the census data set the effects are a bit more in line.

Overall it should be noted that the execution times, although considerably increased over those obtained without item insertions, still remain within acceptable limits. Even with 40 items having an insertion penalty factor of $\frac{1}{8}$ and unlimited insertions, few execution times exceed 180 seconds ($\log_{10}(180) \approx 2.25$). In addition, we can observe the interesting effect on the BMS-Webview-1 data that at the highest parameter settings the execution times become almost independent of the minimum support threshold.

8 Conclusions

In this paper we presented a very simple split and merge algorithm for frequent item set mining, which, due to the fact that it uses a purely horizontal transaction representation, lends itself well to an extension to "fuzzy" or approximate frequent item set mining. In addition, it is a highly recommendable method if the data to mine cannot be loaded into main memory and thus the data has to be processed on external storage or in a (relational) database system. As our experimental results show, the SaM algorithm performs the task of "fuzzy" frequent item set mining excellently on the dense Census data, but shows certain weaknesses on the sparse BMS-Webview-1 data. However, our experiments provided some evidence (to be substantiated on more data sets) that "fuzzy" frequent item set mining is much more useful for dense data sets as more additional frequent item sets can be found. Hence SaM performs better in the (likely) more relevant case. Most importantly, however, one should note that with both SaM and RElim the execution times remain bearable.

Software: SaM and RElim sources in C can be found at:

http://www.borgelt.net/sam.html
http://www.borgelt.net/relim.html

References

[1] R. Agrawal and R. Srikant. Fast Algorithms for Mining Association Rules. *Proc. 20th Int. Conf. on very Large Databases (VLDB 1994, Santiago de Chile)*, 487–499. Morgan Kaufmann, San Mateo, CA, USA 1994

[2] C.L. Blake and C.J. Merz. *UCI Repository of Machine Learning Databases*. Dept. of Information and Computer Science, University of California at Irvine, CA, USA 1998

[3] M. Böttcher, M. Spott and D. Nauck. Detecting Temporally Redundant Association Rules. *Proc. 4th Int. Conf. on Machine Learning and Applications (ICMLA 2005, Los Angeles, CA)*, 397–403. IEEE Press, Piscataway, NJ, USA 2005

[4] M. Böttcher, M. Spott and D. Nauck. A Framework for Discovering and Analyzing Changing Customer Segments. *Advances in Data Mining — Theoretical Aspects and Applications (LNCS 4597)*, 255–268. Springer, Berlin, Germany 2007

[5] C. Borgelt. Keeping Things Simple: Finding Frequent Item Sets by Recursive Elimination. *Proc. Workshop Open Software for Data Mining (OSDM'05 at KDD'05, Chicago, IL)*, 66–70. ACM Press, New York, NY, USA 2005

[6] Y. Cheng, U. Fayyad, and P.S. Bradley. Efficient Discovery of Error-Tolerant Frequent Itemsets in High Dimensions. *Proc. 7th Int. Conf. on Knowledge Discovery and Data Mining (KDD'01, San Francisco, CA)*, 194–203. ACM Press, New York, NY, USA 2001

[7] J. Han, H. Pei, and Y. Yin. Mining Frequent Patterns without Candidate Generation. *Proc. Conf. on the Management of Data (SIGMOD'00, Dallas, TX)*, 1–12. ACM Press, New York, NY, USA 2000

[8] R. Kohavi, C.E. Bradley, B. Frasca, L. Mason, and Z. Zheng. KDD-Cup 2000 Organizers' Report: Peeling the Onion. SIGKDD Exploration 2(2):86–93. ACM Press, New York, NY, USA 2000

[9] J. Pei, J. Han, B. Mortazavi-Asl, and H. Zhu. Mining Access Patterns Efficiently from Web Logs. *Proc. Pacific-Asia Conf. on Knowledge Discovery and Data Mining (PAKDD'00, Kyoto, Japan)*, 396–407. 2000

[10] J. Pei, A.K.H. Tung, and J. Han. Fault-Tolerant Frequent Pattern Mining: Problems and Challenges. *Proc. ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMK'01, Santa Babara, CA)*. ACM Press, New York, NY, USA 2001

[11] M. Zaki, S. Parthasarathy, M. Ogihara, and W. Li. New Algorithms for Fast Discovery of Association Rules. *Proc. 3rd Int. Conf. on Knowledge Discovery and Data Mining (KDD'97, Newport Beach, CA)*, 283–296. AAAI Press, Menlo Park, CA, USA 1997

[12] H. Mannila, H. Toivonen, and A.I. Verkamo. *Discovery of Frequent Episodes in Event Sequences*. Report C-1997-15, University of Helsinki, Finland 1997

[13] C. Kuok, A. Fu, and M. Wong. Mining Fuzzy Association Rules in Databases. *SIGMOD Record* 27(1):41–46. 1998

[14] P. Moen. *Attribute, Event Sequence, and Event Type Similarity Notions for Data Mining*. Ph.D. Thesis/Report A-2000-1, Department of Computer Science, University of Helsinki, Finland 2000

[15] X. Wang, C. Borgelt, and R. Kruse. Mining Fuzzy Frequent Item Sets. *Proc. 11th Int. Fuzzy Systems Association World Congress (IFSA'05, Beijing, China)*, 528–533. Tsinghua University Press and Springer-Verlag, Beijing, China, and Heidelberg, Germany 2005

[16] G.I. Webb and S. Zhang. k-Optimal-Rule-Discovery. *Data Mining and Knowledge Discovery* 10(1):39–79. Springer, Amsterdam, Netherlands 2005

[17] G.I. Webb. Discovering Significant Patterns. *Machine Learning* 68(1):1–33. Springer, Amsterdam, Netherlands 2007

ISBN: 978-989-95079-6-8

Online Recognition of Fuzzy Time Series Patterns

Gernot Herbst, Steffen F. Bocklisch

Department of Electrical Engineering, Chemnitz University of Technology
D-09107 Chemnitz, Germany
Email: {gernot.herbst, steffen.bocklisch}@etit.tu-chemnitz.de

Abstract— *This article deals with the recognition of recurring multivariate time series patterns modelled sample-point-wise by parametric fuzzy sets. An efficient classification-based approach for the online recognition of incompleted developing patterns in streaming time series is being presented. Furthermore, means are introduced to enable users of the recognition system to restrict results to certain stages of a pattern's development, e. g. for forecasting purposes, all in a consistently fuzzy manner.*

Keywords— Fuzzy classification, fuzzy automata, multivariate time series, pattern recognition.

1 Introduction

Fuzzy sets have been successfully used for a long time for a model-free or model-based representation of time series or time series sequences, e. g. for the classification and prediction of nonlinear systems behaviour [1–4]. This article aims at extending the fuzzy time series classification approach of [2, 3] to an *always aware* online recognition system for nonstationary recurring patterns in multivariate streaming time series. *Always aware* refers to the ability of the recognition system presented in this paper to constantly expect to find an (incompleted) pattern in a time series in any of its possible stages of development, which will be achieved by classification of pattern subsequences. It will be shown how the computationally expensive classification results for all possible subsequences can be obtained at almost no additional cost compared to the classification of completed patterns by relating the recognition procedure to ideas from fuzzy automata.

2 Fuzzy Modelling of Time Series Patterns

2.1 A Multivariate Parametric Fuzzy Set

In this section a parametric membership function type will be presented, which will serve as the common basis for the classifiers developed in this article. For the univariate case, the membership function is defined in (1). Based upon a potential function, it has already been successfully used in numerous applications such as process or medical surveillance [1], time series modelling and prediction [2, 3] and classifier networks [5]. One important property of this function is its capability of modelling asymmetric fuzzy sets by individual parameters for the left- and right-hand function branches.

$$\mu(x) = \begin{cases} \dfrac{a}{1+\left(\dfrac{1}{b_l}-1\right)\left(\dfrac{r-x}{c_l}\right)^{d_l}} & : \quad x < r \\ \dfrac{a}{1+\left(\dfrac{1}{b_r}-1\right)\left(\dfrac{x-r}{c_r}\right)^{d_r}} & : \quad x \geq r \end{cases} \tag{1}$$

The maximum truth value a of the fuzzy set occurs at its modal point: $\mu(x = r) = a$. For normalised fuzzy sets, $a = 1$ holds, as will be the case throughout this paper. The six parameters $b_{l/r}$, $c_{l/r}$ and $d_{l/r}$ determine the extent and shape of the class. From Fig. 1, the effect of b ($0 < b < 1$) and c ($c > 0$) can be understood. Semantically, they correspond to the class borders. The d parameter ($d \geq 2$) influences the shape of μ's descent to zero, with increasing d leading to a sharper descent and $d \to \infty$ resulting in a rectangular shape. This is particularly interesting since crisp sets may be represented by (1) as well.

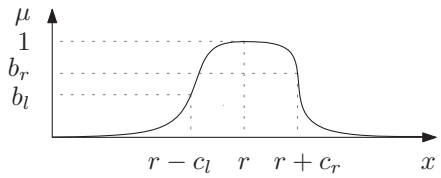


Figure 1: Normalised parametric membership function (1).

The parameters of this membership functions may be chosen based upon expert knowledge on the one hand, and automatically be computed [1] from a set of learning data on the other hand.[1] Both options benefit from the interpretability of these parameters.

A probably unique advantage of this function concept is that a multivariate membership function—based upon a conjunction of one-dimensional sets using a compensatory n-fold Hamacher intersection—may also be obtained in *parametric* form [1]. The n-fold operator is given by (2). The resulting multidimensional membership function is, for a simplified case of symmetric left- and right-hand branches, given by (3). Additionally (and omitted here for brevity), the class space may optionally be rotated in the underlying feature space for each class individually, leading to $(n-1)$ further rotation angle parameters, which can be obtained from a principal component analysis of the learning data. Fig. 2 depicts two examples of membership functions in a two-dimensional feature space.

$$\cap^{n}_{Ham}\,\mu_i = \frac{1}{\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\frac{1}{\mu_i}} \tag{2}$$

[1] The parameters of (1) are computed from data as follows: Firstly, the modal point r is being calculated by averaging the object values. The left- and right-hand side extents $c_{l/r}$ are chosen such that all learning objects, augmented by a so-called elementary fuzziness, lie within these class borders. Finally, the shape parameters b and d parameters are—if not specified manually—being computed to best fit the dispersion of the data.

ISBN: 978-989-95079-6-8

$$\mu(\boldsymbol{x}) = \frac{a}{1 + \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{b_i} - 1\right) \cdot \left|\frac{x_i - r_i}{c_i}\right|^{d_i}} \tag{3}$$

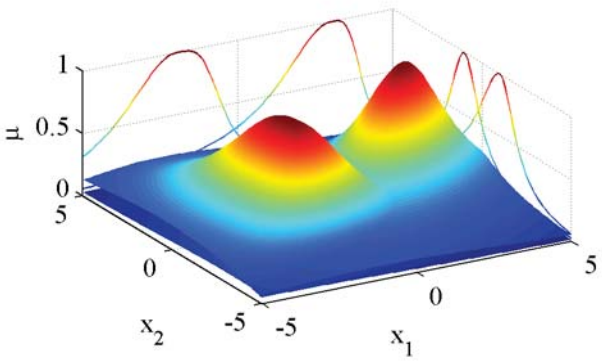


Figure 2: Two examples of two-dimensional fuzzy classes.

2.2 A Fuzzy Model for Time Series Patterns

The membership function concept introduced in section 2.1 forms the basis of a classifier model for multivariate time series sequences [2]. It allows to compute a continuous degree of similarity $\mu \in [0, 1]$ between a measured time series pattern and the fuzzy pattern model. Therefore pattern matching is being treated as a fuzzy classification task. The model described hereafter may be employed for equidistantly sampled univariate or multivariate time series.

Each point $\boldsymbol{x}(i)$ of a time series pattern (length: L sample points) is being described by a membership function $\mu_{P,i}(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^n$, $i = 1, \ldots, L$. This model therefore exhibits a soft tolerance towards noise and other sources of imprecision. To compare a time series sequence against the model, i. e. to classify a pattern given by L points $\boldsymbol{x}(1), \ldots, \boldsymbol{x}(L)$, the individual classification results of each point are combined to an overall degree of similarity $\mu \in [0, 1]$ by means of a fuzzy conjunction:

$$\mu = \mu_{P,1}(\boldsymbol{x}(1)) \cap \cdots \cap \mu_{P,L}(\boldsymbol{x}(L)) \tag{4}$$

If we recall the ideas of section 2.1 which led to the multivariate parametric membership function, it appears reasonable to reuse the same conjunction operator given by (2). In that manner, the overall classifier for the entire time series pattern could also be thought of as *one* fuzzy class created from the conjunction of $(L \cdot n)$ univariate fuzzy sets, forming an $(L \cdot n)$-dimensional membership function. Thusly classification of time series sequences would very well remain in line with the multivariate classification paradigm behind section 2.1. One key advantage of this model—especially in comparison to many *black box* models such as neural networks—is its support for partial classification, i. e. for the classification of any possible subsequence, simply by intersecting the underlying classification results for the sample points available. We will make use of this advantageous property later on.

Fig. 3 depicts an example of an univariate ($n = 1$) pattern (length $L = 286$ sample points) described by this fuzzy time series model, learned from a set of 14 instances of this pattern (*Coffee* dataset, available from [6]). By means of this figure we may also come to the interpretation of the pattern classifier as a corridor with soft boundaries. A measured pattern will, depending on its degree of similarity, more or less follow this corridor. The properties of the underlying membership function ensure that no pattern will be absolutely rejected ($\mu = 0$) by this classifier, thusly enabling a smooth, non-switching behaviour of the recognition system.

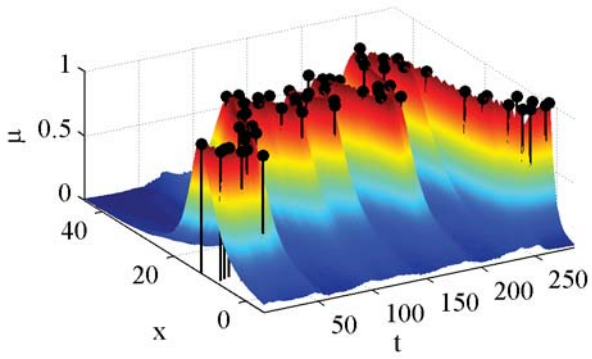


Figure 3: Fuzzy time series model along with a noisy pattern.

While the presented classifier aims at the recognition of recurring patterns that are similar in terms of absolute values, variations such as offsets or scaling may, at least to a certain extent, be covered by suitable preprocessing steps like normalisation. As in section 2.1, the time series pattern model may either be formulated based upon expert knowledge—eased by the interpretability of the individual parameters—or be result of a learning step using a set of pattern instances of the same phenomenon, employing the learning algorithm from section 2.1 for each point of the pattern. The latter is especially interesting in conjunction with clustering or motif mining algorithms. As there is no transform involved, the original properties of the pattern ensemble (energy content, statistics etc.) are mostly being preserved by this model. Furthermore, short-term forecasting of a time series with a partially elapsed pattern is made possible in a straightforward manner based upon the knowledge represented by the model.

3 Online Pattern Recognition

3.1 Task Description

The task of online pattern recognition—as we would like to understand it in the context of this work—can be described by means of Fig. 4. In a streaming univariate or multivariate time series $\boldsymbol{x}(t)$, we are looking for instances of a known pattern. *Online* pattern recognition firstly implies that an occurrence of the pattern has to be permanently expected. But in addition to that, we aim at detecting incompleted patterns, which would enable short-term predictions of the time series based on known patterns. For applications such as machine diagnosis, for instance, this facilitates preventive maintenance actions before a minor damage develops into a severe one.

At the current point in time t_{now} (as sketched in Fig. 4), the model containing the desired pattern has to be adjusted in its relative position τ such that the subsequence $\boldsymbol{x}([t_{\text{now}} - \tau] \ldots t_{\text{now}})$ optimally matches the partial model. τ then represents the time already elapsed in the pattern.

Summarisingly, an online pattern recognition system has to check if a pattern in a time series can be found at the current point in time, and if so, at which stage of development τ.

ISBN: 978-989-95079-6-8

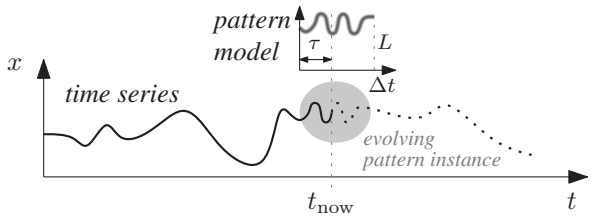


Figure 4: Online matching of evolving patterns.

3.2 Fuzzy Online Recognition

For reasons outlined above, we would like to treat the pattern matching problem as a fuzzy classification task, with time series subsequences (length τ, $\tau \leq L$) being input to a suitable classifier representing a pattern of length L. For each point in time t, this would yield classification results $\mu(t,\tau)$ for every possible value of τ. Semantically, $\mu(t_{\text{now}},\tau)$ represents the similarity of the time series subsequence $\boldsymbol{x}([t_{\text{now}}-\tau]\ldots t_{\text{now}})$ to the first part (length τ) of the pattern model/classifier.

In a truly fuzzy sense[2], *each* pattern would be found at *every* point in time in *all* possible stages of development τ, expressed by non-zero classification results $\mu(t,\tau) > 0\ \forall\tau$, even though $\mu(t,\tau)$ may, of course, be negligibly small. At any point in time t, fuzzy classification of evolving time series patterns therefore results in a *fuzzy set* $\mu(t,\tau)$—instead of just *one* truth value $\mu(t)$, as in a "conventional" fuzzy classification task, cf. Fig. 5. Representing the recognition result as a fuzzy set reflects the fuzziness of the recognition task. It is a joint fuzzy representation of the following information:

1. Could the pattern be found at the current point in time?
2. What is the pattern's state of development τ?

Both questions can—in real-world time series—not be answered precisely and unambigously. $\mu(t,\tau)$ commensurates with this vagueness. Furthermore, it may be interpreted as a fuzzy specification of τ, resembling the concept of fuzzy numbers.[3]

3.3 Representation as Fuzzy Automaton

In section 2.2, a suitable classifier model for patterns in equidistantly sampled time series has been introduced. As this model supports partial classification, it is well suited for the online pattern matching problem. In an online recognition system, however, obtaining $\mu(t,\tau)$ for every possible value of τ by classifying all respective subsequences $\boldsymbol{x}([t-\tau]\ldots t)$ would be a rather time-consuming procedure. More specifically, it appears incommensurate to recompute the whole fuzzy set $\mu(t,\tau)$ if only one new datum $\boldsymbol{x}(t)$ is being presented to the recognition system at time t. We therefore would like to reuse the classification result $\mu(t-1,\tau-1)$ of the subsequence

[2] As an unofficial definition, we would like to understand "true" fuzziness as a property of models and methods which do not (explicitely or implicitely) perform crisp decisions or cut-off operations. For instance, a "truly" fuzzy set has infinite support, thusly all elements x of the underlying universe of discourse belong to this set with $\mu(x) > 0$. As a counterexample, triangular fuzzy sets feature crisp and finite extents, and would therefore not be regarded as "truly" fuzzy models.

[3] Although not being a fuzzy number in the sense of Dubois and Prade [7], especially owing to the very likely multimodality, it matches the spirit of a fuzzy number as a fuzzified representation of a real-valued number (here: τ) quite well.

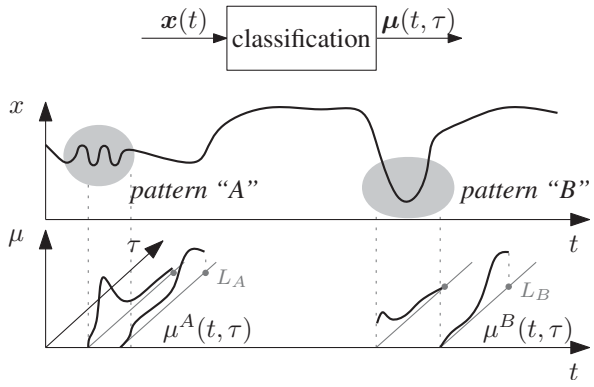


Figure 5: Fuzzy online recognition of two patterns "A" and "B" (length L_A and L_B). At each point in time t, the maximum truth values of the recognition results $\mu^{A/B}(t,\tau)$ w. r. t. τ point to the relative positions τ associated to the best match of the partially elapsed patterns. The fuzzy sets $\mu^{A/B}$ are shown at two different stages of development τ for each pattern.

$\boldsymbol{x}([(t-1)-(\tau-1)]\ldots t-1)$ when computing $\mu(t,\tau)$ for $\boldsymbol{x}([t-\tau]\ldots t)$.

For discrete time patterns of length L, the pattern recognition result $\mu(t,\tau)$ consists of L truth values ($\tau = 1,\ldots,L$). This result may be represented by means of a fuzzy state automaton [8], as sketched in Fig. 6. Being a fuzzy automaton, each state—corresponding to a certain value of τ—is active to a certain degree $\mu(t,\tau)$. At each point in time t, the automaton is triggered with a discrete event in the form of a new input value $\boldsymbol{x}(t)$, leading to new activation levels for each state. An activation level $\mu(t,\tau)$ depends on the current input value $\boldsymbol{x}(t)$ and the former activation level of the previous state:

$$\mu(t,\tau) = f(\mu(t-1,\tau-1),\boldsymbol{x}(t)) \tag{5}$$

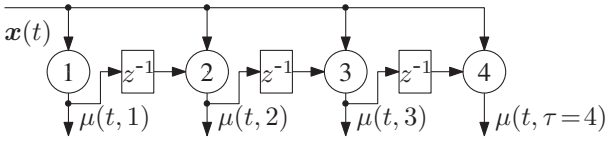


Figure 6: (Informal) fuzzy state automaton representation of a time series pattern recognition system, here for a pattern consisting of four sample points.

Recalling the internals of the time series pattern classifier from section 2.2, the overall classification result was gained from a compensatory Hamacher conjunction (cf. (4)) of individual classification results $\mu_{P,i}$ of time series sample points $\boldsymbol{x}_i$. Thusly $\mu(t-1,\tau-1)$ consists of a conjunction of $\tau-1$ truth values. $\mu(t,\tau)$ could therefore be obtained from a conjunction of $\mu(t-1,\tau-1)$ and the partial classification result $\mu_{P,\tau}(\boldsymbol{x}(t))$ for the τ^{th} point of the time series pattern. However, this conjunction should reflect the weight of the $\tau-1$ truth values that contributed to $\mu(t-1,\tau-1)$, so that $\mu_{P,\tau}(\boldsymbol{x}(t))$ does not outweigh all previous information. A suitable conjunction operator will be derived in the following section, which will enable us to rewrite (5) to such a weighted conjunction:

$$\mu(t,\tau) = \mu(t-1,\tau-1)\ {}^{(\tau-1)}\cap^{1}\ \mu_{P,\tau}(\boldsymbol{x}(t)) \tag{6}$$

ISBN: 978-989-95079-6-8

Equation (6) forms the basis for the mode of operation of the fuzzy automaton from Fig. 6, and will enable us to perform online recognition of time series patterns much more efficiently.

3.4 Weighted Compensatory Hamacher Conjunction

The compensatory Hamacher conjunction of n truth values μ_i $(i = 1, \ldots, n)$ has already been introduced in (2). By rearranging the components of (2) into two sets of truth values, we may formulate the compensatory Hamacher conjunction of $n = n_1 + n_2$ truth values as:

$$\begin{aligned}\mu_1 \cap \cdots \cap \mu_n &= \frac{1}{\frac{1}{n_1+n_2}\left(\sum_{i=1}^{n_1}\frac{1}{\mu_i} + \sum_{j=n_1+1}^{n}\frac{1}{\mu_j}\right)} \\ &= \frac{1}{\frac{1}{n_1+n_2}\left(\frac{n_1}{\mu_1 \cap \cdots \cap \mu_{n_1}} + \frac{n_2}{\mu_{n_1+1} \cap \cdots \cap \mu_n}\right)} \\ &= (\mu_1 \cap \cdots \cap \mu_{n_1})^{\,n_1 \cap^{n_2}}\ (\mu_{n_1+1} \cap \cdots \cap \mu_n)\end{aligned} \tag{7}$$

Based upon (7) we may now define the (n_a, n_b)-weighted compensatory Hamacher conjunction of two truth values μ_a and μ_b, which was in a similar manner also derived in [9]:

$$\mu_a \,^{n_a \cap^{n_b}}\ \mu_b = \frac{1}{\frac{1}{n_a+n_b}\left(\frac{n_a}{\mu_a} + \frac{n_b}{\mu_b}\right)} \tag{8}$$

Thusly it becomes possible to compute the conjunction of new truth values with existing conjunction results without having to store the individual truth values leading to the existing results and while exactly retaining the weight of each truth value that contributed to the overall result.

Finally, we may define the vectorial weighted conjunction of two equal-sized truth vectors $\boldsymbol{\mu}_a$ and $\boldsymbol{\mu}_b$ using weight vectors $\boldsymbol{n}_a$ and $\boldsymbol{n}_b$ containing the individual weights for the respective elements:

$$\boldsymbol{\mu}_a \,^{\mathbf{n_a} \cap^{\mathbf{n_b}}}\ \boldsymbol{\mu}_b = \begin{pmatrix} \vdots \\ \dfrac{1}{\frac{1}{n_{a,i}+n_{b,i}}\left(\frac{n_{a,i}}{\mu_{a,i}} + \frac{n_{b,i}}{\mu_{b,i}}\right)} \\ \vdots \end{pmatrix} \tag{9}$$

3.5 Update Equations for the Fuzzy Automaton

Recalling (6) and employing the weighted conjunction operator from (8), we may now formulate the update equation for each fuzzy state of the automaton representation of the recognition results $\mu(t, \tau)$ from section 3.3. Equation (6) holds for $\tau \geq 2$, as the first state of the automaton does not have a predecessor state to obtain $\mu(t-1, \tau-1)$. Since the weight of the non-existing zero-th state would be zero anyway, (6) can (for $\tau = 1$) be thought of to reduce to solely the classification result for the first sample point, cf. (10). All further results $\mu(t, \tau \geq 2)$ are computed by a weighted conjunction of antecedent recognition results and the classification results $\mu_{P,\tau}$ of the new sample point $\boldsymbol{x}(t)$ available at each point in time.

$$\mu(t, 1) = \mu_{P,1}(\boldsymbol{x}(t)) \tag{10}$$

To obtain a vectorial form of the update equations, we define $\boldsymbol{\mu}_P(t)$ as the vector containing the individual membership values of $\boldsymbol{x}(t)$ to all classes constituting the fuzzy time series pattern model, and $\boldsymbol{\mu}_\tau(t)$ as the state vector containing all levels of activation of the automaton, and thusly the overall pattern recognition results:

$$\boldsymbol{\mu}_P(\boldsymbol{x}(t)) = \begin{pmatrix} \mu_{P,1}(\boldsymbol{x}(t)) \\ \vdots \\ \mu_{P,L}(\boldsymbol{x}(t)) \end{pmatrix} \tag{11}$$

$$\boldsymbol{\mu}_\tau(t) = \begin{pmatrix} \mu_{\tau,1}(t) \\ \vdots \\ \mu_{\tau,L}(t) \end{pmatrix} = \begin{pmatrix} \mu(\tau, 1) \\ \vdots \\ \mu(\tau, L) \end{pmatrix} \tag{12}$$

Summarisingly, the recursive update equation for the recognition result (the fuzzy set $\mu(t, \tau)$) or its vectorial representation $\boldsymbol{\mu}_\tau(t)$ can be condensed from (5), (6) and (9) to (13):

$$\begin{aligned}\boldsymbol{\mu}_\tau(t) &= f(\boldsymbol{\mu}_\tau(t-1)), \boldsymbol{\mu}_P(\boldsymbol{x}(t))) \\ &= \left(\begin{pmatrix}1\\ \mathbf{0}\end{pmatrix} + \begin{pmatrix}\mathbf{0}^T & 0\\ \boldsymbol{I} & \mathbf{0}\end{pmatrix} \cdot \boldsymbol{\mu}_\tau(t-1)\right)^{\,\mathbf{n}_\tau \cap \mathbf{1}}\ \boldsymbol{\mu}_P(\boldsymbol{x}(t))\end{aligned} \tag{13}$$

$$\boldsymbol{n}_\tau = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ L-1 \end{pmatrix} \tag{14}$$

Incorporating the weight vector $\boldsymbol{n}_\tau$, this finally results in:

$$\boldsymbol{\mu}_\tau(t) = \begin{pmatrix} \dfrac{1}{\frac{1}{1}\cdot\left(\frac{0}{1} + \frac{1}{\mu_{P,1}(\boldsymbol{x}(t))}\right)} \\ \vdots \\ 1 \\ \dfrac{1}{\frac{1}{i}\cdot\left(\frac{i-1}{\mu_{\tau,i-1}(t-1)} + \frac{1}{\mu_{P,i}(\boldsymbol{x}(t))}\right)} \\ \vdots \\ 1 \\ \dfrac{1}{\frac{1}{L}\cdot\left(\frac{L-1}{\mu_{\tau,L-1}(t-1)} + \frac{1}{\mu_{P,L}(\boldsymbol{x}(t))}\right)} \end{pmatrix} \tag{15}$$

Apart from the $(L-1)$ weighted conjunctions of truth values that are required at each time step, the online recognition of a (partially elapsed) fuzzy pattern therefore does not require more underlying classification steps and computations of $\mu_{P,\tau}$ than would be needed just for the fuzzy classification and recognition of only completed patterns in streaming time series.

4 Post-Processing Recognition Results

4.1 Windowing of $\mu(t, \tau)$

The fuzzy set $\mu(t, \tau)$—describing the recognition results of a pattern found at the current point in time t in all possible stages of development τ—is a rather complex result containing a wealth of information. In many cases it will thusly have to be narrowed down to particularly interesting or relevant parts. In addition to that, shorter subsequences of the pattern will be found more often in a time series than the whole pattern. This results in a tendency to relatively often obtain high degrees

ISBN: 978-989-95079-6-8

of similary $\mu(t,\tau)$ for small values of τ, which will later on diminish if the time series does not develop further into an instance of the pattern. For longer patterns, a user will therefore always have to gauge and find a trade-off between the necessity of detecting patterns at an early stage of development τ (with a higher risk of a false or premature recognition result) and the importance of a reliable recognition result (which can ultimately only be achieved at the end of an evolving pattern). This trade-off cannot be universally formulated in a crisp manner, and neither does it seem advisable to do so.

These considerations lead us to the idea of letting the user of the recognition system focus on certain stages development τ of a pattern in a soft manner, such that the recognition system only reports matches of the evolving pattern for values of τ falling within certain (fuzzy) boundaries. Subsequently, we will call these boundaries the *fuzzy window of interest for* τ. We will represent this window of interest by means of a fuzzy set $\mu_w(\tau)$ defined over all possible values of τ, with $\mu_w(\tau) \to 0$ indicating low interest in a recognition result for this particular value of τ, and $\mu_w(\tau) \to 1$ indicating high interest.

To obtain a *windowed* pattern recognition result $\tilde{\mu}(t,\tau)$, the window of interest $\mu_w(\tau)$ and the recognition result $\mu(t,\tau)$ are combined as in (16) by a conjunction[4] of the two fuzzy sets. Semantically, this corresponds to a coincidence of a pattern being found at a certain stage of development and the user's subjective assessment of the importance of this particular stage.

$$\tilde{\mu}(t,\tau) = \mu(t,\tau) \cap \mu_w(\tau) \quad \forall\tau \tag{16}$$

Equation (16) in its general form includes some interesting special cases, such as the following: By definition, the truth values in $\mu(t,\tau)$ appear equally important to the user regardless of τ, so that higher values of μ will often be found for smaller values of τ, although the pattern has just begun to evolve. Following the windowing approach would also enable us to (for example: linearly) weigh a recognition result $\mu(t,\tau)$ according to the stage of development τ of the pattern by chosing a suitable $\mu_w(\tau)$. To achieve this, $\mu_w(\tau)$ has—in this special case—to be chosen as a fuzzy set with (linearly) increasing truth values, reaching $\mu_w(\tau = L) = 1$ at the end of the pattern (length L). If we, as an example, choose the algebraic product for the operator $\cap$ in (16), we would simply obtain (17) for a thusly weighted/windowed recognition result.

$$\tilde{\mu}_{linear}(t,\tau) = \mu(t,\tau) \cdot \frac{\tau}{L} \tag{17}$$

4.2 *Deriving Decisions*

In order to further process the (windowed) recognition results $\mu(t,\tau)$ or $\tilde{\mu}(t,\tau)$, e. g. when serving as a rationale for subsequent actions, it will often be necessary to come to a crisp decision regarding the similarity and stage of development of a pattern found in a time series. In these cases, the fuzzy set $\tilde{\mu}(t,\tau)$ has to be condensed to one or more crisp parameters.

Although many defuzzification methods for fuzzy sets are available [10], we would—for the sake of simplicity—like to restrict ourselves in this article to a maximum approach (first of maxima, FOM). The resulting crisp value τ^* from $\tilde{\mu}(t,\tau)$ points to the earliest stage of development of the pattern yielding the highest degree of similarity $\tilde{\mu}^*$ to the respective time series subsequence, both of which may be represented as a fuzzy singleton $\tilde{\mu}^*(t,\tau^*)$ for each point in time t.

5 Example

Fig. 7 shows a time series containing one instance of the pattern modelled and depicted in Fig. 3. For illustrative purposes, the instance has been highlighted, as the surrounding parts of the the time series are relatively similar in amplitude and shape.

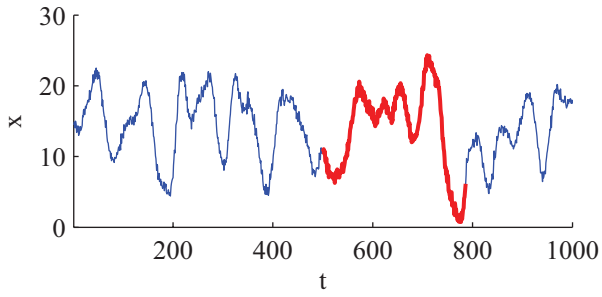


Figure 7: Univariate time series with a pattern instance.

Employing the fuzzy model from Fig. 3, we are now performing an online recognition of this pattern in the time series of Fig. 7. While advancing in time t through all points of the time series, the pattern is constantly being expected to be found in all[5] possible stages of development τ. The classification results for each point in time t—evolving fuzzy sets $\mu(t,\tau)$ defined over τ—are being shown in Fig. 8.

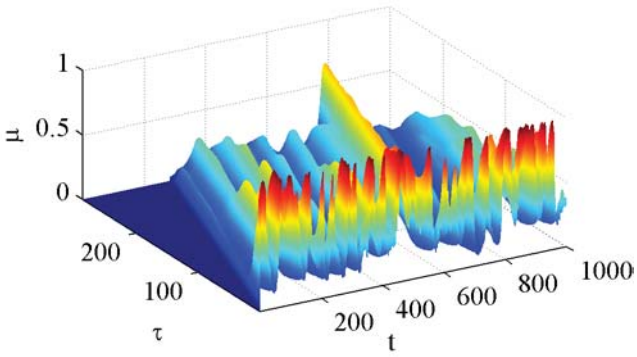


Figure 8: Results for an online recognition of the time series pattern of Fig. 3.

These recognition results $\mu(t,\tau)$ reveal high degrees of similarity of time series subsequences to the first part (i. e. small values of τ) of the pattern, and, in fact, it can be understood from Fig. 7 that the ascending first slope can (with slight variations) be found repeatedly throughout the time series. Apart from that, the progression of a detected instance is well visible in the course of $\mu(t,\tau)$. Only once a completed pattern is being found with distinctly high values for $\mu(t,\tau = L)$, which point to the pattern instance embedded in this time series.

To detect a pattern ahead of time ($\tau < L$), e. g. for forecasting purposes, the recognition result has to be well-founded, and therefore based upon a adequately large number of recognised

[4] As conjunction operator, a non-compensatory operator—such as all T-norm operators—should be chosen to ensure $\tilde{\mu}(t,\tau) \leq \mu(t,\tau)$.

[5] In the first part of the time series, the pattern may only be found at early stages of development $\tau < t$, i. e. $\mu(t,\tau > t) = 0$ must hold.

ISBN: 978-989-95079-6-8

sample points τ on the one hand, whilst τ being small enough to leave sufficient scope for actual forecasting. We will tackle both problems of "premature" pattern recognition and formulating a compromise for a suitable value of τ delivering usable recognition results by following the windowing approach for $\mu(t,\tau)$ presented in section 4.1. As an example, we would like to concentrate on recognition results for pattern instances being at least half-way, but not yet fully completed. Once again employing the unimodal parametric membership function from (1), we formulate a fuzzy window of interest $\mu_w(\tau)$ for this certain stage of the pattern's development, cf. Fig. 9.[6] Due to the fuzzy nature of the presented method, slight variations of $\mu_w(\tau)$ and its parameters will not alter the overall system behaviour and recognition results in a drastic, switching manner. This eases the design of $\mu_w(\tau)$, as no "critical" decisions have to be made when choosing its parameters.

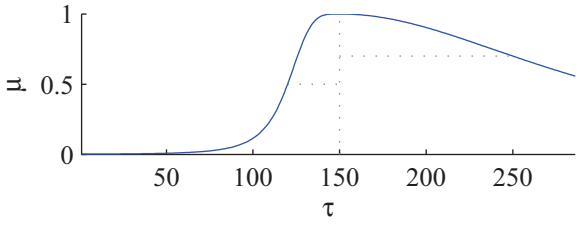


Figure 9: Fuzzy window of interest $\mu_w(\tau)$ for pattern recognition results $\mu(t,\tau)$ from Fig. 8.

For each t, the recognition result (fuzzy set) $\mu(t,\tau)$ was processed as given by (16) using a Hamacher product with $\mu_w(\tau)$. The windowed recognition results $\tilde{\mu}(t,\tau)$ are shown in Fig. 10. As a final step, a simple decision strategy was followed to generate crisp recognition results in singleton form for each point in time t. For each fuzzy set $\tilde{\mu}(t,\tau)$, the modal (maximum) value $\tilde{\mu}^*(t,\tau^*)$ was determined. Any result for τ^* with a corresponding truth value falling below a threshold $\tilde{\mu}^*(t,\tau^*) < 0.5$ was discarded. Results of this decision procedure are marked in Fig. 10. As can be seen in this figure, two candidate pattern instances are being identified at the desired stage of development τ, the first one of which vanishing quickly. The actual pattern instance embedded in the time series of Fig. 7 is being recognised reliably throughout the fuzzy window of interest.

6 Conclusions

This article dealt with the online recognition of patterns in multivariate time series. The vectorial recursive update equation for the recognition results derived in section 3.5 allows to obtain a fuzzy recognition result $\mu(t,\tau)$ for all possible stages of development τ of a pattern at almost the same computational cost as required for the fuzzy classification of only completed patterns. By formulating and applying a fuzzy window of interest, the recognition result $\mu(t,\tau)$ can be narrowed down to parts relevant for the respective application in a fuzzy manner. This also eases a subsequent crisp decision step.

Further work will concentrate on the application of the model and methods presented here to the recognition and prediction

[6] The parameter were chosen manually: $r = 150$, $b_l = 0.5$, $b_r = 0.7$, $c_l = 30$, $c_r = 100$, $d_l = 4$, $d_r = 2$. For $\tau < r$, $d_l = 4$ and $c_l = 30$ ensure a quick decrease of $\mu_w(\tau)$, whereas the right-hand side parameters b_r, c_r, d_r provide a gentle decrease of μ_w for $\tau > r$.

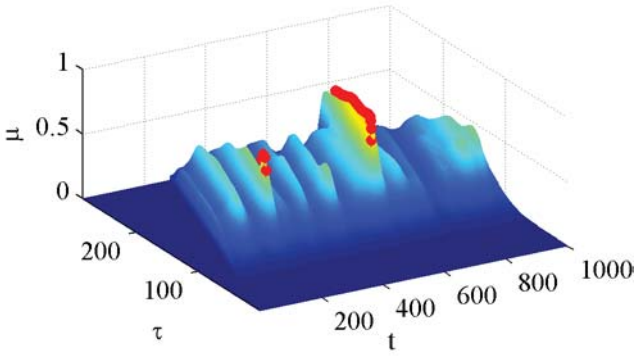


Figure 10: Windowed recognition results $\tilde{\mu}(t,\tau)$. Singleton results $\tilde{\mu}^*(t,\tau^*)$ of the decision procedure are marked with dots (slightly increased for better visibility).

of recurring patterns in energy load time series. Besides that, an enhancement of the decision procedure appears promising. A desirable goal would be a *fuzzy* decision process delivering *parametric* fuzzy information for the stage of development of a detected pattern. This could be more directly employed in subsequent (e. g. prediction) steps without discarding the fuzziness always inherent in a real-world recognition result.

References

[1] Steffen F. Bocklisch and Norman Bitterlich. *Fuzzy systems in computer science*, chapter Fuzzy pattern classification – methodology and application. Vieweg, Wiesbaden, 1994.

[2] Michael Päßler and Steffen F. Bocklisch. *Fuzzy Control: Theory and Practice*, chapter Fuzzy time series analysis, pages 331–345. Springer Verlag, 2000.

[3] Peter Bretschneider. *Ein Beitrag zur Vorhersage musterbasierter nichtlinearer stochastischer Signale*. PhD thesis, TU Ilmenau, 2002.

[4] Bernd Möller and Uwe Reuter. Prediction of uncertain structural responses using fuzzy time series. *Computers and Structures*, 86(10):1123–1139, 2008.

[5] Arne-Jens Hempel and Steffen F. Bocklisch. A conception for networks of fuzzy pattern classifiers. In *Proceedings 13th East West Fuzzy Colloquium*, Zittau, 2006.

[6] Eamonn Keogh, Xiaopeng Xi, Li Wei, and Chotirat Ann Ratanamahatana. The UCR time series classification/clustering homepage. http://www.cs.ucr.edu/~eamonn/time_series_data/, 2006.

[7] Didier Dubois and Henry Prade. Fuzzy numbers: An overview. In James C. Bezdek, editor, *Analysis of Fuzzy Information, Vol.I: Mathematics and Logic*, pages 3–39. CRC Press, Boca Raton, FL, 1987.

[8] William G. Wee and King-Sun Fu. A formulation of fuzzy automata and its application as a model of learning systems. *IEEE Transactions on Systems Science and Cybernetics*, 5(3):215–223, 1969.

[9] Ullrich Scheunert. *Fuzzy-Mengen-Verknüpfung und Fuzzy-Arithmethik zur Sensor-Daten-Fusion*. PhD thesis, Chemnitz University of Technology, 2001.

[10] Werner Van Leekwijck and Etienne E. Kerre. Defuzzification: criteria and classification. *Fuzzy Sets and Systems*, 108(2):159–178, 1999.

ISBN: 978-989-95079-6-8